Journal of Affective Disorders Reports 8 (2022) 100333

Contents lists available at ScienceDirect

Journal of Affective Disorders Reports

journal homepage: www.sciencedirect.com/journal/journal-of-affective-disorders-reports

Research Paper

The Stirling Antenatal Anxiety Scale (SAAS): development and initial psychometric validation

Andrea Sinesi [a,*], Helen Cheyne [b], Margaret Maxwell [c], Ronan O'Carroll [d]

[a] Nursing, Midwifery and Allied Health Professions Research Unit (NMAHP RU), University of Stirling, UK
[b] NMAHP RU, University of Stirling, UK
[c] NMAHP RU, University of Stirling, UK
[d] Department of Psychology, Faculty of Health Sciences and Sport, University of Stirling, UK

ARTICLE INFO

Keywords:
Antenatal anxiety
Pregnancy
SAAS
Diagnostic accuracy
Psychometric properties
GAD-2

ABSTRACT

Background: Anxiety in pregnancy affects approximately 15% of all pregnant women. Antenatal anxiety can impact negatively on subsequent child development and is a strong predictor of postnatal depression. No anxiety measures with sound psychometric properties are currently available for screening pregnant women. This study aimed to develop and evaluate the psychometric properties of a brief questionnaire specifically devised for use by healthcare professionals and researchers to screen for antenatal anxiety.

Methods: A mixed-method study for scale development which included: 1) a review examining the psychometric properties of anxiety measures 2) qualitative interviews with women with experience of antenatal anxiety 3) a Delphi study to achieve consensus amongst experts in relation to questions to be included in the scale. For the psychometric validation, 174 women completed the new 10 item scale and the GAD-2/7. A sub-sample of women were also assessed using a diagnostic interview.

Results: The Stirling Antenatal Anxiety Scale (SAAS) showed very good diagnostic accuracy (sensitivity=91%; specificity=85%). The GAD-2 performed significantly worse (sensitivity=27%; specificity=96% at the recommended cut off-score) than the new antenatal anxiety scale in identifying women with an anxiety disorder.

Limitations: The key limitation of the research was the relatively small sample size of women assessed with a diagnostic interview for the psychometric validation study.

Conclusions: Identifying pregnant women experiencing problematic anxiety is the first, crucial step in providing them with an appropriate level of support. This research provides health professionals with a reliable, brief, and easy-to-complete screening scale for anxiety in pregnancy.

1. Introduction

Anxiety during pregnancy is highly prevalent (Fairbrother et al., 2016; Rubertsson et al., 2014), affecting between 15 and 23% of women (Dennis et al., 2017). Antenatal anxiety increases the risk of negative outcomes for mother and child (Dunkel Schetter and Tanner, 2012; Goodman et al., 2014) and is a strong predictor of postnatal depression (Lee et al., 2007). Its potential detrimental effects on birth and child outcomes include increased rates of low birth weight, premature birth (Ding et al., 2014) and a higher risk of a range of negative child developmental outcomes, including poorer cognitive development (Ibanez et al., 2015) and behavioural and affective problems (Capron et al., 2015; Leis et al., 2014; O'Donnell et al., 2014).

Over the last two decades antenatal anxiety has thus become the focus of growing research and clinical attention (Brouwers et al., 2001; Goodman et al., 2014). Despite its high prevalence, detection rates in maternity care are estimated to be lower than 50% (Bauer et al., 2014; Biaggi et al., 2016). In the UK, the National Institute for Health and Care Excellence [NICE] guidance on perinatal mental health indicated that anxiety disorders often go unrecognised and thus untreated throughout the perinatal period (National Institute for Health and Care Excellence, 2014). In 2014 NICE introduced two screening questions for perinatal anxiety (GAD-2: Generalised Anxiety Disorder-2, Kroenke et al., 2007). For further assessment, NICE also recommended use of the longer GAD-7

* Corresponding author at: Nursing, Midwifery and Allied Health Professions, Research Unit (NMAHP RU), Pathfoot Building, University of Stirling, Stirling, FK9 4LA, UK.
E-mail address: andrea.sinesi@stir.ac.uk (A. Sinesi).

https://doi.org/10.1016/j.jadr.2022.100333
Received 8 November 2021; Received in revised form 18 February 2022; Accepted 24 February 2022
Available online 27 February 2022

(NICE, 2014). This recommendation, however, drew heavily on the evidence base from guidance for screening tools in non-pregnant populations (National Institute for Health and Care Excellence, 2011), as critically there is no psychometric evidence to support the use of the GAD-2 in perinatal women (Nath et al., 2018; Austin et al., 2021).

Recent reviews have reported that anxiety self-report scales used in studies with pregnant women (Evans et al., 2015; Meades and Ayers, 2011; Sinesi et al., 2019) lack satisfactory psychometric properties and have limited applicability to the antenatal period. A key limitation of many anxiety scales developed for the general population relates to their emphasis on physical symptoms and their potential confounding role when questions on somatic symptoms are used to screen for anxiety during pregnancy (Biaggi et al., 2016). A further significant issue is the occurrence of 'pregnancy-related anxiety', which has been proposed as a specific and distinct syndrome (Huizink et al., 2004). In pregnancy-related anxiety (PrA), symptoms of anxiety are specifically focused on pregnancy and childbirth and may include persistent worries about personal appearance and health, health of the foetus, fear around labour and delivery, and about future parenting (Orr et al., 2007; Dunkel Schetter and Tanner, 2012). Some level of worry and anxiety is normal and adaptive during pregnancy (Haines et al., 2015). It is thus important not to consider these common concerns as indicators of pathological or problematic anxiety. However, if these fears and worries become persistent or particularly distressing they can have a detrimental impact on a woman's psychological wellbeing over the course of pregnancy (Wijma and Wijma, 2017). While this anxiety type is not covered by standard diagnostic classifications, it is of clinical significance and has been shown to be an independent predictor of negative outcomes for mother and child, such as lower gestational age and birth weight, and maternal postnatal mood disorders (Blackmore et al., 2016).

Recent increased attention of researchers to PrA has resulted in the development of scales such as the Perinatal Anxiety Screening Scale (PASS: Somerville et al., 2014) and the Pregnancy-related Anxiety Scale (PrAS: Brunton et al., 2019). However, a key clinical limitation of these scales is their length, making the 31 item PASS and the 32 item PrAS unfeasible for use in routine maternity services. In the UK, NICE only considers brief scales, defined as those containing less than 12 items, as potentially feasible to implement in maternity care settings (NICE, 2014).

We thus aimed to develop and validate a brief, psychometrically robust self-report scale specifically to screen for a range of problematic anxiety symptoms in pregnant women, including both anxiety disorders and PrA, that would be feasible to use in research and routine antenatal care.

2. Method

The development and psychometric testing of the new scale was based on guidance on scale development proposed by DeVellis (2012) and Streiner et al. (2008). The three-phase process is summarised in Fig. 1. In Phase three the new scale (Stirling Antenatal Anxiety Scale: SAAS) and the GAD-2/7 were tested against a structured diagnostic interview for anxiety disorders in a cross-sectional study to determine their diagnostic accuracy.

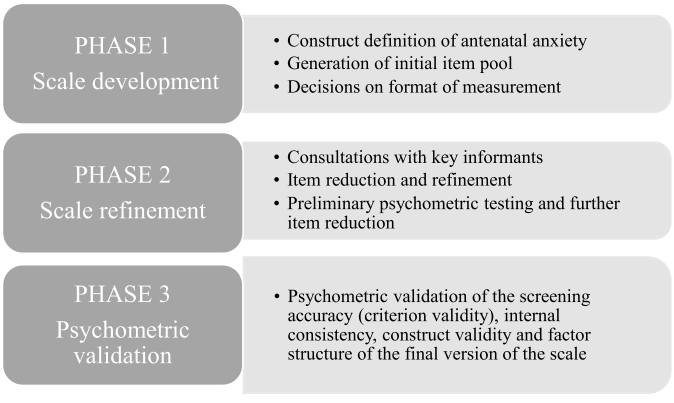

Fig. 1. Phases of the development and psychometric validation of the SAAS.

2.1. Phase 1 – Scale development

The aim of the initial phase in scale development is to define the construct for measurement and to generate an initial pool of candidate items for potential inclusion (Netemeyer et al., 2003). Two studies were conducted to inform this initial phase. 1. A systematic review examined the psychometric properties and item content of anxiety scales used in studies with pregnant populations. The review, reported elsewhere (Sinesi et al., 2019), mapped a set of core anxiety symptoms and domains (i.e. content areas) showing good or excellent evidence of their psychometric value when used to assess general anxiety or PrA in pregnant populations. The anxiety domains identified included symptoms of generalised worry, panic, fear of childbirth and persistent worries specifically related to pregnancy (Sinesi et al., 2019). A full list of anxiety items and domains is presented as supplementary material (S1). 2. Twelve semi-structured interviews were conducted with women with lived experience of problematic anxiety symptoms during pregnancy. The aim was to explore key symptoms of the target construct and the different factors useful to distinguish between normal experiences of anxiety and worries in pregnancy and elevated levels of anxiety. Thematic analysis revealed a range of anxiety symptoms that were identified as potentially important indicators of antenatal anxiety. These could be categorised into five higher-order anxiety domains (Silverman, 2001), as follows: 1) *Worry and anxious apprehension* 2) *Fear/Panic* 3) *Pregnancy-related anxiety* 4) *General distress* and 5) *Anxiety-driven behaviours*. Supplementary material S2 illustrates all the identified themes and anxiety domains.

Findings were used to inform the construct definition of antenatal anxiety (Text box 1) and guide the generation of an initial item pool to reflect the proposed construct. The combination of different sources of evidence (i.e. psychometric literature and target population) ensured a comprehensive coverage of the range of problematic anxiety symptoms that pregnant women may experience, thus contributing to maximise the content and construct validity in the initial item pool (Simms, 2008; Furr, 2011).

Notably, in contrast with several authors (Blair et al., 2011; Huizink et al., 2004) who have proposed that pregnancy-related anxiety should be considered a specific and entirely distinct syndrome, the definition of antenatal anxiety proposed in this study considers pregnancy-related anxiety as one of the possible dimensions of the target construct, which may or may not be present in women experiencing antenatal anxiety. A five-point Likert scale, measuring frequency of symptoms over the past two weeks, with the response options being 'Never' = 1, 'Rarely' = 2, 'Sometimes' = 3, 'Often' = 4 and 'Always' = 5, was chosen as the scale format for the initial item pool.

Candidate items for potential inclusion in the SAAS were generated in order to represent the entire range of anxiety symptoms identified as potentially important indicators of antenatal anxiety. Specifically, 37 items were written de novo and 15 items were included from an existing anxiety item bank (i.e. PROMIS: Patient-Reported Outcomes Measurement Information System; Pilkonis et al., 2011). All newly written items were generated based on the wording used by women interviewed (Kline, 2005; Streiner et al., 2008) and phrased to be consistent with the included PROMIS items. To maximise the content validity of the scale, expert opinion was sought from three clinicians working in the area of perinatal mental health as a further source of item generation. Seven additional items were suggested. A total of 59 items were thus included in the initial item pool (Supplementary material S3).

Text box 1 – Definition of the construct of antenatal anxiety

Antenatal anxiety can be defined as the experience of clinically significant symptoms of anxiety in pregnant women. The term clinically significant is used to indicate that the symptoms are sufficiently problematic to: 1.Be perceived as distressing 2.Have a negative impact on at least one area of individual functioning 3.Be experienced for a sufficiently prolonged period of time (i.e. *not limited to a temporary reaction to a specific event or situation).*

Antenatal anxiety can manifest as the experience of one or more of these symptom domains: a) Excessive or generalised worry b) Repetitive thoughts and rumination c) Feelings of panic or intense fear d) Specific symptoms of general distress e) Specific symptoms of pregnancy-related anxiety f) Behavioural avoidance of specific places or situations.

2.2. Phase 2 – Scale refinement and item reduction

The item pool was initially reviewed for face validity by five women with experience of perinatal mental health problems who provided comments on the clarity and acceptability of items. Four items which scored on average lower than 9 on a 1–10 in a feedback survey in relation to either their clarity or acceptability were reworded. The same women also commented on the design of subsequent versions of the SAAS.

2.2.1. Delphi study

An initial phase of item reduction was performed. A group of 16 experts working in the area of perinatal mental health (including clinical psychologists, perinatal psychiatrists, midwives with special expertise in perinatal mental health and mental health nurses) took part in two rounds of a Delphi study (Hsu and Sandford, 2007). Experts were asked to "*rate each item according to how much you consider it to be an important indicator of problematic anxiety in pregnant women based on your clinical experience*". The Likert scale used to collect data ranged from 1 to 9, with 1–3 indicating 'limited importance', 4–6 indicating an item 'important but not essential', and 7–9 to indicate items considered 'essential'. Through this procedure we discarded items considered to be less relevant for the assessment of the target construct as well as problematic items. Thirty items meeting pre-defined criteria (average rating of at least 6.50 and more than 50% of panellists rating the item in the 'essential', 7–9 category) were selected for the final phase of scale development (see Supplementary material S4).

2.2.2. Preliminary psychometric testing

A cross-sectional pilot study of the 30-question SAAS was conducted with the aim to reduce the scale to a final, shorter and psychometrically robust version. The 30-item SAAS was completed by 62 pregnant women recruited from one NHS Board in Scotland. The sampling, recruitment and part of the data collection procedures of this pilot study were identical to those described for the psychometric validation study (Phase 3) and are thus not discussed here.

Preliminary psychometric testing through item analysis was performed by discarding items which did not significantly contribute to improve the psychometric properties of the scale, while retaining those that were psychometrically robust and appeared to adequately capture the target construct (DeVellis, 2012). Item statistics including mean score, standard deviation and response distribution for each item were inspected to identify items showing floor or ceiling effect. Items for which the lowest or highest response option was endorsed by $> 90\%$ of study participants were discarded (Streiner et al., 2008). In relation to item-total correlations, various scholars agree that items with an item-total correlation < 0.30 should be discarded (Kline, 2005; Abell et al., 2009). This criterion was therefore adopted in this study. For inter-item correlations the criterion used in this study was to inspect all inter-item correlations equal or above 0.75 for suspect redundancy (De Vaus, 2014). Both the analysis of item statistics and considerations related to the contribution of each item to a short measure that retained construct relevance and a sufficiently broad scope were considered. This process resulted in a final version of the SAAS consisting of 10 items, with seven items assessing general anxiety symptoms and three PrA symptoms. The 10-item SAAS is presented in Appendix 1.

2.3. Phase 3 – psychometric validation of SAAS and GAD-2/7

2.3.1. Study design

The final, 10-item version of the SAAS and the GAD-2/7 were psychometrically tested in a sample of pregnant women in a cross-sectional postal survey. The diagnostic accuracy of the scales was assessed against a 'reference standard', a structured diagnostic interview, to examine their case-detection ability in identifying pregnant women experiencing an anxiety disorder. Other psychometric properties that were evaluated for the SAAS included its internal consistency, convergent and discriminant validity and factor structure. The COSMIN (Consensus-Based Standards for selection of Health Measurement Instruments) criteria for evaluating whether a psychometric study meets the standards for good methodological quality (Terwee et al., 2007; Mokkink et al., 2010) were used to inform study design and ensure consistency with the criteria for excellent methodological quality in the evaluation of a range of psychometric properties.

2.3.2. Sample and data collection

A convenience sample of 174 women in their second and third trimester of pregnancy were recruited between February and July 2018. Women attending routine antenatal clinics in the Glasgow area (UK) and meeting the study inclusion criteria were initially approached and given study booklets, including information, questionnaire, consent form and pre-paid return envelope, by their midwives. Women were eligible to take part if they were: *a)* pregnant between 6 and 38 gestational weeks *b)* at least 18 years of age *c)* receiving routine prenatal care *d)* with a level of English sufficient to understand and complete questionnaires in lay language *e)* able to provide written informed consent to take part in the study.

The questionnaire comprised the SAAS, GAD-2/7, Edinburgh Postnatal Depression Scale (EPDS: Cox et al., 1987) and information on women's age, gestation, score on a single pregnancy-related anxiety item, parity, history of obstetric complications, ethnicity, educational level and marital status. Ease of completion and acceptability of the SAAS were evaluated through two questions.

A sub-sample, selected to represent a range of GAD-7 scores, were invited to take part in a structured diagnostic interview (M.I.N.I.: Sheehan et al., 1998) to evaluate the diagnostic accuracy of the SAAS and GAD-2/7. All diagnostic interviews were telephone-administered and conducted by a trained researcher, independent to the study authors and blind to the scores of the scales. A target sample of 60 interviews was required to give a 10% confidence interval for the estimates of sensitivity and specificity (Buderer,1996). Seventy-one women were invited to take part in the diagnostic interview and 37 interviews were completed. The aim was to conduct all interviews within four weeks from completion of the scales. Where this was not possible, the SAAS and the GAD-7 were re-administered at the end of the diagnostic interview to allow a clinically meaningful comparison

between scale scores and the diagnostic interviews.

2.3.3. Measures

The GAD-2 consists of the initial two questions of the GAD-7, which was developed to assess specifically Generalised Anxiety Disorder (Kroenke et al., 2007) but has shown good screening accuracy for a range of anxiety disorders. Respondents are asked to indicate how frequently they "have been bothered" by each of the symptoms listed over the past two weeks, with all items scored on a 4-point Likert scale ranging from "0 = *not at all"* to "3 = *nearly every day*".

The EPDS is a widely used 10-item scale for the assessment of depression in the perinatal period (Cox et al., 1987; Murray and Cox, 1990), with very good sensitivity and specificity (Howard et al., 2018). The EPDS asks respondents about symptoms of depression experienced in the previous week, with four possible response options and a total score range of 0–30. It was used in this study exclusively to assess the discriminant validity of the SAAS and GAD-2/7.

The M.I.N.I. International Neuropsychiatric Interview (M.I.N.I.: Sheehan et al., 1998) is a brief, structured interview used to ascertain the presence of Axis I DSM-IV and ICD-10 psychiatric disorders. It has excellent validity and inter-rater reliability, and it has been validated in a range of populations (Sheehan et al., 1998). The anxiety modules of the MINI PLUS version 5.0, which included Panic Disorder, Agoraphobia, Generalised Anxiety Disorder, Social Anxiety Disorder, Obsessive-Compulsive Disorder, Posttraumatic Stress Disorder and Specific Phobia, were used in this study to determine diagnoses of an anxiety disorder. The MINI PLUS version 5.0, based on the DSM-IV, was chosen rather than a DSM-V based instrument as when the study was conducted in early 2018 the M.I.N.I version 7.0.2, based on DSM-V, had just been published. Consequently, it did not have as much validation data as its previous version which had been widely validated in a range of populations.

2.4. Statistical analysis

Descriptive statistics were used to summarise respondents' demographic and obstetric characteristics. Responses to the two questions enquiring about the ease of completion and acceptability of the SAAS (scored on a 1–10 scale) were assessed using mean scores and frequency distributions.

The primary aim of this psychometric validation study was to assess the diagnostic accuracy (i.e. criterion validity) of the SAAS and GAD-2/7 in identifying women experiencing an anxiety disorder, as determined by M.I.N.I diagnoses. A number of parameters of diagnostic accuracy, including the sensitivity (i.e. true positive rate), specificity (i.e. true negative rate), positive and negative predictive values for each of the three scales were calculated at a range of possible cut-off scores. Values closer to 100% for sensitivity and specificity indicate better discriminative accuracy. In the psychometric literature, 70% is often cited as a minimally acceptable value for both sensitivity and specificity, with values over 70% considered good, $\geq$ 80% very good and $\geq$ 90% excellent (Furr, 2011). 95% lower and upper confidence intervals for all the parameters of diagnostic accuracy are also reported. Analyses of the Area Under the ROC curve (AUROC), which provide a single index of overall diagnostic performance, were also conducted for the SAAS, GAD-2 and GAD-7. A value of 0.90 or above is considered excellent, with an AUROC between 0.80 and 0.90 indicating good discriminative accuracy (Pallant, 2013).

Cronbach's Alpha (Cronbach, 1951) was used to determine the internal consistency of the SAAS and GAD-7. Abell and colleagues (2009), indicate that in the case of clinical applications, internal consistency as measured by Cronbach's alpha (α) should be above 0.80, and other authors have suggested values closer to 0.90 (Hunsley and Mash, 2008). Item statistics for the 10-item SAAS were calculated using descriptive statistics to examine the spread and patterns of items' scores, floor and ceiling effects as revealed by excessive item skewedness, and to check the overall interrelatedness of items by inspecting the correlation matrix. The criteria previously discussed for the pilot study were applied (floor/ceiling effect $\geq$ 90%, item-total correlations $\geq$ 0.30, inter-item correlations $\geq$ 0.20 and $\leq$ 0.80).

Principal Factors Analysis was used to explore the factor structure of the SAAS. The factor structure of a scale is an important aspect of validity, as it provides evidence of whether a scale is unidimensional (i.e. measures a single factor or latent construct) or multidimensional. In the case of the SAAS, the proposed construct of antenatal anxiety was hypothesised to be unidimensional. Kaiser's criterion, also known as the eigenvalue rule (Brace et al., 2013), and the scree test (Cattell, 1966) were used to determine the number of factors in the SAAS.

Convergent and discriminant validity are two psychometric properties that are commonly used to assess the construct validity of a scale (Abell et al., 2009). The widely used recommendations proposed by Cohen (1988) to evaluate the strength of correlations were used as follows: small correlation (0.10 – 0.29), medium correlation (0.30 - 0.49) and large correlation ($\geq$ 0.50). We hypothesised a large correlation between SAAS and GAD-2/7 scores to support convergent validity, while a moderate correlation between the SAAS and the EPDS scores was hypothesised as evidence of discriminant validity.

2.5. Ethics approval

Ethical and management approval was obtained from the South East Scotland Research Ethics Committee 02, and from the Research & Development service of the NHS Greater Glasgow & Clyde (GG&C) Health Board.

3. Results

3.1. Sample characteristics

An estimated 320 study booklets were distributed by midwives recruiting for the study; 178 women returned completed questionnaires (estimated response rate = 56%). Four respondents did not complete the consent form (n = 3) or reply slip (n = 1) and were excluded from the analysis. All the results reported here refer to the remaining 174 women. Thirty-seven women were assessed with a structured diagnostic interview and eleven met the criteria for M.I.N.I diagnoses of anxiety disorders. Specifically, four met the criteria for GAD, two for Posttraumatic Stress Disorder, two for Obsessive-Compulsive Disorder, one for Panic Disorder and two for multiple anxiety disorders.

The socio-demographic characteristics of questionnaire respondents are shown in Table 1. The mean age of study participants was 31.1 years (range 19 – 43). Fifty-four per cent of women were in their first pregnancy (n = 94) while the remaining 46% (n = 80) were in their second or subsequent pregnancy. amongst women who had previously given birth, almost one third reported at least one experience of pregnancy or birth complication (n = 24). The mean gestation of respondents was 28.4 weeks (Table 2).

Two participants missed one item each from the SAAS (item 5) and the GAD-7 (item 6). As missing items were less than 20% of the total number of items in a scale in both cases, item values were replaced with the median score for that scale, as suggested by Wilson and MacLean (2011). Four participants did not answer the two questions on ease of completion and acceptability of the SAAS.

3.2. Diagnostic accuracy of the SAAS, GAD-2 and GAD-7

The sensitivity, specificity, positive and negative predictive values of the three scales (based on M.I.N.I diagnosis) at different cut-off points are presented in Table 3.

The SAAS, at a cut-off score of 8 or above, showed excellent sensitivity (91%) and very good specificity (85%) in this sample. A different cut-off score of 12 or above maximised specificity (92%) at the expenses

Table 1
Respondents' ethnic background, highest level of qualification and marital status.

Ethnicity	Frequency(n = 174)	Response distribution	Scotland's census 2011
White Scottish	142	81.6%	83.9%
White – Other British	7	4.0%	7.8%
White – Any other white ethnic group	9	5.2%	3.9%
Asian/Asian British	8	4.6%	2.6%
Black/Black British/ African/ Caribbean	2	1.1%	0.7%
Mixed/multiple ethnic group	3	1.7%	0.4%
Other ethnic group	3	1.7%	0.6%
Education	Frequency (n = 173)	Response distribution	Scotland's census 2011 *(% amongst all people 16 years or over)*
Level 1 and 2 (e.g. SVQ level 1 o 2, A level)	29	16.7%	37.3%
Level 3 - HNC/HND or equivalent (SVQ Level 4)	27	15.6%	9.7%
Level 4 - Degree, Professional qualification (Above SVQ Level)	110	63.6%	26.1%
Other qualification	6 *(all PhDs)*	3.5%	*Included in Level 4 and above*
No qualifications	1	0.6%	26.7%
Marital status	Frequency (n = 174)	Response distribution	Scotland's census 2011 *(% amongst all people 16 years or over)*
Married	115	66.1%	45.2%
Single	19	10.9%	35.4%
Cohabiting	36	20.7%	*Included in 'Single'*
Divorced	1	0.6%	8.1%
Other	3	1.7%	*Included in 'Single'*

Table 2
Mean, standard deviation (SD) and range for age and gestational week of study participants.

Sample characteristics	Mean(SD)	Range
Age	31.16 (4.46)	19 - 43
Gestation of pregnancy	28 (6)	15 - 40
Of which:	*Frequency (n = 174)*	*%*
15 – 21 weeks	31	18
22 – 28 weeks	55	32
29 - 35 weeks	52	30
36 or more weeks	36	21

of the true positive rate (i.e. sensitivity, 73%).

The GAD-7 showed very good sensitivity (82%) and excellent specificity (96%) at a cut-off of 7 or above. NICE perinatal mental health guidance does not specify a cut-off score for the GAD-7 (NICE, 2014),

Table 3
Sensitivity, Specificity, Positive and Negative Predictive Values for the SAAS, GAD-2 and GAD-7 at a range of cut-off points.

SAAS cut-offscore	**Sensitivity**	**Specificity**	**PPV**	**NPV**
≥ 7	91% (82% - 100%)	81% (68% - 94%)	67% (52% - 82%)	95% (88% - 100%)
≥ 8	**91%** (82% - 100%)	**85%** (73% - 96%)	**71%** (56% - 86%)	**96%** (89% - 100%)
≥ 9	82% (69% - 94%)	88% (77% - 98%)	75% (61% - 89%)	92% (83% - 100%)
≥ 10	73% (58% - 87%)	88% (77% - 98%)	73% (58% - 87%)	88% (77% - 98%)
≥ 11	73% (58% - 87%)	88% (77% - 98%)	73% (58% - 87%)	88% (77% - 98%)
≥ 12	**73%** (58% - 87%)	**92%** (83% - 100%)	**80%** (67% - 93%)	**89%** (79% - 99%)
GAD-2 cut-off score	**Sensitivity**	**Specificity**	**PPV**	**NPV**
≥ 2	**73%** (58% - 87%)	**96%** (89% - 100%)	**89%** (79% - 99%)	**89%** (79% - 99%)
≥ 3	27% (13% - 141%)	96% (89% - 100%)	75% (61% - 89%)	76% (62% - 90%)
≥ 4	18% (6% - 30%)	100% (Not computable)	100% (Not computable)	74% (62% - 90%)
GAD-7 cut-off score	**Sensitivity**	**Specificity**	**PPV**	**NPV**
≥ 7	**82%** (69% - 94%)	**96%** (89% - 100%)	**90%** (80% - 100%)	**92%** (83% - 100%)
≥ 8	73% (58% - 87%)	96% (89% - 100%)	89% (79% - 99%)	89% (79% - 99%)
≥ 9	64% (49% - 80%)	96% (89% - 100%)	87% (76% - 98%)	86% (75% - 97%)
≥ 10	54% (39% - 71%)	96% (89% - 100%)	86% (75% - 97%)	83% (71% - 95%)

Values in bold indicate candidates as optimal cut-off scores for the three scales. 95% confidence intervals are included in parentheses.

which may lead to the assumption that the cut-off recommended for the general population (≥ 8: NICE, 2011) should be used. In this sample, however, this cut-off considerably reduced the sensitivity of the measure (73%). A similar, but arguably more significant problem was found with regard to the GAD-2, which at the NICE-recommended cut-off score of 3 or above showed very poor sensitivity (27%). Its specificity was 96% at this cut-off score.

The three ROC curves for the SAAS, GAD-7 and GAD-2 are presented as supplementary material (S5). The AUROC for the SAAS was 0.94, well above the threshold of 0.90 indicating excellent discriminative ability (Bland, 2000), and the AUROC for the GAD-7 was 0.93, only marginally lower than the SAAS. The AUROC for the GAD-2 was 0.88 (95% CI: 0.75–1), a value appearing to confirm the poorer discriminative

accuracy of this ultra-brief scale in an antenatal sample compared to the GAD-7 and the SAAS.

3.3. Diagnostic accuracy of the SAAS and GAD-2/7 in identifying women with an anxiety disorder and/or pregnancy-related anxiety symptoms

The study questionnaire included a single question to assess women's general level of pregnancy-related anxiety symptoms. Specifically, "*From 1 to 10, how do you feel about your pregnancy and about giving birth?*" with anchor points being "*1= completely calm*" and "*10= extremely anxious*". A score of 7 or above was considered indicative of probable PrA. The mean score was 4.6 (SD 1.91). Scores to this single, PrA question were used for secondary statistical analyses to establish the screening accuracy of the SAAS, GAD-2 and GAD-7 in identifying women experiencing an anxiety disorder and/or elevated levels of PrA. For this specific purpose, women were considered as 'positive' cases if they met diagnostic criteria for an anxiety disorder and/or if they scored ≥ 7 to the single question on PrA. These analyses indicated that the SAAS showed a superior screening accuracy when both anxiety disorders and PrA symptoms were considered (76% sensitivity, 90% specificity) at a cut-off of ≥ 7. The screening performance of the GAD-2 at the NICE-recommended cut-off score of 3 or above was poor with regard to its true positive rate (sensitivity: 18%), with only a moderate performance at the optimal cut-off score of 2 or above (47%). Its specificity was excellent at this cut-off point (95%). Similarly, the GAD-7 at its optimal cut-off scores of ≥ 7 exhibited only moderate sensitivity and excellent specificity (47% and 95% respectively). All values of sensitivity, specificity, PPV and NPV in these secondary analyses are reported as Supplementary material (S6).

3.4. Internal consistency and item statistics

For the SAAS, Cronbach's Alpha was found to be $\alpha = 0.88$. This is an excellent value, which closely approximates the value for clinical applications of a scale ($\alpha \sim 0.90$) often suggested in the literature (Kline, 2005). Internal consistency for the GAD-7 was $\alpha = 0.87$, comparable to the SAAS. This was not calculated for the GAD-2 which only includes two items. Item-total correlations for the SAAS were all above the pre-defined criterion of ≥ 0.30 (range 0.44 – 0.77). Inspection of the inter-item correlation matrix revealed a range of moderate to moderately high inter-item correlations, a desirable pattern for items in a scale (Streiner et al., 2008; Abell et al., 2009), with correlations all above 0.20 and below 0.80 (range 0.24 – 0.65). No items, if deleted, improved the value of Cronbach's Alpha, suggesting a unique contribution of each of the item to the total score. Inspections of response distributions did not reveal any floor or ceiling effect in SAAS items.

3.5. Factor structure

The KMO measure of sample adequacy was excellent (0.902), far exceeding the limit for acceptable sample size of > 0.60 (Kline, 2005) for factor analysis. Principal Factors Analysis was thus conducted, and initially the magnitude of the eigenvalues was inspected. These are reported in Table 4.

Table 4
Eigenvalues from Principal Factors Analysis on SAAS scores.

Factors	EigenvaluesTotal	% of variance	Cumulative variance%
1	4.88	48.9%	
2	0.97	9.7%	58.6%
3	0.80	8%	66.6%
4	0.69	6.9%	73.5%
5	0.63	6.3%	80%
6	0.51	5.1%	85.1%

Only factors with eigenvalues above 0.50 are reported.

The eigenvalue rule recommends that only factors with eigenvalues above 1 are retained (Brace et al., 2013). A second eigenvalue was considerably close to the recommended value for retaining factors. Catell's scree test (1966) was thus also visually inspected to examine factors above the point of inflection, and only one factor was found well above the point of inflection. Consequently, based on the recommended combination of the eigenvalue rule and examination of the scree plot, a single factor was retained and a one-factor solution for the SAAS was proposed.

3.6. Convergent and discriminant validity

Correlation coefficients were calculated between total scores of the SAAS, GAD-7 and EPDS to evaluate convergent and discriminant validity. Spearman's correlations (rs) were used to calculate and report correlation coefficients. A significant, positive correlation was found between the SAAS and GAD-7 ($rs = 0.70$, $n = 174$; $p < 0.01$). The strength of the correlation indicated a large correlation between the two scales, as hypothesised. The correlation between the SAAS and the GAD-2 was only marginally lower ($rs = 0.68$, $n = 174$; $p < 0.01$), also indicating good convergent validity. The magnitude of the correlation between SAAS and EPDS scores was also examined. There was a significant, positive correlation between the two measures ($rs = 0.73$, $n = 173$; $p < 0.01$), which was slightly greater in strength than between the SAAS and the GAD-7. While a moderate to large correlation was hypothesised between the SAAS and the EPDS, this was expected to be lower than the correlation with the GAD-7, which measures a construct arguably more closely related to the target construct of the SAAS. The Spearman's correlation coefficient between GAD-7 and EPDS revealed a correlation coefficient considerably similar ($rs = 0.70$, $n = 173$; $p < 0.01$) to the correlation between SAAS and EPDS ($rs = 0.73$). Different hypotheses, which may contribute to explain these somewhat unexpected findings in relation to the discriminant validity of the SAAS and GAD-7, are briefly considered in the Discussion.

3.7. Ease of completion and acceptability of the SAAS

The ease of completion and acceptability of the SAAS were evaluated with two questions, namely *"How easy was the questionnaire to complete?"* and *"Would you find it acceptable to complete this questionnaire as part of routine antenatal care"*. Of 170 responses, 103 assigned the maximum score of 10 both for ease of completion and acceptability for use in routine antenatal care (61% of the sample). The mean scores for both questions (ease of completion: 8.93; acceptability: 9.48) indicated that the SAAS was considered very easy to complete and acceptable.

4. Discussion

Identifying women with significant levels of anxiety during pregnancy is the first, crucial step in delivering support (Marchesi et al., 2016). There is now clear evidence that anxiety scales developed for the general population cannot be reliably used to screen for problematic anxiety in pregnant women (Meades and Ayers, 2011; Evans et al., 2015). This has resulted in a substantial proportion of women experiencing antenatal anxiety going undetected and thus not being offered any support or treatment (Khan, 2015). Consequently, a number of authors in recent years have advocated the use of a brief scale for the universal screening of antenatal anxiety (Biaggi et al., 2016; Brunton et al., 2015; Rubertsson et al., 2014). This study contributes to fill this gap by developing, through a rigorous and systematic process of scale development and testing, a short and easy-to-complete questionnaire to screen for a range of problematic anxiety symptoms during pregnancy. While further testing will be required, the findings presented in this paper provide an early indication that the SAAS has desirable psychometric properties and is potentially appropriate and acceptable as a

screening tool for use in clinical and research settings.

The methods used have various strengths. The combination of different sources of evidence to inform construct definition and the subsequent generation and selection of items to measure the target construct followed best practice in scale development (DeVellis, 2012). Sources of evidence to inform the construction of the scale included a systematic review of the psychometric literature on anxiety scales used in pregnancy, interviews with women with experience of antenatal anxiety, health professionals with expertise in the area of perinatal mental health, and the intended population of respondents through pilot psychometric testing of candidate items.

A significant strength of this study is that it draws on the expertise both of clinicians and of women with lived experience of problematic anxiety in pregnancy, increasing the content and construct validity of the measure (Clark and Watson, 1995; Abell et al., 2009). The use of a structured diagnostic interview as part of the psychometric validation of the SAAS is a significant strength (Evans et al., 2015). To our knowledge, the SAAS is also the first scale to include both general anxiety and PrA symptoms and can thus potentially be used to identify pregnant women experiencing either an anxiety disorder or PrA. In sum, we argue that the methods and procedures used to develop the SAAS provide evidence of methodological rigour in the attempts to maximise the psychometric properties and the acceptability of the final version of the scale.

In relation to the psychometric properties of the scale, the screening accuracy of the SAAS in this study was very good in the identification of women experiencing an anxiety disorder (sensitivity: 91%, specificity: 85%). The optimal cut-off score for the SAAS which maximised sensitivity and specificity for the identification of anxiety disorders was found to be ≥ 8. The findings also indicated that a different, conservative cut-off of ≥ 12 may be alternatively used if the aim is to maximise the specificity of the scale (92% compared to 85% for ≥ 8) and thus the number of true negative cases. There are many instances in clinical settings in which it might be preferable to prioritise one of the two indexes (Streiner et al., 2008) over the other (i.e. either maximise the proportion of true positive cases identified or the proportion of true negative cases). Some have observed that in clinical settings such as maternity care, the additional resources associated with the management of women incorrectly identified as depressed or anxious (i.e. false positives) are not cost-effective (Paulden et al., 2009). A large number of false positives is, moreover, likely to generate unintentional worry in women. If this approach is favoured, a conservative cut-off of ≥ 12 should be chosen in order to maximise the specificity of the scale, and consequently reduce the proportion of false positives. Others, however, have suggested that a two-stage approach to universal screening for common perinatal mental health problems may be adopted if the aim is to identify as many women as possible that are clinically depressed or anxious (Austin and Kingston, 2016). In this case, the SAAS could be used at a cut-off of 8 or above, in order to ensure that a large proportion of women experiencing problematic anxiety symptoms are identified (sensitivity 91%). In a second stage, a positive score may trigger a conversation with their midwife (e.g. a woman may be asked if the symptoms reported in the scale are something she would like support around), or referral for further assessment, depending on the severity of the problem.

The particularly poor discriminative performance of the GAD-2 at the recommended cut-off score of ≥ 3 was concerning, considering its current use in UK maternity services and in several studies to screen for anxiety in perinatal women. We found that the optimal cut-off for the GAD-2 was a score of 2 or above, yielding good sensitivity (73%) and excellent specificity (96%). However, NICE explicitly recommends a cut-off score of 3 or above in perinatal populations (2014). The sensitivity at this cut-off score was significantly poorer in this study (27%), indicating that a substantial proportion of women with an anxiety disorder were missed at a cut off of ≥ 3 (i.e. had GAD-2 scores of 2 or below). Our findings are consistent with those of Nath and colleagues (2018) who tested the GAD-2 in a sample of over 500 pregnant women, all also assessed with a structured diagnostic interview. At the recommended cut-off of 3 or above, the sensitivity of the scale in identifying women experiencing any anxiety disorder was poor (sensitivity 26%, specificity 91%). When considered together, these findings would strongly appear to indicate that the GAD-2 does not show sufficient psychometric robustness for the assessment of clinically significant anxiety during pregnancy at the recommended cut-off score and may result in a high rate of incorrect classifications with many false negatives.

While the SAAS and the GAD-7 showed comparable sensitivity and specificity in identifying women experiencing an anxiety disorder at their optimal cut-off scores (≥ 8 for the SAAS, ≥ 7 for the GAD-7), the SAAS performed significantly better in the identification of women experiencing PrA symptoms (i.e. as assessed by a single PrA question) than both the GAD-7 and the GAD-2, which showed only poor or moderate sensitivity. Based on these preliminary findings, the SAAS might thus provide a superior screening performance than the GAD-2/7 in identifying pregnant women experiencing a range of problematic anxiety symptoms. These findings also suggest that the GAD-2/7 is not sufficiently accurate to screen for the whole spectrum of problematic anxiety symptoms that women can experience in the antenatal period.

The internal consistency of the SAAS, close to excellent ($\alpha = 0.88$), confirmed the reliability of the scale and suggested a unidimensional construct, a hypothesis that was further supported by the single-factor solution identified by the factor analysis. Evidence of the convergent validity of the SAAS was also supported by a large correlation coefficient with the GAD-7, as hypothesised a priori. The surprising finding of large correlation coefficients between the EPDS and the two anxiety measures (rs range: 0.70 - 0.73) appeared to question the discriminant validity of the SAAS. Considered in combination, however, these findings confirm that the EPDS is not a unidimensional measure of depression, as also indicated in various studies suggesting the existence of a three-item anxiety subscale within the 10-item EPDS (Jomeen and Martin, 2005; Matthey et al., 2013). The large correlation between the EPDS and both the SAAS and the GAD-7 would thus seem to suggest issues with the structural and construct validity of the EPDS (i.e. not a single factor assessing depression), as opposed to issues of construct validity of the SAAS. Alternatively, the large correlations between the EPDS and the anxiety measures may be indicative of cothymia (Tyrer, 2001), also considering that comorbidity of anxiety and depressive symptoms is not uncommon, both in the general population (WHO, 2017) and in perinatal women (Goodman et al., 2014; Staneva et al., 2015).

5. Limitations

The study also has several limitations. A potential limitation of the psychometric validation study was that the response rate was relatively low at 56%, with potential implications for the representativeness of the sample. The response rate, however, is consistent with other studies using a postal survey method (Sahlqvist et al., 2011). Additionally, the sample was found to be representative of the female UK population with regard to most socio-demographic characteristics considered. Compared with maternity statistics in the UK, with 42% of first births in 2017 (NHS Information Services Division [ISD], 2017), women at their first pregnancy were somewhat over-represented in this sample (54%). For ethnicity and marital status their proportions in the study sample (as shown in Table 1) were similar to national-level statistics, while women with a higher level of education were over-represented in the sample. Perhaps the most significant limitation of the psychometric validation of the SAAS and GAD-2/7 against the M.I.N.I was the relatively small sample size for the diagnostic interviews ($N = 37$). The relatively large confidence intervals reported in Table 3 for sensitivity, specificity, PPV, NPV and AUROC of the three scales indicate that some caution is needed when interpreting the parameters of screening accuracy reported in the study. Further testing of the scale in larger samples, with a sufficient number of women assessed with a 'gold standard' diagnostic interview, would thus be highly desirable.

To conclude, it is now clear that current assessment methods of antenatal anxiety, including the one currently recommended by NICE in the UK (2014), are not evidence-based and may thus lead to incorrect identification (Nath et al., 2018), potentially creating unintended worry and anxiety in pregnant women. Timely and effective screening procedures for antenatal anxiety, however, are crucial in order to identify women who would benefit from monitoring, and where appropriate, early intervention, with a high potential for prevention of negative outcomes for mothers and children. The research reported here has developed a screening scale for the assessment of a range of problematic anxiety symptoms during pregnancy that has shown preliminary evidence of very good screening accuracy and is considered easy to complete and highly acceptable to women. The cautionary notes regarding the limitations of the research detailed above need to be considered when interpreting these positive findings. Notwithstanding this important caveat, in an area typically under-resourced as perinatal mental health care has been for many years (Glover, 2014), a psychometrically robust, acceptable and easy-to-complete screening tool that may be used by midwives and other health professionals to identify pregnant women experiencing problematic anxiety symptoms could be a valuable addition, and allow for more efficient targeting of resources and care.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Contributors

All authors contributed to the conception and design of the study. AS coordinated the acquisition, analysis and interpretation of data. AS drafted the manuscript. HC, MM and ROC contributed to the analysis and interpretation of the findings, and critically revised the manuscript. All authors were involved in editing the article and approved the final version of the manuscript.

Acknowledgments

We would like to thank Maternal Mental Health Scotland (MMHS), a forum of health professionals with expertise in perinatal mental health and women with lived experience of perinatal mental illness. Health professionals from MMHS participated in the Delphi study, while a group of women (MMHS Change Agents) took part in the qualitative interviews and provided feedback on the wording and formatting of consecutive versions of the scale.

Role of the funding source

This work was supported by the Chief Scientist Office of the Scottish Government Health Directorates through a Doctoral Training Fellowship to A.S. (DTF/15/03). The funding source had no role in study design; in the collection, analysis and interpretation of data; in writing the manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jadr.2022.100333.

Appendix 1 – Stirling Antenatal Anxiety Scale (SAAS)

The questions below ask about how you have felt in the past two weeks. **Please complete each question by circling or marking ("✔") the number that best describes your experience in the past 14 days.** Please be sure to answer each question. If you make a mistake, simply cross it out and mark the correct answer.

In the past two weeks…

	Never	Rarely	Sometimes	Often	Always
1. My anxiety stopped me from doing things	0	1	2	3	4
2. I felt panicky for no good reason	0	1	2	3	4
3. I felt unable to cope	0	1	2	3	4
4. I worried that something may be wrong with my baby	0	1	2	3	4
5. Thoughts got stuck in my head	0	1	2	3	4
6. I avoided people	0	1	2	3	4
7. I could not control my anxiety	0	1	2	3	4
8. I have had negative thoughts about childbirth	0	1	2	3	4
9. I did not feel worthy of being a mother	0	1	2	3	4
10. My worries overwhelmed me	0	1	2	3	4

Reference list

Abell, N., Springer, D.W., Kamata, A., 2009. Developing and Validating Rapid Assessment Instruments. Oxford University Press, New York. https://doi.org/10.1093/acprof:oso/9780195333367.001.0001 https://doi.org/.

Austin, M.P., Kingston, D., et al., 2016. Psychosocial assessment and depression screening in the perinatal period: benefits, challenges and implementation. In: Sutter Dallay, A., Glangeaud-Freudenthal, N., Guedeney, A., et al. (Eds.), Joint Care of Parents and Infants in Perinatal Psychiatry. First Ed. Cham, Springer International Publishing, Switzerland, pp. 167–195. https://doi.org/10.1007/978-3-319-21557-0_11 https://doi.org/.

Austin, M.P., Mule, V., Hadzi-Pavlovic, D., Reilly, N., 2021. Screening for anxiety disorders in third trimester pregnancy: a comparison of four brief measures. Arch Womens Ment Health. https://doi.org/10.1007/s00737-021-01166-9.

Bauer, A., Parsonage, M., Knapp, M., Iemmi, V., Bayo, A., 2014. The Costs of Perinatal Mental Health Problems. Centre for Mental Health, London.

Biaggi, A., Conroy, S., Pawlby, S., Pariante, C.M., 2016. Identifying the women at risk of antenatal anxiety and depression: a systematic review. J. Affect. Disord. 191, 62–77. https://doi.org/10.1016/j.jad.2015.11.014.

Blackmore, E.R., Gustafsson, H., Gilchrist, M., Wyman, C., O'Connor, T.G., 2016. Pregnancy-related anxiety: evidence of distinct clinical significance from a prospective longitudinal study. J. Affect. Disord. 197, 251–258. https://doi.org/10.1016/j.jad.2016.03.008.

Blair, M.M., Glynn, L.M., Sandman, C.aA., Davis, E.P., 2011. Prenatal maternal anxiety and early childhood temperament. Stress 14, 644–651. https://doi.org/10.3109/10253890.2011.594121.

Brouwers, E.P.M., van Baar, A.L., Pop, V.J.M., 2001. Maternal anxiety during pregnancy and subsequent infant development. Infant Behav. Dev. 24, 95–106. https://doi.org/10.1016/S0163-6383(01)00062-5.

Brunton, R.J., Dryer, R., Saliba, A., Kohlhoff, J., 2015. Pregnancy anxiety: a systematic review of current scales. J. Affect. Disord. 176, 24–34. https://doi.org/10.1016/j.wombi.2018.05.004 https://doi.org/.

Brunton, R.J., Dryer, R., Saliba, A., Kohlhoff, J., 2019. The Initial Development of the Pregnancy-Related Anxiety Scale. Women and Birth 32, e118–e130. https://doi.org/10.1016/j.wombi.2018.05.004 https://doi.org/.

Buderer, N.M., 1996. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. Acad. Emerg. Med. 3, 895–900.

Capron, L.E., Glover, V., Pearson, R.M., Evans, J., O'Connor, T.G., Stein, A., Murphy, S. E., Ramchandani, P.G., 2015. Associations of maternal and paternal antenatal mood with offspring anxiety disorder at age 18 years. J. Affect. Disord. 20–26. https://doi.org/10.1016/j.jad.2015.08.012.

Cattell, R.B., 1966. The scree test for the number of factors. Multivariate Behav Res 1, 245–276.

Clark, L.A., Watson, D., 1995. Constructing validity: basic issues inobjective scale development. Psychol. Assess. 7, 309–319. https://doi.org/10.1037/1040-3590.7.3.309 https://doi.org/.

Cohen, J., 1988. Statistical Power Analysis For the Behavioral Sciences, 2nd ed. Hillsdale, NJ. https://doi.org/10.4324/9780203771587. https://doi.org/.

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 (06), 782–786. https://doi.org/10.1192/bjp.150.6.782 https://doi.org/.

Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. Psychometrika 16, 297–334. https://doi.org/10.1007/BF02310555 https://doi.org/.

Dennis, C.L., Falah-Hassani, K., Shiri, R., 2017. Prevalence of antenatal and postnatal anxiety: systematic review and meta-analysis. Br. J. Psychiatry 210, 315–323. https://doi.org/10.1192/bjp.bp.116.187179.

De Vaus, D., 2014. Surveys in Social Research, 6th ed. Routledge, London. https://doi.org/10.4324/9780203519196. https://doi.org/.

DeVellis, R.F., 2012. Scale development: Theory and Applications, 3rd ed. Sage, Thousand Oaks, CA.

Ding, X.X., Wu, Y.Le, Xu, S.J., Zhu, R.P., Jia, X.M., Zhang, S.F., Huang, K., Zhu, P., Hao, J. H., Tao, F.B., 2014. Maternal anxiety during pregnancy and adverse birth outcomes: a systematic review and meta-analysis of prospective cohort studies. J. Affect. Disord. 159, 103–110. https://doi.org/10.1016/j.jad.2014.02.027.

Dunkel Schetter, C., Tanner, L., 2012. Anxiety, depression and stress in pregnancy: implications for mothers, children, research, and practice. Curr. Opin. Psychiatry 25, 141–148. https://doi.org/10.1097/YCO.0b013e3283503680.

Evans, K., Spiby, H., Morrell, C.J., 2015. A psychometric systematic review of self-report instruments to identify anxiety in pregnancy. J. Adv. Nurs. 71, 1986–2001. https://doi.org/10.1111/jan.12649.

Fairbrother, N., Janssen, P., Antony, M.M., Tucker, E., Young, A.H., 2016. Perinatal anxiety disorder prevalence and incidence. J. Affect. Disord. 200, 148–155. https://doi.org/10.1016/j.jad.2015.12.082.

Furr, R.M., 2011. Scale Construction and Psychometrics For Social and Personality Psychology. Sage, Thousand Oaks, CA. https://doi.org/10.4135/9781446287866 https://doi.org/.

Glover, V., 2014. Maternal depression, anxiety and stress during pregnancy and child outcome; What needs to be done. Best Pract. Res. Clin. Obstet. Gynaecol. 28, 25–35. https://doi.org/10.1016/j.bpobgyn.2013.08.017.

Goodman, J.H., Chenausky, K.L., Freeman, M.P., 2014. Anxiety Disorders During Pregnancy: a Systematic Review. J. Clin. Psychiatry 75, 1153–1184. https://doi.org/10.4088/JCP.14r09035.

Haines, H.M., Pallant, J.F., Fenwick, J., Gamble, J., Creedy, D.K., Toohill, J., Hildingsson, I., 2015. Identifying women who are afraid of giving birth: a comparison of the fear of birth scale with the WDEQ-A in a large Australian cohort. Sex. Reprod. Healthc. 6, 204–210. https://doi.org/10.1016/j.srhc.2015.05.002.

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

Hsu, C.C., Sandford, B.A., 2007. The Delphi technique: making sense of consensus. Pract. Assess. Res. Evaluation. 12, 1–8.

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. https://doi.org/10.1016/j.earlhumdev.2004.04.014.

Hunsley, J., Mash, E.J., 2008. A Guide to Assessment That Work. Oxford University Press, New York. https://doi.org/10.1093/med:psych/9780195310641.001.0001 https://doi.org/.

Ibanez, G., Bernard, J.Y., Rondet, C., Peyre, H., Forhan, A., Kaminski, M., Saurel-Cubizolles, M.J., 2015. Effects of antenatal maternal depression and anxiety on children's early cognitive development: a prospective Cohort study. PLoS ONE 10 (8), e0135849. https://doi.org/10.1371/journal.pone.0135849, 10.

Jomeen, J., Martin, C.R., 2005. Confirmation of an occluded anxiety component within the Edinburgh Postnatal Depression Scale (EPDS) during early pregnancy. J. Reprod. Infant Psychol. 23, 143–154. https://doi.org/10.1080/02646830500129297.

Khan, L., 2015. Falling Through the gaps: Perinatal mental Health and General practice. London. Centre for Mental Health.

Kline, T., 2005. Psychological Testing : a Practical Approach to Design and Evaluation. Sage, Thousand Oaks, CA.

Kroenke, K., Spitzer, R.L., Williams, J.B., et al., 2007. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 146, 317–325.

Lee, A.M., Lam, S.K., Sze Mun Lau, S.M., Chong, C.S.Y., Chui, H.W., Fong, D.Y.T., 2007. Prevalence, course, and risk factors for antenatal anxiety and depression. Obstet. Gynecol. 110, 1102–1112. https://doi.org/10.1097/01.AOG.0000287065.59491.70.

Leis, J.A., Heron, J., Stuart, E.A., Mendelson, T., 2014. Associations Between Maternal Mental Health and Child Emotional and Behavioral Problems: does Prenatal Mental Health Matter? J. Abnorm. Child Psychol. 42, 161–171. https://doi.org/10.1007/s10802-013-9766-4.

Marchesi, C., Ossola, P., Amerio, A., Daniel, B.D., Tonna, M., De Panfilis, C., 2016. Clinical management of perinatal anxiety disorders: a systematic review. J. Affect. Disord. 190, 543–550. https://doi.org/10.1016/j.jad.2015.11.004.

Matthey, S., Fisher, J., Rowe, H., 2013. Using the Edinburgh postnatal depression scale to screen for anxiety disorders: conceptual and methodological considerations. J. Affect. Disord. 146, 224–230. https://doi.org/10.1016/j.jad.2012.09.009.

Meades, R., Ayers, S., 2011. Anxiety measures validated in perinatal populations: a systematic review. J. Affect. Disord. 133, 1–15. https://doi.org/10.1016/j.jad.2010.10.009.

Mokkink, L.B., Terwee, C.B., Knol, D.L., Stratford, P.W., Alonso, J., Patrick, D.L., et al., 2010. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol 10 (22), 1–8. https://doi.org/10.1186/1471-2288-10-22 https://doi.org/.

Murray, D., Cox, J.L., 1990. Screening for depression during pregnancy with the Edinburgh Depression Scale (EDDS). J. Reprod. Infant Psychol. 8, 99–107. https://doi.org/10.1080/02646839008403615 https://doi.org/.

Nath, S., Ryan, E.G., Trevillion, K., Bick, D., Demilew, J., Milgrom, J., Howard, L.M., et al., 2018. Prevalence and identification of anxiety disorders in pregnancy: the diagnostic accuracy of the two-item Generalised Anxiety Disorder scale (GAD-2). BMJ Open 8 (9), e023766. https://doi.org/10.1136/bmjopen-2018-023766 https://doi.org/.

National Institute for Health and Care Excellence, 2011. Common Mental Health Disorders: Identification and Pathways to Care (CG123). National Institute for Health and Clinical Excellence, London.

National Institute for Health and Care Excellence, 2014. Antenatal and Postnatal Mental health: Clinical Management and Service Guidance (CG192). National Institute for Health and Clinical Excellence, London.

Netemeyer, R.G., Bearden, W.O., Sughash, S., 2003. Scaling procedures: Issues and Applications. Sage, Thousand Oaks, CA.

NHS Information Services Division [ISD], 2017. https://www.isdscotland.org/Health-Topics/Maternity-and-Births/Publications/data-tables.asp (accessed 12 September 2020).

O'Donnell, K.J., Glover, V., Barker, E.D., O'Connor, T.G., 2014. The persisting effect of maternal mood in pregnancy on childhood psychopathology. Dev. Psychopathol. 26, 393–403. https://doi.org/10.1017/S0954579414000029.

Orr, S.T., Blazer, D.G., James, S.A., Reiter, J.P., 2007. Depressive symptoms and indicators of maternal health status during pregnancy. J. Women's Health. 16 (4), 535–542. https://doi.org/10.1089/jwh.2006.0116 https://doi.org/.

Pallant, J., 2013. The SPSS Survival Guide, fitth ed. McGraw- Hill Education, Berkshire, England.

Paulden, M., Palmer, S., Hewitt, C., Gilbody, S., 2009. Screening for postnatal depression in primary care: cost effectiveness analysis. BMJ 339, b5203. https://doi.org/10.1136/bmj.b5203 https://doi.org/.

Pilkonis, P.A., Choi, S.W., Reise, S.P., Stover, A.M., Riley, W.T., Cella, D., 2011. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMISTM): depression, anxiety, and anger. Assessment 18, 263–283.

Rubertsson, C., Hellström, J., Cross, M., Sydsjö, G., 2014. Anxiety in early pregnancy: prevalence and contributing factors. Arch. Womens. Ment. Health 17, 221–228. https://doi.org/10.1007/s00737-013-0409-0.

Sahlqvist, S., Song, Y., Bull, F., et al., 2011. Effect of questionnaire length, personalisation and reminder type on response rate to a complex postal survey: randomised control trial. BMC Med Res Methodol 11, 1–8. https://doi.org/10.1186/1471-2288-11-62 https://doi.org/.

Sheehan, D.V., Lecrubier, Y., Harnett-Sheehan, K., Amorim, P., Janavs, J., Weiller, E., et al., 1998. The Mini International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview. J. Clin. Psychiatry. 59, 22–33.

Silverman, D., 2001. Interpreting Qualitative data: Methods for Analysing Talk, Text and Interaction, 2nd ed. Sage, London.

Simms, L.J., 2008. Classical and Modern Methods of Psychological Scale Construction. Soc. Personal. Psychol. Compass 2 (1), 414–433. https://doi.org/10.1111/j.1751-9004.2007.00044.x.

Sinesi, A., Maxwell, M., O'Carroll, R., Cheyne, H., 2019. Anxiety scales used in pregnancy: systematic review. BJPsych Open 5 (e5). https://doi.org/10.1192/bjo.2018.75.

Somerville, S., Dedman, K., Hagan, R., Oxnam, E., Wettinger, M., Byrne, S., Coo, S., Doherty, D., Page, A.C., 2014. The Perinatal Anxiety Screening Scale: development and preliminary validation. Arch Womens Ment Health 17, 443–454. https://doi.org/10.1007/s00737-014-0425-8 https://doi.org/.

Staneva, A.A., Hons, B.A., Psychology, M., Bogossian, F., 2015. The experience of psychological distress, depression, and anxiety during pregnancy : a meta-synthesis of qualitative research. Midwifery 31, 563–573. https://doi.org/10.1016/j.midw.2015.03.015 https://doi.org/.

Streiner, D.L., Norman, G.R., Cairney, J., 2008. Health Measurement scales: A practical Guide to Their Development and Use, 4th ed. Oxford University Press, Oxford. https://doi.org/10.1097/00004356-199112000-00017. https://doi.org/.

Terwee, C.B., Bot, S.D., de Boer, M.R., van der Windt, D.A., Knol, D.L., Dekker, J., et al., 2007. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 60 (1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012 https://doi.org/.

Tyrer, P., 2001. The case for cothymia: mixed anxiety and depression as a single diagnosis. Br J Psychiatry 179, 191–193. https://doi.org/10.1192/bjp.179.3.191 https://doi.org/.

Wijma, K., Wijma, B., 2017. A woman afraid to deliver: how to manage childbirth anxiety. In Bio-Psycho-Social Obstetrics and Gynecology 3–31.

Wilson, S., MacLean, R., 2011. Research Methods and Data Analysis For Psychology. McGraw Hill, Maidenhead, UK.

World Health Organization [WHO], 2017. Depression and Other Common Mental disorders: Global Health Estimates. World Health Organization, Geneva.