THE DEVELOPMENT AND INITIAL

VALIDATION OF A SCREENING SCALE

FOR ANTENATAL ANXIETY

Andrea Sinesi

University of Stirling

Thesis submitted for the degree of Doctor of Philosophy

September 2019

Acknowledgements

Abstract

Background: Anxiety during pregnancy is a strong predictor of postnatal depression and can negatively impact on a range of child developmental outcomes. Recent reviews highlight the lack of anxiety measures with robust psychometric properties for screening use in pregnancy.

Aim: This research aimed to develop a brief self-report scale specifically constructed to identify problematic anxiety symptoms in pregnant women, and conduct preliminary psychometric testing of the scale.

Method: The development and psychometric validation of the SAAS (Stirling Antenatal Anxiety Scale) was informed by five studies. A systematic review of the psychometric literature and interviews with women with experience of problematic anxiety symptoms in pregnancy both contributed to the generation of an initial item pool for the assessment of the target construct. This was subsequently refined and reduced, using a Delphi technique involving key informants (i.e. expert opinion and target population). The screening accuracy of the final, 10-item version of the scale was subsequently tested against a diagnostic interview, and compared to the NICE-recommended screening scales for antenatal anxiety (GAD-2/7). The internal consistency, factor structure and construct validity of the SAAS were also assessed.

Results: 174 women completed the SAAS, GAD-2/7 and Edinburgh Postnatal Depression Scale (EPDS). The SAAS was found to have excellent sensitivity (91%) and very good specificity (85%) at its optimal cut-off score of ≥ 8. It also showed a superior screening performance when compared to both the GAD-2 and the GAD-7 at their NICE-recommended cut-off scores. Its internal consistency was close to excellent ($\alpha = 0.88$), and the scale exhibited a single-factor structure. The SAAS was also considered highly acceptable to pregnant women (mean score = 9.48; range 1-10).

Conclusion: The SAAS shows promise as a brief, acceptable and effective screening tool for antenatal anxiety, which may improve identification and aid appropriate targeting of resources and care.

Word count: 300

Publications

This work has resulted in one research paper at the time of submission of the thesis. Chapter 4 is a partial reproduction of this paper, which was published in January 2019. The paper is also included as an appendix to this thesis (Appendix 18).

Sinesi, A., Maxwell, M., O'Carroll, R., & Cheyne, H. (2019). Anxiety scales used in pregnancy: Systematic review. *British Journal of Psychiatry Open*, 5(1), 1–13. https://doi.org/10.1192/bjo.2018.751

Table of contents

CHAPTER 5: WOMEN'S PERSPECTIVE: QUALITATIVE INTERVIEWS ON THE EXPERIENCE OF ANXIETY SYMPTOMS IN PREGNANCY 99

CHAPTER 6: CONSTRUCT DEFINITION, GENERATION OF AN ITEM POOL AND PRELIMINARY ITEM REDUCTION 131

List of tables

List of figures

Chapter 1   Introduction and overview of the thesis

1.1 Introduction

While the physical care of women during pregnancy and in the months after giving birth has substantially improved over the course of the last century, it is only in recent decades that their emotional care has received some attention in clinical practice and research (Glover, 2014). The perinatal period, which includes pregnancy and the first postnatal year (Milgrom & Gemmill, 2015), is a major time of transition for women, and is often characterised by substantial physiological, psychological and social changes (Grant, McMahon, & Austin, 2008). It is thus perhaps not surprising that a considerable proportion of women experience increased psychological vulnerability and morbidity at this time ((Heron, O'Connor, Evans, Golding, & Glover, 2004; Goodman, Watson, & Stubbs, 2016). Among the mental health difficulties that women can experience during pregnancy and in the first year after giving birth, a growing body of research has indicated that women experiencing elevated levels of anxiety during pregnancy are at increased risk of developing postnatal depression (Lee et al., 2007; Verreault et al., 2014). Maternal antenatal anxiety has also been found to be associated with a number of adverse obstetric and birth outcomes, including low birth weight and premature birth (Ding et al., 2014). In addition, evidence now exists suggesting that clinically significant anxiety during pregnancy increases the risk of a range of negative child developmental outcomes (Talge et al., 2007; O'Donnell, Glover, Barker, & O'Connor, 2014). In Chapter 2, which is specifically focused on antenatal anxiety, current evidence in relation to its potential negative impact on mother and child is reviewed and discussed in detail. In conclusion of this overview of antenatal anxiety, a rationale for the need to develop a new screening scale for the identification of pregnant women experiencing problematic anxiety symptoms is also presented.

This brief, introductory chapter aims to introduce the reader to the area of perinatal mental health research and clinical practice. The chapter initially provides an overview of perinatal mental health problems, their prevalence and common assessment methods. The issue of under-recognition of common mental health problems in routine maternity care is also briefly discussed. In 1.3, the study aims and research questions are presented, while the final section (1.4) provides a brief summary of the nine chapters included in the thesis.

1.2 Setting the context

1.2.1 An overview of perinatal mental health problems

Perinatal mental health problems and perinatal mental illness are general terms commonly used to refer to the range of mental health difficulties that women can experience during pregnancy and in the first postnatal year (Austin, 2004; Glover, 2014). These range from moderate symptoms of depression or anxiety to more severe conditions such as postpartum psychosis. Overall, perinatal mental health problems are estimated to affect between 10 and 20% of women (Gavin et al., 2005; Dennis, Falah-Hassani, & Shiri, 2017; Howard et al., 2018). Historically, the majority of research studies in the area of perinatal mental health have focused on the postnatal period (i.e. the first year after giving birth), and in particular on postnatal depression (Cox, Holden, & Sagovsky, 1987; Goodman et al., 2016). Consequently, much less attention has been devoted to other forms of psychological distress that women experience during the perinatal period (Tebbe, Terluin, & Koelewijn, 2013). The reasons for the historical emphasis on postnatal depression, both in research and clinical settings, are likely to be varied. Some authors have suggested that the lack of attention towards the antenatal period (i.e. pregnancy) may have resulted from the incorrect assumption that women are somewhat hormonally protected from poor mental health at this time (Lee et al., 2007; Biaggi, Conroy, Pawlby, & Pariante, 2016). Others have argued that during the antenatal period the clinical focus is predominantly on the physical health of the woman and the fetus, and psychological complaints may be dismissed or attributed to the physiological changes occurring during gestation (Talge et al., 2007). As a result, clinical symptoms of anxiety and depression in the antenatal period were, until recent years, neglected by the research literature on perinatal mental health (Martin, 2012).

However, a growing body of literature has emerged over the course of the last two decades documenting the common occurrence of mental health problems other than postnatal depression, both in pregnancy and postnatally (Olde, van der Hart, Kleber, & van Son, 2006; Spinelli, 2009; Rubertsson, Hellström, Cross, & Sydsjö, 2014). A considerable body of evidence now exists indicating that during pregnancy women have an increased susceptibility to poor mental health compared to the general population (Grant et al., 2008; Dennis et al., 2017), and a number of studies and systematic reviews have shown that the prevalence of perinatal mental health problems may in fact be higher in pregnancy than during the postnatal period (Heron et al., 2004, Goodman et al., 2016). The majority of these

studies also appear to indicate that anxiety disorders and self-reported symptoms of anxiety have higher prevalence than any other mental health problem during the perinatal period (Ross & McLean, 2006; Lee et al., 2007; Rubertsson et al., 2014; Fairbrother, Janssen, Antony, Tucker, & Young, 2016), as detailed later in the chapter. As a result, a conceptual shift has occurred in the area of perinatal mental health research and clinical practice, from the traditional emphasis on postnatal depression to an increased attention to the whole spectrum of mental health problems that women can experience during pregnancy and in the first postnatal year (Austin, Tully, & Parker, 2007; Glover; 2014).

Identifying women who experience poor mental health at this time is the first, crucial step in order to provide them with effective and timely support and treatment (Colin, 2012; Nath et al., 2018). Different assessment methods currently used in research and clinical settings are discussed in the following section.

1.2.2 Assessment methods of perinatal mental health problems

In research, the 'gold standard' method of assessment of psychological morbidity is a structured diagnostic interview conducted by a mental health professional or a trained researcher (DeVellis, 2012; Tolin et al., 2018). These clinical interviews are commonly based on diagnostic guidelines such as the Diagnostic and Statistical Manual of Mental Disorders (DSM–5: American Psychiatric Association [APA], 2013) or the ICD-10 Classification of Mental and Behavioural Disorders (ICD-10: World Health Organization [WHO], 1992). However, this type of assessment is time-consuming, potentially expensive and therefore often impractical in studies conducted in maternity care settings. Consequently, a substantial proportion of studies have used self-report rating scales such as the Edinburgh Postnatal Depression Scale (EPDS: Cox et al., 1987) and the State-Trait Anxiety Inventory (STAI: Spielberger et al., 1983) for the assessment of common mental health problems in perinatal populations.

With regard to perinatal depression, when this is assessed with a clinical diagnostic interview the occurrence of depressive symptomatology is commonly determined by meeting diagnostic criteria for Major Depressive Disorder or a Depressive Episode (Scottish Intercollegiate Guidelines Network [SIGN], 2012; Gavin, Meltzer-Brody, Glover, & Gaynes, 2015). In cases when a self-report scale is used, the Edinburgh Postnatal Depression Scale is by far the most widely used and well-validated measure available (O'Connor,

Rossom, Henninger, Groom, & Burda, 2016), and it is also used to assess antenatal depression (Matthey, Fisher, & Rowe, 2013). Another ultra-brief screening tool which has recently been used in research settings (Howard et al., 2018) is the Patient Health Questionnaire-2 (PHQ-2), sometimes referred to as the Whooley questions (Kroenke, Spitzer, & Williams, 2003; Bosanquet et al., 2015).

In contrast to depression, which is a well-defined and somewhat unitary construct (Widiger & Clark, 2000), various anxiety disorders exist according to formal diagnostic criteria (WHO, 1992; APA, 2013). An overview of the various anxiety disorders is provided in Chapter 2. Here it is important to highlight that assessing and screening for clinically significant symptoms of anxiety during the perinatal period is more complicated than screening for perinatal depression, as observed by various authors (Meades & Ayers, 2011; Evans, Spiby, & Morrell, 2015). Perinatal anxiety has been assessed with a range of self-report rating scales, including measures of general state and trait anxiety, scales assessing specific anxiety disorders and other self-report measures which focus on pregnancy-specific anxiety and worries. Some investigators would appear to use the term anxiety and stress interchangeably in perinatal mental health research (Van Den Bergh, Mulder, Mennes, & Glover, 2005; Dunkel Schetter & Tanner, 2012), thus adding to the uncertainty about how perinatal anxiety should be conceptualised and measured (Grant et al., 2008). Moreover, the majority of scales used have no or very limited evidence of their psychometric properties in perinatal populations (Evans et al, 2015; Brunton, Dryer, Saliba, & Kohlhoff, 2015). These significant limitations are also explored in more detail in Chapter 4.

In the UK, the National Institute for Health and Care Excellence regularly publishes recommendations for the identification and management of antenatal and postnatal mental health problems (NICE, 2007; NICE, 2014). As part of general screening procedures, NICE recommends that all women should be asked at the first antenatal assessment visit (also known as the booking visit) about family history and their own past and present experience of severe mental illness. In relation to screening for more common mental health problems such as perinatal anxiety and depression, a significant change in the most recent version of the NICE guidance (2014) was the introduction, for the first time, of two screening questions to assess the presence of perinatal anxiety (GAD-2: Generalised Anxiety Disorder-2, Kroenke et al., 2007). The previous guidelines (NICE, 2007) only included two screening questions for depressive symptoms (PHQ-2: Kroenke et al., 2003). An overview of the

NICE-recommended screening questions that health professionals should consider asking at each antenatal appointment and early in the postnatal period is presented below in Table 1.

Table 1: Depression and anxiety screening questions in the perinatal period (NICE, 2014)

Target condition	Scale	NICE screening questions	Further assessment if positive screen
Perinatal depression	PHQ-2 (also known as Whooley question)	• *During the past month have you often been bothered by feeling down, depressed, or hopeless?* • *During the past month have you often been bothered by having little interest or pleasure in doing things?*	If positive answer to one or both questions, consider: -Using other validated measures such as the PHQ-9 or EPDS - Referral to GP/mental health professional, depending on severity of problems
Perinatal Anxiety	GAD-2	• *During the past month have you been feeling nervous, anxious, or on edge?* • *During the past month have you not been able to stop or control worrying?*	If score is 3 or more on the GAD-2 scale, consider: - Using the GAD-7 for further assessment - Referral to GP/mental health professional, depending on severity of problems

With the recommendation to use the GAD-2 as part of standard screening procedures, NICE thus appears to acknowledge the clinical importance of identifying and providing support to women experiencing problematic anxiety symptoms during the perinatal period. Critically, however, the GAD-2 has no evidence to recommend its use in perinatal populations, and in particular during the antenatal period. This key limitation is considered in detail in Chapter 2, as part of a wider discussion on issues in screening for problematic anxiety during pregnancy.

1.2.3 The prevalence of perinatal mental health problems

The bulk of the evidence on the prevalence of perinatal mental health problems is available for perinatal anxiety and depression, with other disorders such as eating disorder and postpartum psychosis receiving considerable less attention in the research literature (Howard et al., 2014). In relation to perinatal depression, a meta-analysis which only included studies using formal diagnostic criteria reported a period prevalence (i.e. the rate over a defined period of time) of 12.7% during pregnancy and 5.7% over the first two postnatal months (Gavin et al., 2005). The authors, however, specified that for the postpartum period only a small number of high-quality studies were available. A large American study ($n > 10,000$) using DSM-IV diagnoses (APA, 2000) reported that 9.1% of women during pregnancy and 10.2% postnatally met the diagnostic criteria for a Major Depressive Episode (Hoertel et al., 2015). More recently, a systematic review and meta-analysis of the global prevalence of perinatal depression was conducted by Woody and colleagues (Woody, Ferrari, Siskind, Whiteford, & Harris, 2017). The authors reported an overall pooled prevalence of 11.9% (95% CI: 11.4–12.5%) for perinatal depression, with prevalence being slightly higher postpartum compared to the antenatal period. Overall, the studies discussed above appear to indicate a prevalence of perinatal depression of 10-12%, with comparable rates in the antenatal and postnatal period. This is consistent with other recent, methodologically robust, studies (Howard et al., 2018) and reviews (Schmied et al., 2013). Higher prevalence estimates are usually found in low and middle income countries and when depression is assessed with a self-report scale (Fisher et al., 2011; Woody et al., 2017).

As noted earlier, anxiety in the perinatal period has become the focus of growing research and clinical attention over the last two decades (Brouwers, van Baar, & Pop, 2001; Lee et al., 2007; Goodman, Chenausky, & Freeman, 2014). In relation to anxiety during pregnancy, Chapter 2 presents numerous studies examining the prevalence of general antenatal anxiety and specific anxiety disorders, as well as the risk factors and potential detrimental effects of antenatal anxiety on mother and child. Here only one systematic review is discussed, which arguably provides the most comprehensive and up-to-date evidence available on the prevalence of antenatal and postnatal anxiety (Dennis et al., 2017). This thorough review and meta-analysis of the literature examined over 100 studies from 34 countries and included both research using anxiety scales and studies employing structured diagnostic interviews. The review indicated that between 15 and 23% of women experience problematic anxiety symptoms during pregnancy. The prevalence of women with significant self-reported

anxiety symptoms (i.e. scoring above a predefined cut-off score on an anxiety scale) was found to be 18.2% in the first trimester, increasing to 24.6% in the third trimester. The pooled prevalence across trimesters was 22.9%. For anxiety disorders based on formal diagnostic criteria (WHO, 1992; APA, 2013), the overall prevalence during pregnancy was 15.2% (95% CI: 9.0%-21.4%). In the postnatal period, the overall prevalence of self-reported anxiety symptoms over the first 6 months after giving birth was 15.0%. Anxiety disorders meeting diagnostic criteria over the same period were estimated to have a prevalence of 9.9%. This recent review clearly illustrates that problematic anxiety symptoms are common throughout the perinatal period, and in particular during pregnancy. As noted above, other psychological disorders that women can experience during the perinatal period include eating disorder (Meltzer-Brody et al., 2011), posttraumatic stress disorder (Ayers, Bond, Bertullies, & Wijma, 2016) and postpartum psychosis. Postpartum psychosis is a particularly severe condition that can affect women in the weeks following childbirth, with symptoms such as thought disorganisation, memory loss and delusional or suicidal ideation (Sit, Rothschild, & Wisner, 2006). It is relatively rare, estimated to occur in 1-2 cases per 1,000 births (Spinelli et al., 2009).

The prevalence of perinatal mental health problems discussed above refer to those found in research settings. However, in routine maternity care detection rates are significantly lower (Bauer, Parsonage, Knapp, Iemmi, & Adelaja, 2014), especially for the most common mental health problems such as perinatal anxiety and depression (NICE, 2014). This has been identified as a major issue in recent clinical guidelines (SIGN, 2012; NICE, 2014), and the under-recognition of perinatal mental health problems in the UK has been indicated as a clinical priority to be urgently addressed by the Royal College of General Practitioners (RCGP, 2017). The lack of identification of women experiencing poor mental health in the perinatal period thus currently remains a significant problem.

This brief overview of perinatal mental health problems, their prevalence and common assessment methods aimed to provide a brief introduction to to this area of research and clinical practice. As noted in 1.1, the reasons and rationale for focusing specifically on the development of a screening scale for antenatal anxiety are discussed extensively in Chapter 2. In the next page, the study aim and research objectives are presented.

1.3   Study aim and research questions

The primary aim of the programme of work presented in this thesis was to develop a brief and psychometrically robust self-report scale to screen for a range of problematic anxiety symptoms in pregnancy, and conduct a preliminary psychometric validation of the scale.

Since the early phases of this programme of work, the aim was also to develop a scale that was feasible to use both in research settings and in routine antenatal care. For this reason, the target was to produce a final version of the scale which contained less than 12 items (i.e. questions). NICE, in its most recent guidance on perinatal mental health clearly indicates that this is a prerequisite for self-report scales to be considered for use as a screening tool in maternity care in the UK (NICE, 2014).

The development and psychometric validation of the screening scale for antenatal anxiety presented in this thesis was guided by five main research questions, as listed below:

- *Research question 1:* What should a construct definition of antenatal anxiety include in order to cover the core domains of problematic anxiety symptoms in pregnancy?

- *Research question 2:* Which items are the most appropriate to operationalise the proposed construct of antenatal anxiety into a self-report rating scale?

- *Research question 3:* Which items are considered clear, relevant and acceptable by the target population and experts, and can thus be used to create a short and psychometrically robust self-report scale for the assessment of antenatal anxiety?

- Research question 4: What is the evidence in relation to the convergent and discriminant validity, internal consistency and factor structure of the final version of the scale?

- Research question 5: How does the new scale perform when compared to the measure currently recommended by NICE (GAD-2/7), and to expert assessment using a structured diagnostic interview; and what are the optimised cut-off points for maximising sensitivity and specificity of the scales?

1.4 Overview of the thesis

The organisation of the thesis is presented here. Chapter 1 has provided a brief overview of the prevalence and assessment methods of common perinatal mental health problems. The study aim and research questions were also presented.

Chapter 2, following an introduction to the construct of general anxiety and anxiety disorders, focuses entirely on antenatal anxiety. Evidence with regard to its prevalence at different stages of pregnancy and its role in increasing the risk for adverse maternal and child outcomes in the postnatal period is discussed. The second part of the chapter examines the current issues in screening for antenatal anxiety and makes the case for the need for early identification and support. The chapter concludes with a rationale for the development of a new screening scale for the assessment of problematic anxiety symptoms in pregnant women.

Chapter 3 begins with the presentation of a number of theoretical and methodological considerations related to scale development and psychometric testing. The chapter then describes the research methodology and the specific research methods that were used to develop and conduct preliminary psychometric testing of the scale, and concludes by examining the potential ethical issues arising from the research.

Chapters 4 and 5 presents the two studies that were conducted to inform the initial phase of scale development and generation of an initial item pool. A systematic review of the psychometric literature of anxiety scales used in pregnancy is reported in Chapter 4. The findings from this study were complemented by qualitative interviews with women with experience of problematic anxiety symptoms during pregnancy (Chapter 5).

In Chapter 6, based on the findings of the previous two chapters, a conceptual and an operational definition of the construct of antenatal anxiety are proposed. The pool of items generated to operationalise the proposed construct is subsequently presented, including a preliminary phase of item refinement through the contribution of key informants. Finally, a Delphi study is presented in which expert opinion was used to reduce the initial number of items from 59 to 30.

Chapter 7 presents the pilot psychometric testing of the preliminary, 30-item version of the scale using a cross-sectional survey. This survey, which aimed to reduce further the number

of items through a process of item analysis, resulted in a final, 10-item version of the scale named as the Stirling Antenatal Anxiety Scale (SAAS).

Chapter 8 presents the preliminary psychometric validation of the SAAS. A second cross-sectional survey conducted in a larger sample of pregnant women is described, in which a range of psychometric properties of the new scale were assessed. Its screening performance was evaluated against a structured clinical interview, and it was also compared to the GAD-2/7.

In the Discussion (Chapter 9), the study aim and research questions are re-examined in light of the findings. The strength and limitations of the research are critically discussed, and the potential implications for policy and clinical practice are examined. The thesis concludes by indicating possible directions for future research, and briefly highlighting the original contribution of this research to the field of perinatal mental health.

Chapter 2   Perinatal mental health problems: the case of antenatal anxiety

2.1 Introduction

As noted in the introductory chapter, Chapter 2 is almost entirely focused on problematic anxiety symptoms during pregnancy (i.e. antenatal anxiety). First, however, it was considered appropriate to define and briefly discuss the general construct of anxiety, and provide an overview of the most common anxiety disorders. This chapter thus initially presents a brief introduction to the construct of anxiety and anxiety disorders (2.2). In 2.3 several key studies documenting the prevalence of antenatal anxiety, and indicating specific groups of women which may be at increased risk, are reviewed. A specific type of anxiety that women can experience during the antenatal period (pregnancy-related anxiety) is also discussed. Subsequently, a number of studies which examined the potential detrimental effects of antenatal anxiety on mother and child are summarised (2.4), and the importance of early identification and support is highlighted (2.5). The chapter continues with a discussion on the current issues in screening for clinically significant anxiety symptoms in pregnant women (2.6), and concludes by providing a rationale for developing a self-report screening scale specifically constructed to identify women experiencing antenatal anxiety.

2.2 A brief overview of anxiety and anxiety disorders

Anxiety has been described as a universal human emotion (Simpson, Neria, Lewis-Fernandez, & Schneier, 2010). It is commonly characterised by a set of physiological, affective, behavioural, and cognitive responses to internal or external stimuli which are perceived as potentially dangerous or threatening (Barlow, 2002). Its evolutionary advantages are well documented in the literature, and it is now clear that anxiety has evolved primarily as a defence mechanism to alert us to react promptly to imminent or future dangers (Simpson et al., 2010). It is thus a common and normal emotion that serves important signal functions (Vanin, 2008). However, when feelings of anxiety become frequent, pervasive, or chronic regardless of the presence of an imminent or future threat, they can have a considerable negative impact on an individual's psychological wellbeing and daily functioning (APA, 2013). Anxiety symptoms can include excessive worry, heightened vigilance, a feeling of pervasive uneasiness, intense fear, physiological arousal, muscle tension and behavioural avoidance of stimuli perceived as potentially dangerous (Remes, Brayne, van der Linde, & Lafortune, 2016). In this thesis, when the general term "anxiety" is used, this will refer to the broad definition provided by the American Psychiatric Association (2013) indicating that anxiety is an emotion characterised by feelings of tension, worried thoughts, and physical changes like increased blood pressure. With regard to the diagnostic approach to anxiety, when anxiety symptoms cause significant distress or become chronic they are typically categorised as one of a range of anxiety disorders, according to the classifications provided by the DSM (APA, 2013) or the ICD (WHO, 1992). The most prevalent and well-known anxiety disorders are Generalised Anxiety Disorder (GAD), Social Anxiety Disorder, Panic Disorder, Agoraphobia and Specific Phobia (Remes et al., 2016). For reasons of brevity, an overview of the core symptoms associated with different anxiety disorders is provided in the following page in Table 2.

Table 2: Core symptoms of anxiety disorders in the DSM-V and ICD-10

Anxiety disorder	Core symptoms
Generalised anxiety disorder (GAD)	• Excessive worry and anxiety related to a range of situations/activities • Difficulty controlling the worry • Physical symptoms may include restlessness and muscle tension
Social anxiety *(Social phobia in ICD-10 classification)*	• Marked worry or fear about being the focus of attention/acting in a way that may be embarrassing in one or more social situations • Exposure to social situations invariably causes marked anxiety and is avoided or endured with significant distress
Panic disorder	• Unexpected and recurrent panic attacks, characterised by intense fear or discomfort (physical symptoms may include pounding heart, sweating, difficulty breathing). • Persistent worry around experiencing further panic attacks
Agoraphobia	• Marked and disproportionate fear in situations such as public transport and open spaces • Exposure to the feared situation provokes almost invariably a panic attack • Avoidance behaviours or anticipatory anxiety with significant impact on routine and social activities
Specific phobia	• Intense anxiety or fear related to a specific situation or object, which is disproportionate to the actual danger • Exposure to the feared stimuli provokes almost invariably a panic attack • The anxiety-provoking object or situation is avoided or endured with marked distress
Posttraumatic stress disorder *(Included in the 'Trauma and Stressor-Related Disorders' chapter in DSM-5)*	• Personal or indirect exposure to traumatic situation such as death, threatened death or severe injury • Trauma-related flashbacks, nightmares or physical reactivity following exposure to stimuli reminding the trauma • Negative affect, loss of interest in activities • Hypervigilance, marked irritability, difficulty sleeping
Obsessive-Compulsive Disorder *(Included in the 'Obsessive-Compulsive and Related Disorders' chapter in DSM-5)*	• <u>Obsessions</u> - Presence of recurrent and unwanted thoughts that cause marked anxiety or distress and that the individual tries to ignore or suppress with some specific thought or action AND / OR <u>Compulsions</u> – Presence of repetitive behaviours or mental acts, such as checking or ordering that are performed in response to an obsession or following particular rules.

By examining the core symptoms of the anxiety disorders summarised in Table 2, it is evident that many common features are shared by the majority of the anxiety disorders. While the distinction of anxiety disorders in a number of different diagnostic categories is useful for clinical purposes, in recent years a trans-diagnostic approach to the classification and treatment of anxiety disorders is emerging in the literature (Norton & Paulus, 2017). Notably, Obsessive-Compulsive Disorder (OCD) and Posttraumatic Stress Disorder (PTSD), which were classified as anxiety disorders in the DSM-IV-TR (APA, 2000), are no longer considered as such in the DSM-V. However, the clinical utility of this reclassification has been questioned as OCD and PTSD share many distinctive features of other anxiety disorders (Emmelkamp & Ehring, 2014). In this thesis OCD and PTSD will be considered under the umbrella of the anxiety disorders, consistently with the ICD-10 classification which includes all the anxiety disorders under the same category of *Neurotic, stress-related and somatoform disorders*, and in consideration of the clinical importance of these conditions in perinatal women (Chaudron & Nirodi, 2010; Ayers, Meades, & Matthey (2015).

Global epidemiological evidence in the general population indicates that anxiety disorders are the most common class of psychiatric disorders worldwide (Kessler et al., 2005; Remes et al., 2016). One of the most rigorous systematic reviews of prevalence studies of anxiety disorders to date indicated a lifetime prevalence of 28.8% worldwide (Kessler et al., 2005). In an epidemiological review of reviews (Remes et al., 2016), women were found to have an almost two-fold increased risk of experiencing an anxiety disorder compared to men, with a female-to-male ratio of 1.9: 1. This observation is supported by other recent data indicating a global time-point prevalence of 4.6% for women compared to 2.6% for men (WHO, 2017). Moreover, anxiety disorders are particularly prevalent under 35 years of age (Baxter, Scott, Vos, & Whiteford, 2013), with onset of symptoms typically occurring in early adulthood (Kessler et al., 2012). In relation to specific anxiety disorders, the highest prevalence is found for Generalised Anxiety Disorder and Specific Phobia, both with a lifetime prevalence of approximately 6% (Remes et al., 2016). It is also important to note that comorbidity of anxiety and depressive disorders is not uncommon, both in the general population (Haug, Mykletun, & Dahl, 2004; WHO, 2017) and in perinatal women (Goodman et al., 2014; Staneva, Bogossian, Pritchard, & Wittkowski, 2015). This is consistent with the widely influential tripartite model of anxiety and depression proposed by Clark and Watson (1991). This model postulates that, while some of the core symptoms of anxiety and depressive

symptomatology are clearly distinct (i.e. physiological arousal in anxiety and absence of positive affect in depression), anxiety and depressive disorders also share a common component of general distress that the authors named negative affect (Clark & Watson, 1991). Following this brief overview of anxiety and anxiety disorders, the rest of the chapter is specifically focused on discussing anxiety during pregnancy, which will be mostly referred to as antenatal anxiety.

2.3 Antenatal anxiety

In the following pages, the term antenatal anxiety will be used to refer to problematic anxiety symptoms experienced during pregnancy. This will include both women meeting diagnostic criteria for one or more anxiety disorders (WHO, 1992; APA, 2013) during the antenatal period, as well as pregnant women scoring above a predefined cut-off on a self-report measure of anxiety. A more detailed definition of antenatal anxiety will be provided later in this chapter, based on recent evidence that anxiety disorders are only one of the types of problematic anxiety symptoms that women can experience during pregnancy. A formal definition of the construct of antenatal anxiety is subsequently proposed in Chapter 6.

2.3.1 Antenatal anxiety: common but under-recognised

As noted in the previous chapter, antenatal anxiety appears to be more common than antenatal depression (Grant et al., 2008). The systematic review on the global prevalence of antenatal anxiety reported earlier indicated prevalence ranging from 15 to 23%, for diagnosed anxiety disorders and self-reported anxiety symptoms respectively (Dennis et al., 2017). Although accurate prevalence estimates for antenatal anxiety remain challenging because of a number of factors, including the heterogeneity of assessment methods used and timing of assessment (i.e. different trimesters), the largest longitudinal study (n > 8000) conducted in the UK on the prevalence of self-reported symptoms of anxiety and depression in the perinatal period found that approximately 14% of women experience elevated levels of anxiety during pregnancy (Heron et al., 2004), with similar prevalence at 18 and 32 gestational weeks. This study also provided strong evidence for the relative stability of anxiety symptoms across the perinatal period, as illustrated by the fact that two-thirds of women reporting significant anxiety symptoms in the postpartum period had already experienced anxiety during pregnancy. A number of other studies have reported similar prevalence, indicating that problematic anxiety symptoms affect approximately 15% of women, both in early pregnancy (Rubertsson et al., 2014) and in later stages (Lee et al., 2007; Grant et al., 2008; Goodman et al., 2014). The well-documented high prevalence of antenatal anxiety is of particular concern when its significant under-recognition in routine maternity care is considered. Both NICE (2014) and SIGN (2012) guidance on perinatal mental health have observed that common perinatal mental health problems, and in particular

anxiety disorders, often go unrecognised and thus untreated throughout pregnancy and the postnatal period. This is certainly due to a combination of factors, but it has been suggested that the traditional emphasis on depression in perinatal mental health care and a hierarchical diagnostic custom in clinical practice may lead to prioritise depressive over anxiety symptomatology (Matthey et al., 2013a; Dennis et al., 2017). It is also important to note that there is evidence indicating that a proportion of women experiencing mental health difficulties during pregnancy are likely to present with comorbid disorders (Grant et al., 2008; Fairbrother et al., 2016).

In relation to specific groups of women at higher risk of antenatal anxiety, the research literature has identified a number of obstetric, psychosocial and contextual factors that increase the risk of experiencing antenatal anxiety. The most commonly reported include a previous history of mental health problems (Rubertsson et al., 2014; Marchesi et al., 2014; Bayrampour, McDonald, Fung, & Tough, 2015), low educational level (Glazier, Elgar, Goel, & Holzapfel, 2004; Bodecs et al., 2013), lack of perceived social support (Lee et al., 2007; Grant et al., 2008; Martini et al., 2015), a history of domestic abuse (Fisher et al., 2011) and a past experience of pregnancy loss or obstetric complications (Armstrong, 2004M; Waqas et al., 2015). There is also evidence, although more limited, for other variables to be predictive of antenatal anxiety, including low income (Prady et al., 2013) and being a single mother (Faisal-Cury & Rossi Menezes, 2007). Contradictory findings have been reported in relation to other factors such as age (Biaggi et al., 2016), with some studies indicating that younger women are at higher risk (Lee et al., 2007; Rubertsson et al., 2014) and others reporting that women over 35 years of age are more likely to experience antenatal anxiety (Nasreen, Kabir, Forsell, & Edhborg, 2011; Fisher et al., 2011). The observation that women who have experienced a miscarriage and those with a present or past history of obstetric complications are at higher risk of clinically significant anxiety during pregnancy is significant. Considering that approximately 10 to 15% of pregnant women have at least one experience of previous miscarriage (Royal College of Obstetricians and Gynaecologists [RCOG], 2017), these women could be specifically targeted with regular monitoring of their emotional wellbeing throughout pregnancy. Pregnant women can obviously experience any of the anxiety disorders that are also found in the general population. However, anxiety disorders are not the only type of problematic anxiety that women can experience during pregnancy, as illustrated in the following section.

2.3.2 The case of pregnancy-related anxiety

During the antenatal period, women can experience significant worries and fears that are specific to pregnancy (Blackmore, Gustafsson, Gilchrist, Wyman, & O'Connor, 2016). The occurrence of pregnancy-related anxiety has been proposed as a specific and distinct syndrome (Huizink, Mulder, Robles De Medina, Visser, & Buitelaar, 2004) and a growing body of empirical evidence, briefly discussed below, would appear to support its clinical distinction from other forms of anxiety that women can experience during pregnancy (Brunton et al., 2015; Blackmore et al., 2016; Witteveen et al., 2016). Pregnancy-related anxiety (hereafter often referred to as PrA) can be defined as a particular anxiety response in which symptoms of anxiety are specifically focused on pregnancy and childbirth, and may include persistent worries regarding the health of the woman and fetus, fear around labour and delivery and concerns around physical appearance and future parenting (Orr, Blazer, James, & Reiter, 2007; Dunkel Schetter & Tanner, 2012). It is reasonable to expect that most expectant women will experience a level of worry and anxiety regarding these aspects of pregnancy and childbirth, which are also thought to serve important functions in protecting the fetus from potential harm and preparing for parenthood (Haines et al., 2015). It is thus particularly important not to consider these common and normal concerns as indicators of pathological or problematic anxiety. However, similarly to anxiety disorders, 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 & Wijma, 2017). While this anxiety type is not covered by standard diagnostic classifications, it is of clinical significance as it can lead to negative outcomes for mother and child, as discussed later in this section.

Initial psychometric evidence of PrA as a psychological construct that can be distinguished from general antenatal anxiety was provided by Huizink and colleagues (2004) who developed a self-report scale, the Pregnancy-Related Anxiety Questionnaire – Revised (PRAQ-R), based on previous attempts at measuring PrA (Levin, 1991). The scale was tested on nulliparous women (i.e. in their first pregnancy), who completed both the PRAQ-R and a measure of general anxiety (STAI: Spielberger et al., 1983). The authors found that the STAI and factors of the PRAQ-R only shared low to moderate variance (8-27%), and interpreted this as an indication that PrA should be considered as a distinct syndrome, which can be reliably distinguished from general anxiety and anxiety disorders during pregnancy.

Comparable findings, indicating only a moderate overlap between PrA and general antenatal anxiety using the PRAQ-R, were replicated in some other studies, both in nulliparous and multiparous women (Arch, 2013; Westerneng, de Cock, Spelten, Honig, & Hutton, 2015). Several other scales to measure PrA have been developed, including the Pregnancy-Specific Anxiety Scale (PSAS: Roesch et al., 2004) and the Wijma Delivery Expectancy/Experience Questionnaire (W-DEQ: Wijma, Wijma, & Zar 1998), the latter specifically focused on fear of childbirth as a domain of PrA. Other authors have, however, argued that while PrA describes worries specifically related to pregnancy and childbirth, the nature of the symptoms (e.g. excessive worrying) closely resembles anxiety symptoms as experienced in the general population, and consequently that pregnancy-related anxiety should not be conceptualised as a distinct syndrome (Bar-Shai, Gott, Kreinin, & Marmor, 2015). In sum, although the literature on PrA is still relatively limited, several authors have proposed that the psychological construct of PrA is at least partially distinct from general anxiety during pregnancy (Blackmore et al., 2016). An important implication of this is that women experiencing problematic PrA symptoms may not meet diagnostic ICD or DSM criteria for an anxiety disorder or score above cut-off point on a measure of general anxiety (Meades & Ayers, 2011; Brunton et al., 2015), and consequently go unrecognised in routine antenatal care.

Fear of childbirth

Within the research literature on pregnancy-related anxiety, fear of childbirth has been recognised as an important dimension of PrA (Heimstad, Dahloe, Laache, Skogvoll, & Schei, 2006; Rubertsson et al., 2014). Fear of childbirth (FoC), sometimes also referred to as tokophobia, can be defined as intense anxiety or fear related to the expectation of giving birth (Heimstad et al., 2006; Lukasse, Schei, & Ryding, 2014). It may include fears about uncontrollable pain during labour and medical interventions, and intense anxiety that the woman or baby may die during delivery (Klabbers, 2016). FoC is estimated to affect approximately 10% of pregnant women when measured with specific scales such as the W-DEQ (Wijma et al., 1998), and it has been documented that at least 5-6% of all pregnant women experience fear related to childbirth that is severe or disabling (Heimstad et al., 2006; Lukasse et al., 2014). The increased interest towards FoC is mainly due to the observation that FoC appears to increase significantly the chance of a woman opting for an elective

caesarean section (elective C-section), when there is no medical indication to support this choice (Handelzalts et al., 2012). Lukasse and colleagues (2014) used the W-DEQ to conduct a large study aimed to examine the prevalence of FoC in six European countries and found that severe fear of childbirth affected between 4.5% and 15.2% of women (respectively in Belgium and Sweden). Evidence that FoC increases the risk of elective C-section deserves particular attention, considering that C-sections are known to increase the risk of negative health outcomes for women and babies (RCOG, 2012). Rates of caesarean sections in the Western world have consistently risen over the last few decades (Betran et al., 2018). In the UK, figures published by the UK National Maternity and Perinatal Audit (RCOG, 2017) and referring to the period 2015-2016 show that the mode of birth was caesarean in 25% of cases. 10.8% of women opted for an elective C-section (emergency 14.2%), compared to approximately 4% of women in 1980. If women who are experiencing moderate or severe FoC were identified at the earliest opportunity during pregnancy and given the appropriate support to manage these fears and worries related to giving birth, this could arguably result in a reduction of women opting for an elective C-section, with a considerable potential for improved birth and postnatal outcomes for mother and child. However, no measures of PrA or fear of childbirth are currently used in routine antenatal care in the UK.

2.3.3 A construct in search of a definition

It is evident from the above that the current research on antenatal anxiety is characterised by the use of a considerable variety of anxiety scales, including measures of general anxiety, specific anxiety disorders (e.g. GAD-7) and pregnancy-related anxiety. This heterogeneity of screening scales appears to reflect substantial differences in how researchers have defined and conceptualised the psychological construct of anxiety during the antenatal period (Grant et al., 2008; Brunton et al., 2015).

Based on the body of literature reported above and in Chapter 1 in relation to different types of clinically significant anxiety symptoms that pregnant women can experience, antenatal anxiety will be used in the following chapters to refer to pregnant women who meet at least one of these criteria:

- Diagnosed with one or more anxiety disorders, based on standard diagnostic classification
- Scoring above a validated cut-off score on a self-report anxiety measure

- Experiencing pregnancy-related anxiety or fear of childbirth as assessed by a PrA measure

Each of the criteria above is related to problematic anxiety symptoms during pregnancy. It is thus important to consider that anxiety in the antenatal period may present in different forms, all of which can cause significant distress to women and deserve clinical attention (Ayers et al., 2015). The general term antenatal anxiety will thus be used here to refer to the range of problematic anxiety symptoms that women can experience during pregnancy. When pregnancy-related anxiety is specifically discussed, this will be specified in the text. A more detailed definition of antenatal anxiety will be provided in Chapter 6.

2.4 Antenatal anxiety and the risk of negative outcomes for mother and child

The following sections provide a brief summary of the growing body of evidence which has examined the short- and long-terms effects of antenatal anxiety on a range of maternal and child outcomes.

2.4.1 Maternal postnatal disorders

Antenatal anxiety has been consistently found to be an important predictor of postnatal anxiety and depression in numerous studies conducted over the last two decades (Sutter-Dallay, Giaconne-Marcesche, Glatigry-Dallay, & Verdoux, 2004; Austin et al., 2007; Milgrom et al. 2008; Grant et al., 2008; Verreault et al., 2014). A large prospective study conducted in the UK by Heron and colleagues (2004), which was briefly discussed earlier, examined the patterns of self-reported anxiety and depressive symptoms from the second trimester of pregnancy up to eight months postnatally. The authors found that a considerable proportion of cases of postnatal anxiety and depression could be predicted antenatally. In particular, almost half (47%) of women experiencing postnatal depression had already reported elevated depressive symptoms in pregnancy. Anxiety in the perinatal period was even more stable, with 64% of women reporting elevated levels of anxiety symptoms both in pregnancy and in the postpartum period. A further, significant finding from this large study ($n > 8000$) was that antenatal anxiety predicted not only postnatal anxiety but also postnatal depression, both at 8 weeks and 8 months postpartum. Women who experienced anxiety at 32 weeks gestation had a more than three-fold increased risk to experience persistent postnatal depression (Odds ratio [OR] = 3.22), even after accounting for antenatal depression (Heron et al, 2004).

Antenatal anxiety as a predictor of postnatal depression

While there is evidence of antenatal anxiety as a risk factor for both postnatal anxiety and depressive symptoms, evidence regarding the association between antenatal anxiety and postnatal depression is particularly robust. A considerable number of studies have consistently indicated that women experiencing antenatal anxiety have a three- to five-fold increased risk of developing postnatal depression (Lee et al., 2007; Verreault et al., 2014;

Norhayati et al., 2015). Antenatal anxiety appears to predict postpartum depression both when measured with a self-report rating scale and when formal diagnostic criteria are used to determine the presence of an anxiety disorder in pregnancy. For example, Grant and colleagues (2008) found that symptoms of anxiety in pregnancy meeting diagnostic criteria led to significantly greater odds (OR: 4.97) of experiencing a depressive or anxiety disorder in the months following childbirth. A recent systematic review has confirmed that antenatal anxiety is an important risk factor for the development of postpartum depression (Goodman et al., 2014). The authors of this review found a significant association in all of the reviewed studies investigating the role played by antenatal anxiety symptoms in predicting maternal postnatal problems. While most of the studies discussed above used general anxiety measures, when pregnancy-related anxiety scales were employed, they were also predictive of postnatal depression (Blackmore et al., 2016). A somewhat surprising finding in a number of studies is that antenatal anxiety appears to predict postpartum depression more accurately than antenatal depression (Matthey, Barnett, Howie, & Kavanagh, 2003; Heron et al., 2004; Verreault et al., 2014). The combination of these findings would appear to indicate that screening for antenatal anxiety provides a key opportunity for targeting in pregnancy, through early identification and support, women at risk of postnatal mental health disorders.

2.4.2 The effects on child: Possible mechanisms of transmission

Before discussing the range of negative birth and child developmental outcomes found to be associated with antenatal anxiety, it is important to briefly outline the main hypothesis that has been proposed to explain the mechanisms by which maternal antenatal anxiety can negatively affect fetal development, birth outcomes and child developmental trajectories. Gestation is a time of rapid cell division and organ development for the fetus, which during this period is particularly sensitive to both beneficial and detrimental influences from the maternal environment, with potential consequences on health outcomes across the lifespan (Davis & Sandman, 2012; Cao-Lei, Laplante & King, 2016). This is commonly known as the fetal programming hypothesis (Barker, 1998), which postulates that fetal development is strongly influenced by responses of the fetus to intrauterine conditions throughout the prenatal period (Hocher, 2014). Maternal physiological responses to anxiety or significant stressors can affect the release of specific hormones which can, in turn, alter the fetal environment (Aizer, Stroud, & Buka, 2009; Dunkel-Schetter & Lobel, 2011; Staneva et al., 2015; Van den Bergh, 2016).

Glucocorticoids, and in particular cortisol, have been indicated as primary candidates for fetal programming in the context of maternal antenatal anxiety (Wadhva, 2005; Blair, Glynn, Sandman, & Davis, 2011). Cortisol is secreted by the hypothalamic-pituitary-adrenal (HPA) axis, and a large body of evidence exists indicating the association between exposure to stressors or persistent anxiety and increased levels of this glucocorticoid (Kirschbaum et al., 1995; Hunter, Minnis, & Wilson, 2011). It is estimated that approximately 40% of in-utero levels of cortisol are determined by concentration levels of maternal cortisol (Gitau et al., 2001). This is of particular importance as there is now evidence that in utero exposure to elevated levels of glucocorticoids, including cortisol, can have detrimental effects on the fetus and on brain development (Reynolds, 2013). One of the mechanisms by which excessive levels of fetal cortisol may have a detrimental effect on the developing fetus is by affecting fetal growth, thus increasing the risk of premature birth and lower birth weight (Goodman et al., 2014), with potentially poorer short- and long-term health outcomes (Barker, 1998; Dueker, Chen, Cowling, & Haskin, 2016).

It has been suggested that the detrimental effects on the fetus and the developing child are also likely to be mediated by timing and length of exposure to maternal prenatal distress (Stein et al. 2014). In the context of antenatal anxiety, this hypothesis is supported by a number of studies, which found that maternal anxiety is a better predictor of adverse birth outcomes when it is prolonged during pregnancy, as well as in the third trimester of gestation (DiPietro, Hilton, Hawkins, Costigan, & Pressman, 2002; Lobel et al., 2008; Blair et al., 2011). This observation is of critical importance, as it suggests that early identification and treatment of women experiencing antenatal anxiety might considerably increase the chance to prevent, or at least reduce, the risk of poorer fetal and child developmental outcomes.

In sum, it has thus been proposed that the potential detrimental effects of maternal antenatal anxiety on the child may begin in utero (Staneva et al., 2015). Although the specific mechanisms in action to explain the underlying association between prenatal exposure to maternal anxiety and negative birth and child outcomes remain uncertain, evidence now exists documenting the specific association between antenatal anxiety and increased risk for a range of adverse neonatal and child developmental outcomes. A brief overview of this body of research is provided in the following sections.

2.4.3 Obstetric and birth outcomes

In recent years, an increasing number of studies have investigated the association between antenatal anxiety and a range of fetal and neonatal outcomes (O'Connor, Heron, Golding, & Glover, 2002; Heron et al., 2004; Dunkel Schetter et al., 2012). The research literature has focused mainly on two neonatal outcomes: gestational age at birth and infant birth weight.

Gestational age at birth is an important indicator of neonatal wellbeing (Tucker & McGuire, 2004). Premature birth, commonly defined as birth before the completion of the 37th gestational week (RCOG, 2017), is a well-known risk factor for subsequent health problems throughout the lifespan (Barker et al., 1998; Hagberg & Wennerholm, 2000). The association between maternal antenatal anxiety and greater risk of premature birth has been examined in numerous studies ((Mancuso, Schetter, Rini, Roesch, & Hobel, 2004; Berle et al., 2005; Dunkel-Schetter and Lobel, 2011, Ibanez et al., 2015). In a systematic review and meta-analysis of the effects of maternal anxiety in pregnancy on length of gestation (Ding et al., 2014), twelve studies were reviewed and a statistically significant, although modest, association was found between antenatal anxiety and preterm delivery, with a pooled risk ratio = 1.50. On the other hand, the authors of a narrative summary of the literature conducted in the same year argued that the evidence regarding the association between antenatal anxiety and premature birth is still inconclusive, because of the small number of studies and the frequent comorbidity of depressive and anxiety symptoms which may be a confounding factor (Stein et al., 2014). However, a more recent systematic review (Staneva et al., 2015) reported consistent associations between elevated levels of anxiety in pregnancy and preterm birth, with the majority of reviewed studies supporting this hypothesis.

There is also evidence suggesting an association between elevated anxiety in pregnancy and lower birth weight. Low weight at birth, similarly to premature birth, is a significant predictor of poorer short- and long-term health outcomes (Phillips, 2002). Two recent systematic reviews investigated the effects of maternal anxiety during pregnancy on birth weight. A study by Ding and colleagues (2014) reviewed six studies and concluded that there was evidence to indicate that exposure to maternal anxiety in pregnancy increases the risk of low birth weight (pooled risk ratio = 1.76). A second review (Goodman et al., 2014) applied more stringent inclusion criteria, by examining only studies in which the presence of antenatal anxiety was established according to formal diagnostic criteria (WHO, 1992; APA, 2013). Three studies found that infants of mothers who had experienced specific anxiety disorders, namely specific phobia, PTSD and panic disorder, were more likely to

have lower birth weight when compared to infants of women in a control group. In contrast, three other investigations did not show any significant association between lower weight at birth and Generalised Anxiety Disorder or any anxiety disorder. These two reviews indicate that a modest association between antenatal anxiety and lower birth weight may exist, although mixed findings were reported.

In conclusion to this brief overview of the effects of antenatal anxiety on neonatal outcomes, the body of research discussed above would appear to support the hypothesis that maternal anxiety during pregnancy is associated with higher risk for negative fetal and birth outcomes. Despite the relative paucity of studies which have examined these adverse neonatal outcomes in the context of antenatal anxiety, the overall evidence shows the potential detrimental effects of antenatal anxiety, particularly on length of gestation. The next section summarises the literature on the effects of antenatal anxiety later in life, from the early years up to late adolescence.

2.4.4   Association with child development outcomes

If antenatal anxiety can negatively affect fetal development and birth outcomes, it may also have an adverse impact on child development outcomes. A growing body of research has emerged in recent years investigating this potential association, and numerous studies have been conducted to examine a range of neurodevelopmental, cognitive, affective, and behavioural outcomes for children whose mothers had experience of anxiety in pregnancy (Van den Bergh et al., 2005; Talge et al., 2007; Glover, 2016). An investigation conducted on Australian women in their third trimester of pregnancy (Austin, Hadzi-Pavlovic, Leader, Saint, & Parker, 2005) indicated that elevated maternal anxiety at this time predicts difficult infant temperament, an index of emotional reactivity, at 4 and 6 months of age (OR = 2.56). Importantly, antenatal depression was not found to be a risk factor for problematic infant temperament. An American cohort of women and their babies were assessed from pregnancy up to twelve months postpartum to examine the effects of antenatal anxiety on cognitive functioning and fine and gross motor skills (Keim et al., 2011), and found that elevated levels of anxiety in pregnancy were associated with poorer overall infant cognition, after controlling for a number of other variables. Early cognitive and motor development was also recently investigated in a French study conducted with over 1300 mother-baby dyads (Ibanez

et al., 2015). The researchers found strong associations between maternal anxiety assessed at 24-28 weeks gestation and poorer developmental trajectories at two and three years of age, particularly in the domains of communication, fine motor and personal-social skills as measured by the Ages and Stages questionnaire. Once again, this association was not found to be significant for antenatal depression. This appears to highlight the crucial and specific, adverse role played by antenatal anxiety on child developmental outcomes. Similar findings were reported in other recent studies (Glover et al., 2016; Lin et al., 2017).

A number of studies have also shown significant associations between antenatal anxiety and behavioural and affective problems during childhood and pre-adolescence. Van den Bergh and colleagues (2004) showed that antenatal anxiety during the late first trimester and second trimester was predictive of ADHD symptoms and poorer emotional regulation in 8- and 9-year olds, with 22% of the variance in these symptoms explained by maternal antenatal anxiety. The significant association remained even when a range of factors, including postnatal maternal mental health problems and parental educational level, were controlled for. In another investigation, 7-year olds of mothers who experienced anxiety during pregnancy had a two-fold increased risk for emotional and behavioural problems (O'Connor et al., 2003). A modest association between antenatal anxiety and poorer behavioural and affective outcomes in children aged 10 and 11, as reported by mothers and teachers, was also found by Leis and colleagues (2014).

Several large, longitudinal studies have also reported poorer developmental outcomes and worse self-reported mental health during adolescence, up to age 18. These investigations are commonly based on large cohorts, as in the example of the ALSPAC (Avon Longitudinal Study of Parents and Children) study, a prospective cohort study which recruited pregnant women in the Avon region (UK) during 1992 and followed up their children for an extended period of time to investigate the association between maternal characteristics and a wide range of health and social outcomes. Capron and colleagues (2015), analysing data from a subset of over 4000 adolescents from the ALSPAC population cohort, showed that maternal anxiety in pregnancy increases the risk of self-reported anxiety or depression at 18 years of age (adjusted odds ratio = 1.39). Other studies have indicated an association between anxiety during pregnancy and increased risk for poorer outcomes in adolescence, including higher impulsivity (Van den Bergh et al., 2005) and higher risk of experiencing anxiety or depression (Betts, Williams, Najman, & Alati, 2014). A criticism that could be raised regarding the body of research discussed above relates to the potential role of a wide range

of confounders in determining child developmental trajectories. However, at least one study based on a large dataset from the ALSPAC cohort showed that, even when a variety of potential confounders such as maternal postnatal mental health problems, socio-economic status, and level of education were controlled for, the negative effects of antenatal anxiety on offspring persisted (O'Donnell et al., 2014).

Some authors have suggested that, based on a review of the literature, antenatal anxiety appears to account for approximately 10-15% of adverse developmental outcomes (Glover, 2014). It could thus be argued that, although uncertainty remains about the real magnitude of the effects of antenatal anxiety on child development, the relatively common occurrence of antenatal anxiety and the consequent high number of children potentially affected imply that even a relatively small negative effect on birth and child developmental outcomes can still result in a considerable impact on public health (Ding et al., 2014; Bauer et al., 2014). In sum, an increasing body of literature documenting the negative impact of antenatal anxiety on maternal postnatal mental health and on a range of offspring outcomes now exists. Early recognition and timely treatment of women experiencing antenatal anxiety may thus have significant, beneficial effects in preventing a wide range of negative outcomes for mother and child, as discussed in the following section.

2.5 The importance of early identification and support

The findings presented above have important clinical implications. First, there is substantial evidence documenting the association between antenatal anxiety and postnatal depression and anxiety (Verreault et al., 2014; Goodman et al., 2014). Maternal antenatal anxiety can thus be considered an early marker that health professionals could use to identify during pregnant women at higher risk of poor mental health in the postnatal period (Grant, 2008). This is also true for the negative impact that antenatal anxiety, probably through the mechanism of fetal programming, can have on birth and child developmental outcomes. While further research is needed to improve our understanding of the interplay of a range of biological, psychological and contextual factors in determining adverse birth and developmental outcomes, the evidence reviewed above clearly indicates that antenatal anxiety is at least moderately associated with potentially serious consequences for the child, both short- and long-term. Of particular importance is the observation of the cumulative effect of antenatal anxiety in predicting child outcomes, with some studies showing that prolonged periods of exposure to maternal antenatal anxiety are associated with worse outcomes for the child (Lobel et al., 2008; Blair et al., 2011). These adverse effects are, however, not inevitable. The antenatal period, with its frequent contacts between women and healthcare professionals, provides important opportunities for prevention. If women experiencing problematic anxiety in pregnancy are identified early, they can be offered the appropriate support which is likely to result in a reduction of symptoms and improved outcomes, given that evidence-based interventions and forms of treatment exist (Marchesi et al., 2016). At a time when public funds for health services are under restraint, prevention is recognised by the NHS as a strategic objective, with the *NHS Five-year Forward View* (NHS England, 2014) emphasising the need for "a radical upgrade in prevention and public health" (p. 9). Perinatal mental health care is arguably one of the rare clinical areas in which preventative strategies and interventions can potentially improve outcomes in two individuals at the same time, and reduce the significant health, social and economic costs associated with perinatal mental health problems. This was clearly illustrated in a report published by the London School of Economics (Bauer et al., 2014), which estimated the costs for healthcare services and the wider society of neglecting perinatal mental health problems in the UK. This analysis was the first of its kind, as it estimated the economic costs of perinatal depression, anxiety and postpartum psychosis, taking into account the adverse effects of maternal mental health problems on children as well as women. The report

indicated a staggering figure of approximately £8.1 billion as the total cost of the impact of untreated perinatal mental health problems for each one-year cohort of births in the UK, with nearly three-quarters (72%) of this cost related to adverse impact on the child rather than the mother (Bauer et al., 2014).

A strong economic argument thus also exists to promote the early assessment and treatment of perinatal mental health problems. However, as previously discussed, under-recognition of mental health problems throughout the perinatal period remains a major issue. This is particularly true for common mental health difficulties such as perinatal anxiety and depression (NICE, 2104), with a recent report from the Centre of Mental Health estimating that in the UK only half of all cases of perinatal depression and anxiety are identified, and even less receive evidence-based forms of treatment. (Khan, 2015). In the context of antenatal anxiety, while the introduction of the GAD-2 as a brief screening measure is certainly a positive step, this scale and many other anxiety scales currently used in research and clinical settings have a number of important limitations when used to assess specifically antenatal anxiety.

2.6 Issues in screening for antenatal anxiety

The assessment and identification of women experiencing antenatal anxiety is problematic for a number of reasons. As noted in the first chapter, both in research settings and in busy clinical practice self-report scales are commonly favoured over clinical diagnostic interviews, which are time-consuming and require specific training. In the vast majority of cases, self-report scales used to assess antenatal anxiety were originally developed for use in the general population (Meades & Ayers, 2011). This is also the approach adopted by NICE (2014) to identify women experiencing poor mental health during the perinatal period in routine antenatal care, with the recommendation to use brief screening tools developed for the general population such as the PHQ-2 and the GAD-2 to screen for perinatal depression and anxiety respectively. It was also noted earlier that the GAD-2 consists of two questions *("During the past month have you been feeling nervous, anxious, or on edge?"* and *"During the past month have you not been able to stop or control worrying?"*), with scores ranging from 0 = *'not at all'* to 3 = *'nearly every day'*. A cut-off score of 3 or above is recommended by NICE to screen for problematic anxiety symptoms. By examining the content of these two questions, it could be argued that while women with more severe anxiety symptoms are likely to score ≥ 3, this cut-off may not be appropriate to discriminate reliably between the common feelings of anxiety that can be experienced during pregnancy, and problematic anxiety symptoms that would require appropriate support. The GAD-2 consists of the initial two questions of the GAD-7, which was developed to assess specifically Generalised Anxiety Disorder. NICE recommends use of the longer GAD-7 for further assessment, or referral to a GP or mental health practitioner, if a woman scores ≥ 3 on the GAD-2 (NICE, 2014). However, perhaps surprisingly, at the time of the publication of the guidelines no studies on the screening accuracy of the GAD-2 in pregnant or perinatal populations existed. Since then, very limited psychometric evidence for the GAD-7 (Zhong et al., 2015) and the GAD-2 (Nath et al., 2018) has been published. The NICE Guideline Development Group in the full version of the guideline commented that, considering the inadequate evidence on the screening accuracy of case identification tools specific to the assessment of perinatal anxiety, the recommendation to use the GAD-2/7 drew heavily on the evidence base from other guidelines for screening tools in non-pregnant populations (NICE, 2011).

Over the course of the last decade, three systematic reviews have examined the psychometric properties of anxiety scales used in pregnant or perinatal populations (Meades & Ayers,

2011, Evans et al., 2015, Brunton et al., 2015). Meades and Ayers (2011) reviewed general anxiety measures with published psychometric data in the perinatal period, while Evans and colleagues (2015) examined both general anxiety and pregnancy-related anxiety measures, but limited their review to the antenatal period. On the other hand, Brunton and colleagues (2015) focused exclusively on PrA scales. Two findings, common to all these reviews, are of particular relevance here. First, the authors of the three reviews highlighted the general lack of satisfactory psychometric properties (i.e. indexes of scale reliability and validity, discussed in Chapter 3) of anxiety scales used in the antenatal period, as well as the lack of measures which take into account both general antenatal anxiety and PrA. In second place, the reviews revealed the considerable variety of anxiety scales used to assess antenatal anxiety which, as previously observed, would seem to reflect substantial differences in how researchers have defined and conceptualised this psychological construct (Meades & Ayers, 2011; Brunton et al., 2015). It has also been observed that in numerous cases researchers have opted for a specific anxiety scale primarily because of its good psychometric properties in the general population (Grant et al., 2008). However, an anxiety scale developed for use in other populations may not retain its psychometric properties when used to assess anxiety in pregnant women, for a number of reasons which are briefly discussed here.

Confounding role of physical symptoms

One of the main limitation of numerous 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). Physical complaints such as dizziness, sleep problems or difficulty to relax can be relatively common experiences in pregnancy, without necessarily being indicators of poor mental health (Tebbe et al., 2013). Some of these symptoms, however, resemble closely symptoms of anxiety disorders and are thus included in many anxiety scales. This can potentially lead to inflated scores and high numbers of false positives when these scales are used to screen for anxiety in pregnant women (Lee et al., 2007; Ayers, Coates & Matthey, 2015).

Inaccurate cut-off scores

A further, significant issue in the use of general anxiety measures to identify women experiencing antenatal anxiety is that cut-off scores validated for other populations to

distinguish 'cases' from 'non-cases' might not be accurate when used in the antenatal period. The body of literature on antenatal anxiety provides numerous examples of studies in which researchers have used a specific cut-off score simply based on its widespread use in the general population, without further psychometric testing specific to pregnant women (Lee et al., 2007; Rubertsson et al., 2014). However, cut-off scores require recalibration when they are used in populations with substantial differences from the one they were originally validated for (Jomeen & Martin, 2005a). Meades and Ayers (2011), for instance, observed that the STAI has been used in different studies on antenatal anxiety with cut-off scores ranging from >40 to >48 (Grant et al., 2008; Field et al., 2010), without any prior validation of these cut-offs in pregnant populations. This can lead to inaccurate prevalence estimates, as in the case of Lee and colleagues (2007) who used the Hospital Anxiety and Depression Scale (HADS) with the cut-off validated in the general population (> 7 for the anxiety subscale) and found that over 35% of women in their sample experienced antenatal anxiety based on this measurement. In sum, anxiety scales developed for the general population require further psychometric testing and validation when used in the antenatal period. However, only a paucity of studies have provided any evidence of the measurement properties of scales when they are used to measure antenatal anxiety, and this can lead to misinterpretation of scores and inaccuracy of findings (Evans et al., 2015). The use of the GAD-2 as recommended by NICE (2014) is thus only one of various examples of the application of scales developed for the general population to the perinatal period. While in research studies this choice may lead to inaccurate findings, the implications of using a potentially inaccurate measure in clinical settings are obviously more problematic.

General antenatal anxiety vs pregnancy related anxiety scales

There is also uncertainty in relation to whether to use measures for anxiety disorders or scales developed to assess specifically pregnancy-related anxiety (Meades & Ayers, 2011) to identify women experiencing problematic anxiety symptoms during pregnancy. Symptoms related to pregnancy-specific anxiety and worries may not be detected by general anxiety measures. Conversely, PrA instruments developed exclusively to identify anxiety symptoms related to a current pregnancy will not identify women experiencing different anxiety disorders (e.g. Social Anxiety). However, as discussed previously in this chapter, both anxiety disorders and pregnancy-related anxiety are of clinical importance and deserve attention. While the screening accuracy of the GAD-2 in identifying women experiencing

general antenatal anxiety still needs to be established, this brief measure certainly does not appear to be appropriate to screen for pregnancy-related anxiety.

To conclude this section, it would appear clear from the above that current assessment methods of antenatal anxiety, including the one currently recommended by NICE (2014), are not evidence-based and may thus lead to incorrect identification, potentially creating unmotivated worry and anxiety. Following this overview of the current issues in screening for antenatal anxiety, this chapter concludes by providing a brief rationale for the development of a novel screening scale for antenatal anxiety.

2.7 The development of a screening scale for antenatal anxiety

The lack of self-report scales with an adequate evidence base to assess antenatal anxiety documented above constitutes a significant barrier to the recognition of women experiencing problematic anxiety symptoms during pregnancy, and a number of authors in recent years have advocated the use of a brief scale for the universal screening of antenatal anxiety (Rubertsson et al., 2014; Brunton et al., 2015; Biaggi et al., 2016). It could be argued that, ideally, such a scale should be developed specifically for the antenatal period and take into account symptoms of both general antenatal anxiety and PrA. NICE, in its review of scales to be considered for the identification of perinatal mental health problems which informed the most recent guidelines (2014), only considered brief scales as potentially feasible to implement in maternity care settings. As observed in Chapter 1, brief scales were defined as those containing less than 12 items (NICE, 2014).

Timely and effective screening procedures for antenatal anxiety are crucial in order to identify women who would benefit from monitoring, and where appropriate, early intervention, with a high potential for prevention. However, recent reviews on the topic (Meades & Ayers, 2011; Brunton et al., 2015) have showed a general lack of satisfactory psychometric properties for screening tools currently used to assess anxiety in pregnancy. Therefore, the availability of a short, reliable, and easy-to-complete screening scale for significant antenatal anxiety is pivotal in order to improve the detection of pregnant women experiencing various forms of problematic anxiety symptoms.

The PhD programme of work discussed in this thesis aimed to contribute to fill this gap by developing a self-report measure of anxiety specifically constructed to be used with pregnant women in research and clinical settings. The rest of this thesis will be devoted to report on the methods used to develop this measure and on the different stages of scale development and preliminary psychometric testing. The experimental chapters conclude with Chapter 8, which presents a psychometric validation study in which the new antenatal anxiety screening scale and the GAD-2/7 were validated against a gold standard clinical diagnostic interview.

Chapter 3 Methodology

The main aim of the research documented in this thesis was the development and initial psychometric validation of a self-report scale to screen for antenatal anxiety. In this chapter, the initial sections discuss a number of theoretical and methodological considerations in scale development and psychometric testing, including the important role of psychometric properties in the evaluation of the quality and accuracy of a scale. The chapter continues by providing an overview of the research methods and study design of this programme of work, and a rationale is given for the choice of Classical Test Theory as the guiding theoretical framework of the research. The chapter concludes by discussing the potential ethical issues arising from the research in relation to the participation of pregnant women as study subjects, and how these were addressed to minimise any potential discomfort or distress for study participants and safeguard the confidentiality of the sensitive information they provided.

3.1 Theoretical and methodological considerations in psychometrics and scale development

3.1.1 The measurement of psychological constructs

Measurement is a central component of scientific research (Kline, 2005). In its simplest form, measurement involves the collection of data and the assignment of a numerical value to a characteristic or attribute of the phenomenon under investigation (Furr, 2011). In the natural sciences, techniques and instruments to measure physical quantities have become extremely accurate over the centuries. However, in the fields of social, behavioural and psychological sciences, many of the phenomena of interest are often intangible and elusive (DeVellis, 2012). This is also at least partly true for healthcare research and clinical practice, in which the measurement of intangible phenomena such as quality of life, depression or stress has received increasing attention over the last decades and has become an important aspect of providing evidence-based and patient-centred care (Stewart, 2001; Rattray & Jones, 2005).

In social and psychological research, intangible phenomena that cannot be directly observed are commonly referred to as *constructs* (Messick, 1995; Streiner & Norman, 2008). These include phenomena as varied as emotional states (e.g. anxiety), personality characteristics (e.g. introversion), personal needs (e.g. autonomy) and many others (Atkinson et al., 2004). In this thesis the term 'construct' will be used specifically in the context of psychological research to indicate phenomena such as cognitive or emotional states (e.g. symptoms of anxiety or depression) that cannot be observed or measured directly. Other interchangeable terms such as latent construct or variable, target construct or underlying construct (DeVellis, 2012) are also used in the literature, and at times in this thesis, to highlight the intangible and variable (i.e. not stable) nature of constructs.

The field of psychometrics has emerged over the course of the last century to provide standardised and objective ways of assessing psychological constructs. While a range of measures can be used to assess constructs, including clinician-based behavioural observations, psychophysiological devices, and structured diagnostic interviews, it has been indicated that the most common method of assessment and measurement of psychological constructs is the self-report rating scale (Simms, 2008). This type of scale is the focus for the remainder of this chapter.

Although several approaches exist within the field of psychometrics, the theoretical framework that has dominated scale development and psychometrics until recently is Classical Test Theory (Loevinger, 1957; Kline, 2005). The research presented in this thesis was based on this theory, and consequently the following section is mainly focused on the key principles and assumptions of this approach. In more recent years, Item Response Theory (IRT) has also been used as a somewhat different approach to scale development and validation. A brief description of IRT is provided later in the chapter (3.3.1), which also discusses the strengths and limitations of both approaches and provides a rationale for the choice of Classical Test Theory as the overarching theoretical framework of this programme of work.

3.1.2 Classical Test Theory and the use of self-report rating scales

One of the key assumptions of Classical Test Theory (CTT) is that although constructs cannot be directly observed, they can be assessed and measured by means of effect indicators (Streiner & Norman, 2008). Effect indicators are essentially behavioural, cognitive or affective manifestations of the latent construct (DeVellis, 2012). For example, when assessing depression, self-reported hopelessness and loss of interest in activities are often used by clinicians as effect indicators of the underlying construct of depression. Similarly, self-report measurement instruments (i.e. scales) consisting of a number of effect indicators can be used to measure levels of a latent construct in an individual. In CTT, effect indicators are thus considered to share a common, underlying cause (i.e. the latent construct) and are used to identify and quantify the experience of a subject in relation to the target construct (Rattray & Jones, 2005). Self-report rating scales have become increasingly popular over the course of the last decades and have been extensively used in both research and clinical settings to assess a wide range of constructs. Their simplicity, speed of completion, and the fact that they can be administered to large numbers of individuals at a relatively low cost, are all likely to have contributed to their popularity (Clark & Watson, 2003).

While a variety of terms have been used for instruments measuring constructs, including rating scale, inventory, screening tool, and patient-reported outcome measure (PROM), in this thesis the term *scale* will be generally used to refer to self-report measurement instruments used to assess and measure levels of a latent construct. In a scale, effect indicators take the form of statements or questions (e.g. *"I feel tense"*) which are typically

referred to as scale items or more simply *items*. Respondents are asked to score each item (i.e. assign a numerical value to it based on a specific attribute such as its frequency or severity) and a total score is produced by summing the scores of all items as a measure of the overall level of the target construct. A scale can thus be broadly defined as "a collection of items combined into a composite score and intended to reveal levels of theoretical variables not readily observable by direct means" (DeVellis, 2012, p11). This type of scale, also typically known as a 'summated rating scale' or 'Likert scale' is by far the most commonly used in research and clinical settings (Simms, 2008), and its properties are further discussed below and in Chapter 6. Other types of scales also exist, such as visual analogue scales (a graphic rating method) or Guttman scales (a hierarchical type of scale typically used in achievement tests). This thesis will make use of the 'summated rating scale'. The focus will also be on self-report scales as opposed to observer-based or clinician-administered scales.

A further key concept in CTT is the distinction between observed score and true score, which is particularly relevant to summated rating scales. CTT postulates that an individual's observed score on an item or a full scale is the result of two components: the true score plus some random error of measurement (Netemeyer, Bearden, & Sughash, 2003; Streiner & Norman, 2008). This can be simply illustrated by the formula below:

$$O \text{ (observed score)} = T \text{ (true score)} + E \text{ (random error)}$$

As unsystematic errors in observed scores are assumed to be random and normally distributed, CTT implies that multiple observations of the construct (e.g. as tested by different items in a summated rating scale) will result in a reduction of the overall error component (Spector, 1992). Many efforts in CTT have been devoted to develop techniques to minimise the random error component of the observed score. The reliability of a scale, for instance, can be seen as an index of the proportion of variance of the observed score that is attributable to the true score (Tavakol & Dennick, 2011). A scale that is perfectly reliable would thus be purely a reflection of the true score and would not be influenced by measurement error (DeVellis, 2012). Scale reliability is only one of a range of measurement properties, also commonly known as psychometric properties, which are used to evaluate the accuracy and overall quality of a scale. A brief overview of different forms of psychometric properties is provided in the following section.

3.1.3 Psychometric properties: the important role of reliability and validity

Psychometric theory is specifically concerned with the assessment of the psychometric properties of scales of latent constructs (Bowling, 2014). Evaluating the psychometric properties of a scale is a crucial aspect of scale development and validation. Psychometric properties, in fact, indicate the extent to which a scale measures what it purports to measure (validity) in a consistent and reproducible fashion (reliability) (Cook & Beckman, 2006). Reliability and validity are the two main types of psychometric properties that can be evaluated in a scale. In turn, different forms of reliability and validity exist. In general terms, reliability refers to the consistency, stability or repeatability of scale scores (Kumar, 2015). Validity is a broader concept consisting of multiple and somewhat different lines of evidence, all contributing to support claims that a scale measures what is intended to measure (Tavakol & Dennick, 2011). A brief outline and definition of different psychometric properties commonly evaluated in a scale is provided below. Throughout the thesis, psychometric properties will be further discussed in relation to the development and validation of a screening scale for antenatal anxiety, and the procedures that were used to construct a psychometrically robust scale will be described.

Reliability

As noted above, scale reliability refers to the consistency and repeatability of scale measurements (Kline, 2005; DeVellis, 2012). Scale reliability is an important psychometric property and a necessary prerequisite for scale validity. A scale thus cannot be considered valid unless its reliability has also been established (Tavakol & Dennick, 2011; Furr, 2011). Different forms of scale reliability exist, which include internal consistency (consistency of items within a scale), inter-rater reliability (scale consistency when used by multiple raters) and test-retest reliability (scale consistency over time). The relative importance of these different forms of reliability depends on the type of scale and its intended purpose (Cook & Beckman, 2006). The most commonly examined and reported form of reliability is internal consistency reliability (Abell, Springer, & Kamata, 2009). Internal consistency is essentially an index of inter-item correlations within a scale, and it can be formally defined as the degree to which the scale score is free from measurement errors (APA, 2000). DeVellis (2012) explains clearly the rationale behind internal consistency by observing that if items composing a scale all have a strong relationship to the latent construct, they will also have a

strong relationship to one another. A scale is thus considered to be internally consistent when inter-item correlations are robust, indicating that all items within the scale are measuring the same underlying construct. This is commonly determined by calculating Cronbach's Alpha (Cronbach, 1951), a widely used measure of scale reliability expressed as a number between 0 and 1, with 1 indicating perfect internal consistency. This important index of internal consistency is discussed further in Chapter 7 and 8.

Two other forms of reliability mentioned above, namely inter-rater and test-retest reliability, are not directly relevant to a self-report scale which aims to assess the construct of antenatal anxiety at a fixed moment in time and were thus not considered in the present research. Inter-rater reliability is commonly examined when an observer-based scale is used, and the agreement between different raters is calculated as a measure of the consistency of scale scores among different observers (Kline, 2005). This index is, however, clearly not applicable to self-report scales. Test-retest reliability examines the consistency of scale scores when a scale is completed by the same individual at least two times following an intervening period of time (DeVellis, 2012), and should only be measured when the target construct is considered to be stable over time (e.g. trait anxiety, extroversion, intelligence). In the case of antenatal anxiety, Chapter 2 documented that fluctuations in anxiety levels typically occur throughout pregnancy and thus it was considered not appropriate to examine the consistency over time of the scale as it would be incorrect to assume that scores should remain stable over consecutive administrations.

Validity

Establishing validity is a crucial step in scale development and psychometric validation. While internal consistency indicates the degree to which items in a scale are all measuring the same construct, it does not provide any indication of whether the construct measured is in fact the latent construct that the scale developers originally intended to measure (Netemeyer et al., 2003). This is the role of validity estimates, which overall can be described as the degree to which scale scores can be confidently interpreted as a reflection of the true level of the target construct (Norman & Streiner, 2008). This makes validity arguably the most important form of psychometric robustness of a measure (Furr, 2011). As previously noted, several different types of validity exist, and no single form of validity is sufficient to determine whether a scale can be considered a valid measure of the construct of interest. It

follows that scale validity is a matter of degree rather than an "all-or-none" property of a scale (Furr, 2011). Traditionally, a distinction between three different types of validity, namely content, construct and criterion validity, was made by different scholars (Landy, 1986). In more recent decades, a number of other forms of scale validity have been proposed, as in the case of construct validity which some authors suggested should be based on evidence of both convergent and discriminant validity (DeVellis, 2012). Others have argued that the various types of validity are all essentially forms of construct validity (Abell et al., 2009; Streiner & Kottner, 2014). This thesis follows the approach indicated by Streiner & Norman (2008) who note that, while it is important to evaluate different aspects of validity, the terminology used to describe different forms of validity is somewhat secondary. The key consideration is that scale validity should be established through a process of hypothesis testing which is performed through a number of procedures, each dealing with different aspects of scale validity. For clarity of exposition, a brief outline of different types of validity is provided here, and each of these different forms of validity is discussed further in this and subsequent chapters of the thesis.

Face validity: This is arguably the simplest form of scale validity, and it refers to whether a set of items appears to be measuring the construct of interest "on the face of it" (Streiner & Norman, 2008, p5) as judged by experts or the intended respondents. This intuitive form of validity is rarely evaluated through empirical approaches, and has thus been criticised as superficial and subjective (Cook & Beckman, 2006). Some, however, have argued that scales composed of items that are clear and open regarding what they asses are more likely to increase motivation and cooperation of respondents to complete the scale (DeVellis, 2012).

Content validity: This type of validity is strongly linked to the definition of the target construct articulated by the scale developers, and is concerned with evaluating whether a set of items composing a scale collectively reflect the target construct and are all relevant to its measurement (Netemeyer et al., 2003; Furr, 2011). Content validity is typically established by consulting experts in the area of the construct of interest, who can judge whether they consider the items to reflect the construct of interest. In order to make this process more objective, it is recommended that the feedback of experts is subject to some form of quantification (Abell et al., 2009). Experts can also identify content areas of the construct that have been omitted in a scale (i.e. that are not represented by any items) and potentially suggest further items for inclusion (Watson & Clark, 2003; Hunsley & Mash, 2008).

Construct validity: Construct validity is commonly evaluated by testing whether a scale sufficiently correlates with other scales measuring theoretically similar constructs and is relatively independent from scales purported to measure different constructs (DeVellis, 2012). As noted above, however, some authors have suggested that the construct validity of a scale should be evaluated more broadly, by examining and considering the procedures used throughout the scale construction process. The two following forms of validity are often both considered as types of construct validity (DeVellis, 2012):

Convergent validity: The degree of correlation of a scale with existing scales measuring constructs that are theoretically related to the target construct (Furr, 2011) is calculated to determine convergent validity. Correlation indexes are used (Pearson's *r* or Spearman's rho, depending on whether the assumption of normal distribution is met), with the hypothesis that the scales will exhibit moderately large to large correlations.

Discriminant validity: The degree of correlation of a scale with other scales which measure constructs that are anticipated to be relatively or entirely unrelated to the target construct is calculated to establish discriminant validity. Similarly to convergent validity, Pearson's *r* or Spearman's rho are used, and small to moderate correlations are used as evidence of the discriminant validity of a scale.

Structural validity: This type of validity refers to the factor structure of a scale and is determined by administering the scale to a validation sample and conducting Principal Component Analysis, Exploratory Factor Analysis or Confirmatory Factor Analysis on scale scores. Factor analysis is commonly used to reduce variables (i.e. single items) that share common variance into set of clusters (i.e. factors) (Bartholomew et al., 2011). Factors can thus be described in terms of percentage of the total variance explained by each factor (Netemeyer et al., 2003). Examining 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.

Criterion validity: As the name suggests, criterion validity is concerned with the correlation of a scale with a criterion measure or 'gold standard' of the target construct (Streiner & Norman, 2008). In psychological scale validation, the gold or reference standard is generally considered to be a clinical diagnostic interview based on well-established diagnostic criteria (Gibson, McKenzie-McHarg, Shakespeare, Price, & Gray, 2009), such as DSM or ICD criteria (WHO, 1992; APA, 2013). It is important to note that, although diagnostic interviews

are commonly regarded as an accurate method of assessment of psychological morbidity (Reed et al., 1998), they are not perfect diagnostic tools and may be affected by issues of inter-rater reliability (Pinninti, Madison, Musser, & Rissmiller, 2003). Despite this observation, diagnostic interviews are typically considered the best available approximation to a 'gold standard'. Criterion validity is arguably the most powerful indicator of scale validity for scales assessing psychological disorders, as it describes how precise the scale score is in distinguishing between individuals with the target condition and those without the target condition. Various statistical indexes can be used to evaluate criterion validity, as discussed in the next section.

3.2 Processes and stages in scale development and validation

The development of a self-report scale for the assessment of a psychological construct is a sequential process and a number of steps and procedures are required in order to develop a scale with robust psychometric properties (Kline, 2005; DeVaus, 2014). In this section, the key principles and techniques of scale development and validation recommended by various scholars are discussed and an overview of different factors that should be taken into account when developing a self-report rating scale is provided. Although no universally recommended procedures exist for scale development, a number of guidelines and frameworks are available to guide investigators in the construction of a psychometrically sound scale (Nunnally & Bernstein, 1994; Kline, 2005; Streiner & Norman, 2008; Abell et al., 2009; National Institutes of Health [NIH], 2012: DeVellis, 2012). Consistently with the view that scale validity is established through the accumulation of different lines of evidence, ideally a number of sources should be consulted during the scale development phase. These can include, among others, the research literature related to the construct of interest, members of the target population and individuals with specific expertise in the area of the construct. An overview of the main stages in the development and psychometric testing of a scale, as recommended by both DeVellis (2012) and Streiner & Norman (2008), is summarised in the next page in Figure 1 and each stage is discussed in further detail in the rest of this section.

Figure 1 – Key stages in scale development and psychometric validation, based on Streiner & Norman (2008) and DeVellis (2012).

```
┌─────────────────────────────────┐
│  Definition/conceptualisation of │
│        target construct          │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│   Generation of an initial item  │
│      pool based on proposed      │
│             construct            │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│    Item reduction and refinement │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│    Preliminary psychometric      │
│    testing and further item      │
│             reduction            │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│      Psychometric validation     │
└─────────────────────────────────┘
```
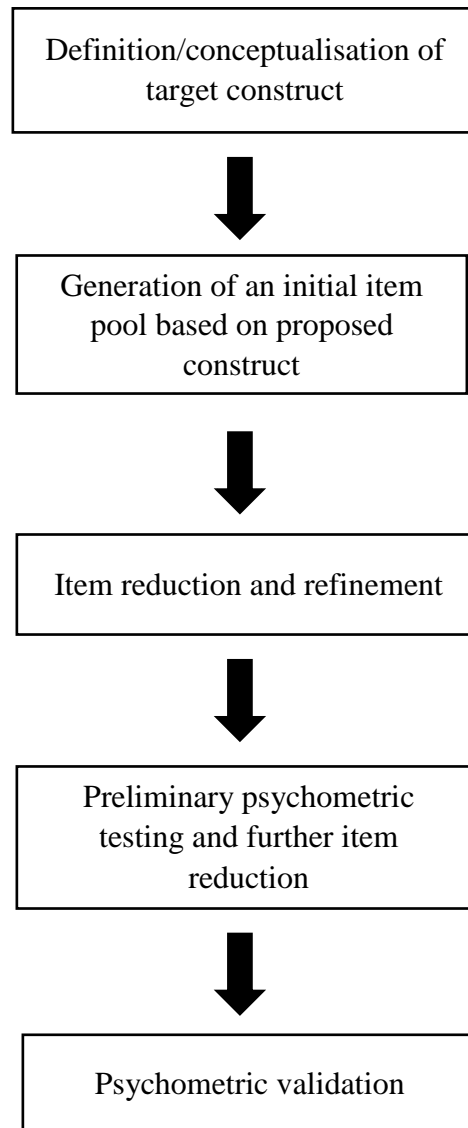
Definition of the construct: The initial stage in the development of a scale is to clearly define the construct that the scale intends to measure (Netemeyer et al., 2003; Kline; 2005; Streiner & Norman, 2008). Although it may appear obvious that a scale needs to be unambiguous with regard to what it purports to measure, it has been indicated that it is not uncommon for scale developers to underestimate the importance of a a clear definition of the target construct (Abell et al., 2009; DeVellis, 2012). The validity of a scale, however, cannot be accurately assessed unless the nature of the construct is delineated and its boundaries are clarified (Keedwell & Snaith, 1996). A definition of the phenomenon of interest (i.e. the target

construct) serves a number of purposes. First, it informs and guides the generation and selection of an initial pool of items for potential inclusion in the scale. Without a well-defined construct, it is difficult to develop good items to measure it, with the potential risk that the scale will have poor reliability and validity (DeVellis, 2012). Second, defining the construct is useful in clarifying its boundaries, thus making it clear not only what the scale intends to measure, but also what it should not measure (Clark & Watson, 2003). Furthermore, once a final version of the scale is developed, a number of validation hypotheses can be tested with regard to the relation of the target construct to other scales, thus contributing to establish the construct validity of the scale. The research literature on the construct of interest is often recommended as a starting point that can guide the conceptualisation of the target construct (Kline, 2005; Streiner & Norman, 2008). Theories related to the construct of interest, prior conceptual definitions and empirical evidence regarding the nature of the construct can all provide a foundation to define the phenomenon that the scale developer wants to measure (Spector, 1992). In the case of psychological traits or disorders, individuals with direct experience of the condition of interest can also be consulted and contribute to determine which domains (i.e. content areas) should be included in the construct definition.

Item generation and format of measurement: The second stage of scale development is the generation of items (i.e. questions or statements) that reflect the construct of interest. This phase is commonly referred to as the operationalisation of the construct (Streiner & Norman, 2008), which essentially corresponds to the translation of the construct definition into an initial item pool for potential inclusion in the scale (Kline, 2005). While the key aim is to formulate items that accurately reflect the proposed construct, in scale development it is desirable to start with a large pool of items which is initially over-inclusive, by also including items that are only tangentially related to the target construct (Netemeyer et al., 2003; Furr, 2011). By allowing some items with marginal content, scale developers improve the chance that all domains of the target construct are represented in the initial item pool (Abell et al., 2009). A second, important consideration in item generation is that, at this stage, item redundancy (i.e. items with similar or overlapping content) is also a desirable characteristic of the initial item pool (Clark & Watson, 2003; Furr, 2011; DeVellis, 2012). In particular, for domains that are deemed to be central in the construct definition, multiple items can be generated which tap into the same content area in slightly different ways (e.g. minor differences in wording). At later stages of scale development, quantitative analyses and inputs from experts and potential respondents regarding the relevance, wording and clarity

of individual items can be used to determine which version of the item can be considered superior. In order to ensure that items in the initial item pool have good face and content validity, they can be generated from a number of sources. The scale developer can formulate items de novo based on the construct definition, and draw on specific items from extant scales with demonstrated psychometric properties for the assessment of the construct of interest (Netemeyer et al., 2003; Kline, 2005; Bowling, 2014). Key informants such as individuals from the target population and experts in the field can also be consulted as sources for the generation of items (Spector, 1992; Rattray & Jones, 2007; Streiner & Norman, 2008; DeVellis, 2012). Norman & Streiner (2008) point out that these different potential sources for the generation of items, and the corresponding methods used to gather relevant information, are not mutually exclusive. Rather, different sources can all contribute to enhance item clarity and relevance to the target construct and consequently the psychometric properties of the final scale.

During the stage of item generation, the scale developers also need to make a number of decisions regarding the format of measurement, including the type of response format (e.g. Likert scale, Guttman scale), the stem question, the number of response categories and how they are worded, and the timeframe assessed by the scale (e.g. past week, past month). These decisions should be taken at the same time of item generation, and need to be guided by carefully considering both the target population and those who will typically score and interpret the measure (e.g. health professionals) (Kline, 2005; Abell et al., 2009). These aspects of scale design are discussed in detail in Chapter 6, which documents the item formulation phase and the decisions taken in relation to these more practical aspects of scale development.

Item reduction and refinement: Once a comprehensive, initial item pool is generated, an assessment of the relevance of items to the target construct and a refinement of their quality can be carried out through qualitative and quantitative procedures. Problematic items can be identified at this stage and they can either be modified or discarded (DeVellis, 2012). One key objective in this phase is to reduce the number of items in the initial item pool to a scale of reasonable length before preliminary psychometric testing on a validation sample (Streiner & Norman, 2008; Furr, 2011). This is typically achieved by discarding items that are considered to be less relevant to the assessment of the construct of interest. Various authors recommend that all items are reviewed by 'judges' with specific expertise in the area of the target construct in order to make this process as objective and rigorous as possible

(Clark & Watson, 2003; Abell et al., 2009; DeVellis, 2012). In the case of psychological constructs, health professionals with clinical expertise of the condition of interest are clearly well-placed to assess the appropriateness and relevance of specific questions (i.e. items) to assess the condition of interest.

Potential issues related to the wording and clarity of items can also be addressed at this stage. It is crucial that scale items need are not only relevant and accurate indicators of the target construct, but also comprehensible and acceptable to the population of respondents (Clark & Watson, 2003). The readability of scale items is a key aspect that must be taken into account by scale developers. In this regard, a useful rule of thumb is thus to keep items as short and simple as possible, as length and complexity tend to negatively impact on item clarity (Netemeyer et al., 2003; DeVellis, 2012). For example, Streiner & Norman (2008) recommend that scales aimed at groups whose educational level is unknown should require reading skills that are not beyond the level of a 12-year old. In addition, well-written items should be unambiguous and designed to be interpreted by respondents in the same way (Furr, 2011). Moreover, the use of jargon, technical terms and double-barrelled items (i.e. asking more than one question at the same time) should all be avoided (Kline, 2005). While scale developers should attempt to address all these potential issues at the initial stage of item writing, further quality checks can be conducted at this stage. Individuals from the target population can be asked to pre-pilot scale items and provide feedback and comments on item clarity, wording, and acceptability, so that problematic items can be modified or dismissed.

Preliminary psychometric testing: Once scale developers have established which items show sufficient face and content validity and discarded or modified problematic items, ideally based on a number of sources, it is considered best practice to administer the obtained preliminary version of the scale to a small sample of individuals from the target population for preliminary psychometric testing (Rattray & Jones, 2007; DeVellis, 2012). This administration of a pilot version of the scale is mostly concerned with addressing any issues with the internal consistency reliability of the scale and selecting items for the final version of the scale. As discussed earlier, internal consistency is an essential psychometric property as a high internal consistency provides evidence that all scale items are measuring the same construct. At this stage, the contribution of individual items to the overall scale can thus be statistically analysed, and items that do not contribute significantly to the internal consistency of the scale can be considered for deletion. Different item and scale statistics can be examined (e.g. item response distributions, item-total correlations, inter-item

correlations) for this purpose (Streiner & Norman, 2008; Tavakol & Dennick, 2011). This procedure typically results in a shorter, final version of the scale that has good preliminary evidence of internal consistency. Two other important forms of validity, specifically construct and criterion validity, are commonly tested on a larger validation sample, in the final phase of the initial psychometric validation of a scale.

Psychometric validation: As noted earlier, criterion validity refers to the correlation of a scale with a criterion measure or reference standard of the target construct (Streiner & Norman, 2008). For scales that intend to measure a psychological trait or disorder, this corresponds to establishing the accuracy of the scale in discriminating between individuals with the condition of interest and those who do not have the condition of interest (i.e. cases and non-cases) at the optimal threshold or cut-off score (Kline, 2005). This is typically examined by conducting a specific type of psychometric validation study, commonly referred to as a study of diagnostic or screening accuracy in the medical and psychological sciences. In this type of study, subjects are asked to complete the scale under scrutiny, and are also assessed using the clinical "gold standard". A range of statistical indexes of screening accuracy, including sensitivity, specificity and positive and negative predictive values can be subsequently calculated. This procedure is discussed in detail in Chapter 7. While the main focus of a psychometric validation study is typically to evaluate the criterion validity of a scale, at this stage it is also considered best practice to evaluate other psychometric properties of the scale, which can include its factor structure, construct validity (i.e. by testing the correlation of the scale with other measures of related or unrelated constructs) and internal consistency as index of scale reliability. The next section discusses the rationale for the choice of CTT as the underpinning theoretical framework of this research and provides an overview of the study design.

3.3 Research methodology

3.3.1 Theoretical framework and rationale for research methodology

The overall aim of the programme of work presented in this thesis was the development and initial psychometric validation of a self-report rating scale to screen for antenatal anxiety. This was conducted using the theoretical framework of Classical Test Theory, whose key principles and assumptions were presented above. As noted earlier, over the last few decades Item Response Theory (IRT) models have provided an alternative to the CTT approach (Rose & Devine, 2014). CTT and IRT share the assumption that, when assessing a latent construct, every person can be described in terms of the true level of the latent variable, called theta in IRT models (Hambleton, Swaminathan, & Rogers, 1991). However, the focus of IRT models is at the item-level rather than at the scale-level as in CTT, and a set of mathematical equations are used to predict the probability of choosing a specific response to an item as a function of theta (Embretson & Reise, 2000). CTT was preferred to IRT as the overarching framework of the present research for a number of reasons, some related to theoretical considerations and others essentially pragmatic. First, as different authors have observed, while techniques based on IRT have become well established in the field of education, as they lend themselves particularly well to the testing of knowledge or abilities (Simms, 2008), CTT remains the dominant paradigm in the construction of scales measuring psychological traits or disorders (DeVellis, 2012; Petrillo, Cano, McLeod, & Coon, 2015). Kline (2005), for instance, notes that a number of psychometrically excellent scales in use in healthcare research and clinical settings were designed and validated based on the principles of Classical Test Theory. This may also be partly due to the fact that the theoretical assumptions and statistical techniques of CTT are somewhat more accessible to health researchers compared to IRT, which uses complex mathematical modelling to describe and analyse the relations between the latent construct and item responses in order to inform scale development (Simms, 2008). A further disadvantage of IRT is that it typically requires considerably large samples, with a number of authors indicating that at least 500 subjects are required for statistical analyses based on IRT models (Hambleton et al., 1991; DeVellis, 2012; Kean & Reilly, 2014). Considering the practical limitations and the relatively limited timeframe of a PhD, the recruitment of such a large number of subjects would have been highly challenging and probably unrealistic. In contrast, the principles of CTT enable scale developers to conduct initial psychometric validation of a newly developed scale with relatively smaller sample sizes (Clark & Watson, 2003; Abell et al., 2009). Moreover,

DeVellis (2012) argues that in most cases CTT and IRT produce relatively comparable scales when scale developers carefully consider issues related to the reliability and validity of the final scale at all stages of scale development and validation.

With regard to the study design, a sequential mixed-method design was used to achieve the research aim and objectives presented in Chapter 1. Mixed-method research can be broadly defined as the integration of quantitative and qualitative research methods (Wilson & MacLean, 2011). Traditionally, quantitative and qualitative approaches were characterised by considerable differences in relation to their epistemological paradigms, with quantitative methods commonly based on positivist paradigms and qualitative methods typically drawing on the philosophical assumptions of constructivism (Pluye & Hong, 2014). The use of mixed-method research has been consequently at times criticised for combining approaches derived from somewhat opposing epistemological backgrounds (Doyle, Brady, & Byrne, 2009). However, it has been proposed that mixed-method research provides a "third paradigm", in which the combination of quantitative and qualitative methods can lead to a more thorough and elaborated understanding of the phenomenon under investigation, thus enhancing the credibility and validity of the findings (Johnson, Onwuegbuzie, & Turner, 2007; Bowling et al., 2014). While the theoretical principles of CTT underpinned the psychometric aspects of scale development and validation in this study, the philosophical stance of pragmatism informed the mixed-method design and guided the choice of specific research methods used at different stages of the research. Pragmatism rejects the argument that quantitative and qualitative research methods are incompatible (Reichardt & Rallis, 1994; Bowling et al., 2014) and its epistemological foundations are based on the idea that a researcher should choose "the combination or mixture of methods and procedures that works best for answering your research questions" (Johnson et al., 2004, p17). The philosophical approach of pragmatism is particularly well-suited to scale developers and consistent with the theoretical framework of CCT, in which issues related to different forms of reliability and validity are typically addressed by using different research methods, with each stage of the research building on previous steps to maximise the psychometric properties of the final scale (DeVellis, 2012). This is clearly illustrated in the next section of this chapter, which provides an outline of the phases of the present research and the range of research methods that were used to address different research questions.

3.3.2 Summary of research design

The present research is broadly based on the guidelines for scale development and validation proposed by DeVellis (2012), who provides clear guidance on the development and psychometric testing of scales to measure social and psychological constructs, and Streiner & Norman (2008), who in their book focus specifically on the development and use of health measurement scales. Whereas these two sources served as the basis for designing the study presented here, specific techniques and procedures used also drew on numerous other guidelines and psychometric textbooks (Spector, 1992; Nunnally & Bernstein, 1994; Netemeyer et al., 2003; Kline, 2005; Simms, 2008; Abell et al., 2009; Furr, 2011). As previously noted, the development of a scale for psychological assessment is a stepwise, sequential process (Streiner & Norman, 2008). Consequently, a number of research methods were used in the construction and validation of a self-report rating scale for antenatal anxiety documented in this thesis. These included, in the initial phases of scale development and refinement, a systematic review, semi-structured qualitative interviews, a Delphi study with a group of experts in perinatal mental health and psychometric testing of a preliminary version of the scale on a sample of pregnant women. Subsequently, the final version of the scale was tested to examine its screening accuracy and other psychometric properties by carrying out a psychometric validation study. The research was thus designed to answer the research questions presented in 1.3, which are also included in Table 3 presented in conclusion to this section.

For illustrative purposes, the present research can be subdivided into three main phases, as illustrated in Figure 2 and discussed in the next page.

Figure 2 – Phases and stages of the research

PHASE 1 Scale development	• Construct definition of antenatal anxiety • Generation of initial item pool • Decisions on format of measurement
PHASE 2 Scale refinement	• Consultations with key informants • Item reduction and refinement • Preliminary psychometric testing and further item reduction
PHASE 3 Psychometric validation	• Psychometric validation of the screening accuracy (criterion validity), internal consistency, construct validity and factor structure of the final version of the scale

These three phases essentially correspond to the various steps in scale development and psychometric validation outlined in 3.2. A brief outline of the different phases of the research is presented below to provide a general overview of the research design and of the research methods that were used. A table (Table 3) is also included at the end of the section to illustrate how a combination of quantitative and qualitative research methods were used to address the research questions presented above. Detailed descriptions of each phase, with regard to the research method, recruitment, data collection and analysis are presented in the following chapters of this thesis.

In Phase 1 of this research, Research Questions 1 and 2 were addressed by carrying out a systematic review of existing anxiety scales used with pregnant populations and conducting qualitative interviews with women with experience of problematic anxiety symptoms during pregnancy. Both the systematic review and the qualitative interviews were instrumental in informing the construct definition of antenatal anxiety and guiding the generation of an initial item pool to reflect the proposed construct. The systematic review was conducted specifically to examine the psychometric properties and content of anxiety measures used in pregnancy, in order to map a set of core anxiety symptoms and domains identified to be

psychometrically sound for the assessment of anxiety during pregnancy (Chapter 4). Evidence from the systematic review was subsequently combined with the findings of semi-structured, qualitative interviews with women with experience of significant anxiety symptoms during pregnancy (Chapter 5). The interviews aimed to explore the experience of problematic anxiety symptoms in the antenatal period and, similarly to the systematic review, identify anxiety symptoms and domains which could be considered relevant indicators of antenatal anxiety. Based on the findings from these two studies, a formal construct definition of antenatal anxiety was articulated, and items were generated reflecting the relative importance of different content areas in the assessment of the construct of interest (Chapter 6). At the stage of item generation, the measurement format of the scale was also determined, and decisions were made on the scaling response, the type and number of response options and the timeframe assessed by the scale.

During Phase 2, which addressed Research Question 3, key informants were consulted in order to refine the initial item pool and reduce the total number of items by discarding those which were considered not sufficiently relevant, clear, or acceptable to the population of potential respondents (Chapter 6). Key informants included both women with experience of mental health problems during the perinatal period, who reviewed all items in the initial item pool in relation to their wording, clarity and acceptability, and a group of health professionals with expertise in the area of perinatal mental health, who participated in a Delphi study and were asked to rate all items in relation to their importance as indicators of antenatal anxiety. As a result of this phase, only items reaching an adequate level of expert consensus regarding their relevance to the assessment of the construct of antenatal anxiety were selected to be included in a preliminary version of the scale. This preliminary version of the scale was subsequently completed by a sample of pregnant women in a cross-sectional survey to examine its psychometric properties and further reduce the total number of items in order to produce a brief and psychometrically robust final version of the scale, as detailed in Chapter 7.

In the third and final phase of the present research, a 10-item version of the scale was generated based on the findings from the previous phase. The final stage of the research consisted in a psychometric validation study, which aimed to address Research Question 4 and 5. The final version of the scale, along with the GAD2/7 (Spitzer, Kroenke, Williams, & Löwe, 2006) and the Edinburgh Postnatal Depression Scale (EPDS: Cox et al., 1987), were completed in a second cross-sectional survey by a different sample of pregnant women.

The convergent and discriminant validity of the scale, both indexes of construct validity, were evaluated by examining the correlations between the new scale, the GAD-7 (convergent validity) and the EPDS (discriminant validity). The factor structure of the new scale was examined by carrying out exploratory factor analysis. Internal consistency of the scale was also determined. Additionally, a subsample of women was also assessed via a structured diagnostic interview for anxiety disorders. The screening accuracy (i.e. criterion validity) of the new scale and the GAD-2/7 in discriminating between women who were or were not clinically anxious was subsequently determined by calculating a number of statistical parameters (i.e. sensitivity, specificity, positive and negative predictive values) for the new scale, the GAD-2 and the GAD-7. Analysis of Receiver Operating Characteristic (ROC) curves were also conducted to determine the optimal cut-off score for the new scale and the GAD-2/7 The psychometric validation study is presented in Chapter 8.

Table 3 illustrates the different phases of the research, the research methods used and the forms of validity addressed at each stage.

TABLE 3 – Overview of research methods used and type of validity addressed in subsequent stages of the research

Phase	Research question	Research method (Thesis chapter)	Type of validity addressed
Definition of target construct + item formulation *(Phase 1)*	RQ1: What should a construct definition of antenatal anxiety include in order to cover the core domains of significant anxiety symptoms in pregnancy?	Systematic review (Chapter 4)	Content validity Construct validity
Definition of target construct + item formulation *(Phase 1)*	RQ1: *As above* RQ2: Which items are the most appropriate to operationalise the proposed construct of antenatal anxiety into a self-report rating scale?	Semi-structured, qualitative interviews (Chapter 5)	Face validity Content validity Construct validity
Decisions on format of measurement (type of scale, response options, time frame assessed) *(Phase 1)*	RQ2: *As above*	Review of relevant literature, consultation with intended respondents (Chapter 6)	N/A
Item refinement and reduction *(Phase 2)*	RQ3: Which items are considered clear, relevant and acceptable by the target population and experts, and can thus be used to create a short and psychometrically robust self-report scale to screen for antenatal anxiety?	Delphi study, consultation with intended respondents, (Chapter 6)	Face validity Content validity Construct validity
Preliminary psychometric testing of a 30-item version of the scale and further item reduction) *(Phase 2)*	RQ3: *As above*	Cross-sectional survey - Psychometric testing (Chapter 7)	Internal consistency Content validity Construct validity

Psychometric validation – Final 10-item version *(Phase 3)*	RQ4: What is the evidence in relation to the convergent and discriminant validity, internal consistency and factor structure of the final version of the scale? RQ5: How does the new scale perform when compared to the measure currently recommended by NICE (GAD-2/7), and to expert assessment using a structured diagnostic interview; and what are the optimised cut-off points for maximising sensitivity and specificity of the scales?	Cross-sectional survey - Psychometric validation (Chapter 8)	Criterion validity Internal consistency Structural validity Construct validity (convergent and discriminant validity)

3.4 Ethical considerations

In this research, study participants were recruited either to be interviewed about their experience of anxiety symptoms during pregnancy (qualitative interviews in Phase 1 or structured clinical interviews in Phase 3) or to complete one of the two versions of the scale in the cross-sectional surveys. As noted above, only in the final psychometric validation study all participants were also asked to complete the GAD-7 and the EPDS. Before any women were recruited, ethical and management approval was sought and obtained from the South East Scotland Research Ethics Committee 02 and the Research & Development service of the NHS Greater Glasgow & Clyde (GG&C) Health (see Appendices 1 and 2). Internal approval was also received from the School Research Ethics Committee (SREC) of the Faculty of Health Sciences and Sport at the University of Stirling. In the next pages, ethical issues related to recruiting and obtaining informed consent from women taking part in the research are discussed in 3.4.1. Potential risks and burdens for research participants were also carefully considered at all stages, and a number of procedures and measures were in place in order to keep participants' discomfort and distress to a minimum, as explained in section 3.4.2. Finally, the issue of safeguarding the confidentiality of any sensitive information provided by study participants was also addressed (3.4.3). Further details on ethical considerations are discussed in Chapter 5, 7 and 8.

3.4.1 Recruitment and informed consent

Individuals invited to take part in research studies as study subjects need to be able to make an informed decision about their participation (Department of Health [DoH], 2005). The general principles related to obtaining informed consent as detailed in the most recent version of the Declaration of Helsinki (World Medical Association, 2013) were followed when designing this study. Consequently, the information sheets that all potential participants were given (see Appendix 3 for an example) included information about the purpose of the study, what it involved, the potential risks and benefits of taking part, and the way in which information they provided was going to be managed and stored. It was also made clear that participation was entirely voluntary. No participants were approached directly by the researcher in the initial recruitment phase, thus preserving their right to be free of intrusion (De Vaus, 2014). With regard to the qualitative interviews in Phase 1, all potential recruits were approached either by the coordinator of the Maternal Mental Health Scotland Change Agents (a group of women with lived experience of perinatal mental health problems who contributed to the research at various stages, as further detailed in Chapters 5 and 6) or by a Nurse Consultant in perinatal mental health who gauged the suitability and interest of potential participants to be interviewed. Women taking part in one of the two cross-sectional surveys were, in contrast, approached during routine antenatal clinics by midwives who were part of the participants' direct healthcare team Further details about the recruitment process and how this was designed to enable potential participants to make an informed decision regarding their participation to the research are discussed in the relevant chapters of this thesis.

3.4.2 Minimising risks and burdens for study participants

It was acknowledged that both the interviews on the experience of anxiety in pregnancy and the psychological scales that study participants were asked to complete contained questions potentially sensitive and distressing for women. In this cases, there is an ethical duty to inform study participants of this (Bowling, 2014). In relation to the qualitative interviews, all potential recruits were informed verbally (i.e. by the Change Agents' coordinator or the clinician) and in written form through the information sheet about the content of the interview before they decided whether to participate, and it was made clear that the interview looked into their experience of anxiety symptoms during pregnancy. In order to minimise any burdens and inconvenience for women who consented to be interviewed, who all lived in the Greater Glasgow & Clyde area, they were given the opportunity to choose a date and time that best suited them for the interview, and were offered a choice of location which included the interviewees' home and local NHS and university sites (the Mother & Baby Unit within the Leverndale Hospital and Glasgow Caledonian University). The author conducted all the interviews and reminded participants before commencing that they could pause or withdraw from the interview at any time without giving a reason. Time was also spent at the end of each interview to reflect on how participants found the process and briefly go through the information sheet, which included information on what to do in case a participant felt distressed following the interview, as also detailed in Chapter 5. For the two cross-sectional surveys, women were asked to complete one (preliminary psychometric testing) or three (psychometric validation) scales asking about symptoms of anxiety and depression. Self-report scales to screen for perinatal mental health problems are commonly used in routine maternity care and the potential distress caused to participants by completing these scales can generally considered to be minimal (Meades & Ayers, 2011). However, it was appreciated that the potential distress caused to participants by answering questions which asked about sensitive topics partially overlapped with that discussed for the qualitative interviews above and for the structured clinical interviews. Consequently, the information sheets given to all women taking part in the research contained information and contact details for accessing support in case they felt distressed as a result of discussing their own experience of anxiety in pregnancy or completing the self-report scales. The information sheets also included advice in relation to health professionals that could be contacted (GP or midwife) in order to discuss further any potential issues.

3.4.3 Confidentiality and safeguarding of sensitive information

While the recent EU General Data Protection Regulation (GDPR) was not in place yet when this research was conducted, previous legislative frameworks on data privacy regulation such as the Data Protection Act (1998), the NHS Research Governance Framework (DoH, 2005) and the Caldicott principles (DoH, 1997) were consulted when planning each of the studies contained in this thesis. The procedures that were followed to safeguard the confidentiality of sensitive information gathered from study participants at all stages of the research are discussed respectively in Chapter 5 (qualitative interviews), Chapter 7 (pilot psychometric testing) and Chapter 8 (psychometric validation).

The following chapters (4-8) document the three experimental phases of the research outlined earlier, commencing with the systematic review of the psychometric properties and content of scales used to assess anxiety in pregnant women presented in Chapter 4.

Chapter 4 Systematic review of anxiety scales used in pregnancy

4.1 Introduction and rationale for the review

The systematic review presented in this chapter was the initial, instrumental step in the development of a self-report scale to screen for antenatal anxiety. As discussed in the previous chapter, in scale development a review of the existing literature related to the construct of interest is recommended by various authors (Spector, 1992; Kline, 2005; DeVellis, 2012) as an essential part of the process of scale construction. Such a review appeared to be particularly appropriate for a construct such as antenatal anxiety, which has been measured using a variety of different instruments for general anxiety, specific anxiety disorders and pregnancy-related anxiety (Meades & Ayers, 2011). As noted in 2.3.3, this heterogeneity of measures used would appear to reflect substantial differences in construct conceptualisation. Consequently, a review of these previous attempts to measure the construct of antenatal anxiety was considered important.

The aim of this study was to systematically examine and synthesise both the psychometric properties and content of self-report scales used to assess anxiety in pregnancy in order to identify a core set of anxiety symptoms and domains with sound psychometric performance in pregnant populations. The study thus contributed to answer Research Questions 1 and 2:

- *What should a construct definition of antenatal anxiety include in order to cover the core domains of significant anxiety symptoms in pregnancy?*

- *Which items are the most appropriate to operationalise the proposed construct of antenatal anxiety into a self-report rating scale?*

This was achieved by conducting a systematic review of studies reporting at least one psychometric property (i.e. one aspect of reliability or validity) of a self-report measure used to assess antenatal anxiety, and by appraising and summarising the best available evidence in the form of a narrative synthesis. In order to guarantee that the conclusions were only based on the strongest evidence available, only studies of good or excellent methodological quality were included in the best-evidence synthesis of this review.

4.2 Method

The review was conducted based on guidance for undertaking reviews of clinical tests from the Centre of Reviews and Dissemination (CRD, 2009) and COSMIN (COnsensus-based Standards for the selection of health status Measurement INstruments) recommendations for systematic reviews of measurement properties (Terwee, 2011), and was reported according to the PRISMA statement (Moher et al., 2009).

4.2.1 Search strategy and selection criteria

Computerised searches were performed to query the following electronic bibliographic databases: MEDLINE, PsycINFO, and the Cumulative Index to Nursing and Allied Health Literature (CINAHL). The initial objective of the review was to locate primary research articles reporting psychometric properties of self-report rating scales used to assess anxiety symptoms in a pregnant population. The databases were searched from 1991 up to and including February 2017 and searches were restricted to articles published in peer-reviewed journals and available in English. A combination of four main themes was used in the search, as it was considered the most appropriate to retrieve relevant articles while keeping the scope of the search broad enough, thus reducing the risk of missing potentially relevant studies (CRD, 2009). Specifically, the major concepts searched were "anxiety", "pregnancy", "measurement" and "psychometrics" and search terms included both free text and MeSH terms. Major concepts and related synonyms for the four main themes were searched in the title and abstract fields, with several key terms also searched as a major concept within each database (see Appendix 4 for the search strategy and all search terms). Reference lists and citation records of papers included in the review were also inspected for potential inclusion of additional studies. Reports, commentaries, conference proceedings and other grey literature were not searched. Methodological search filters were not applied as there is evidence that, because of the variety of designs used in studies of diagnostic or screening test accuracy, applying methodological filters is likely to result in the omission of a significant number of relevant studies (Leeflang, Scholten, Rutjes, Reitsma, & Bossuyt, 2006; Whiting et al., 2011). A predefined list of inclusion and exclusion criteria was applied in relation to type of study, population, construct of interest and type of measurement. A complete list of inclusion and exclusion criteria is provided in Table 4.

Table 4 – Inclusion and exclusion criteria for the review

	Inclusion criteria	Exclusion criteria
Type of study	- Primary research articles reporting at least one psychometric property of a self-report rating scale used to assess anxiety symptoms in pregnancy - Published in a peer-reviewed journal in English in or after 1991	- Studies conducted in countries with substantial cultural differences with the UK (i.e. African and Asian countries) for which cultural equivalence cannot be assumed - Qualitative studies on the experience of anxiety symptoms during pregnancy
Population	- Pregnant or perinatal sample (for perinatal samples, subgroup analyses of psychometric properties of the measure available for the subsample of pregnant women)	- Sample composed exclusively of women with high-risk pregnancies, because of obstetric complications (e.g. pre-eclampsia, ectopic pregnancy) or high psychosocial risk
Construct of interest	- A specific anxiety disorder, as determined by DSM-5 or ICD-10 diagnostic criteria - General "anxiety" or "worry" - Pregnancy-related anxiety (PrA), as defined by Huizink and colleagues (2004) - Fear of childbirth, as this is deemed to be a relevant component of PrA	- Any other construct, as for example general mental health, mental disorders other than anxiety during pregnancy, stress or postnatal anxiety
Type of measurement	- All studies with published psychometric data using self-report rating scales to assess anxiety symptoms in pregnancy and meeting the inclusion criteria for type of study, population, and construct of interest detailed above were eligible for inclusion in the review	- Anxiety symptoms assessed exclusively with methods other than a self-report rating scale (e.g. open questions by a health professional, physiological measures of anxiety) - Scale designed to be completed by a health professional after observation (i.e. not self-report) - Scale not developed to generate a total score or single-item measures

4.2.2 Study selection and data extraction

All articles resulting from the electronic bibliographic database searches were imported into RefWorks and duplicates were removed. Titles and abstracts of articles resulting from the initial search were reviewed to identify potentially relevant studies. When there was an indication that an article may have met the inclusion criteria for the review, the full-text publication was obtained and reviewed. I screened titles and abstracts of all retrieved articles to determine their appropriateness for inclusion in the review. My first supervisor (HC) independently screened a sample (10%) of all retrieved articles to establish an index of inter-rater agreement determined as percent agreement (McHugh, 2012), which was 98% for titles and abstracts screened by both reviewers. Discrepancies were discussed and resolved by applying the relevant study eligibility criteria to reach consensus. The PRISMA flow diagram (Moher et al., 2009) was used to document the different stages of the study selection process. In relation to data extraction, the full-text article of all studies included in the review was inspected and the full version of the rating scale used was obtained in order to extract information relevant to the review. Data extraction forms and summary tables were developed and piloted on a small number of studies ($n = 6$) identified as eligible for inclusion at an early stage of the review. For each included study the following information was extracted: (a) author/s (b) year of publication (c) country (d) name of index test (e) sample size (f) timing of assessment (expressed as trimester or mean gestational week) (g) construct of interest. For each of the rating scales, I extracted: (a) number of items (b) type and number of response options (e.g. Likert scale, dichotomous) (c) timeframe assessed (e.g. past week, past month) (d) score range (e) total possible score (f) cut-off score (if available). In order to determine which psychometric properties were evaluated in each study, the COSMIN taxonomy and definitions of measurement properties were used (Mokkink et al., 2010a). The following psychometric properties were extracted: internal consistency reliability, convergent and discriminant validity, structural (i.e. factorial) validity, content validity (which was narratively summarised when reported) and criterion validity.

4.2.3 Quality assessment

An assessment of the methodological quality of each study included in the review was conducted using the COSMIN checklist, specifically developed to evaluate the study quality and risk of bias in systematic reviews of studies on the measurement properties of health

measurement instruments (Mokkink et al., 2010b). In this review, five of the nine possible boxes in the checklist were employed as they were considered to be relevant to evaluate the methodological quality of studies assessing the construct of anxiety in pregnancy. Specifically, these were box A (internal reliability), D (content validity), E (structural validity), F (hypotheses testing) and H (criterion validity). Each measurement property is scored on a four-point rating scale as "poor", "fair", "good" or "excellent". An overall score for the methodological quality of a study is determined by using a "worse score counts" system (Terwee et al., 2012). I performed the quality assessment for all studies included in the review, with my first supervisor (HC) assessing a random sample of studies (n=5) to confirm the accuracy of the scoring system. It was decided that only studies which achieved an overall rating of good or excellent were considered in the best-evidence synthesis in order to guarantee the quality of the conclusions reached by the review. The rationale for this decision was informed by the developers of the COSMIN checklist, who argue that low quality studies present a high risk of bias, which in a systematic review of measurement properties can lead to incorrect conclusions (Mokkink et al., 2010b)

4.2.4 Best evidence synthesis

The main aim of this review was to examine the psychometric properties and content of anxiety scales used in pregnancy, both at the scale and at the item level, in order to identify specific items (i.e. questions) or anxiety domains with established psychometric properties in this population. A synthesis of the best available evidence is presented for each scale in a narrative form. At the scale level, the psychometric properties discussed above were examined and synthesised. The number of studies, their methodological quality and the consistency of findings were taken into account. Specifically, the following criteria were used to classify the strength of evidence from one or more studies, based on COSMIN recommendations for quality criteria (Terwee et al., 2007): (a) *strong evidence*: consistent findings in multiple studies of good or excellent methodological quality or in one study of excellent quality (b) *moderate evidence*: consistent findings in multiple studies of good or excellent quality, except for one study with contrasting findings (c) *limited evidence*: one study of good methodological quality (d) *unclear or conflicting evidence*: contrasting results in multiple studies of good quality. Only items and anxiety domains with moderate or strong psychometric evidence of being accurate indicators of anxiety symptoms in pregnancy were

considered sufficiently robust, and contributed to inform the generation of items for the scale developed in this research, as discussed in Chapter 3 and detailed in conclusion of this chapter.

At the item level, the analysis was primarily based on factor analysis, and specifically on the examination and comparison of coefficients of item loadings on specific anxiety factors for each scale. In psychometrics, the examination of item loadings is recommended in order to determine which items within a scale possess the strongest psychometric performance in terms of their discriminative power (Streiner & Norman, 2008), and can be therefore considered to detect an important aspect of the construct assessed (DeVellis, 2012). Factor analysis in the psychometric testing of scales is used to reduce variables (i.e. single items) that share common variance into set of clusters (i.e. factors) (Bartholomew et al., 2011). In this review, the criteria proposed by Tabachnick and Fidell (2007) and listed as follows were adopted to evaluate the strength of item loading coefficients: (a) $0 - 0.44$ = poor; (b) $0.45 - 0.54$ = fair; (c) $0.55 - 0.62$ = good; (d) $0.63 - 0.70$ = very good; (e) > 0.70 = excellent. Only items which showed very good or excellent loadings (i.e. 0.63 or above), and for which the strength of evidence from one or multiple studies was moderate or strong according to the criteria discussed above, were considered to be psychometrically sound in measuring anxiety symptoms in pregnant women. When items forming a factor were found to be particularly homogeneous in relation to their content, the entire dimension or domain that the factor represented rather than individual items was selected as a domain identified as psychometrically sound. Secondary indexes that were examined at the item level when factor analysis was not conducted were the correlations between individual items and the remainder of items within a scale (corrected item-total correlations) and item discrimination parameters for analyses based on item response theory models.

4.3 Results

The initial search yielded 2879 citations, which were reduced to 1756 following deduplication in RefWorks. The titles and abstracts of remaining articles were screened for potentially eligible studies, resulting in 74 publications for which the full text article was retrieved. At this stage 47 studies were excluded and two publications were added from hand searches of reference lists of included studies. This resulted in a final sample of 29 studies included in the review. The main reasons for excluding studies after retrieving the full text were: (i) no psychometric data available (ii) construct of interest different from inclusion criteria (e.g. antenatal stress, general mental health) (iii) study participants recruited exclusively from high-risk samples. The study selection process is summarised in the PRISMA flowchart shown in Appendix 5.

The 29 included studies used 9 different scales as index tests to measure antenatal anxiety. The most commonly reported psychometric properties were internal consistency reliability ($n = 27$; 93% of studies), concurrent validity ($n = 21$; 72%) and structural validity ($n = 16$; 55%). Included studies showed a considerable degree of heterogeneity in relation to the construct assessed (i.e. general anxiety or an anxiety disorder or pregnancy-related anxiety), gestational age of participants, sample size and type of psychometric properties reported. As discussed in the method section, a quality assessment of all included studies was performed and only studies achieving a rating of good or excellent in relation to their methodological quality and risk of bias were included in the best evidence synthesis. Seven studies were given a rating of poor (Öhman, Grunewald, & Waldenström, 2003; Tendais, Costa, Conde, & Figueiredo, 2014; Haines et al., 2015) or fair (Levin, 1991; Jomeen & Martin, 2005b; Garthus-Niegel, Størksen, Torgersen, Von Soest, & Eberhard-Gran, 2011; Simpson, Glazer, Michalski, Steiner, & Frey, 2014) for their methodological quality and were thus not considered in the synthesis. The quality assessment of all 29 studies included in the review is presented in Appendix 6. The characteristics of all included studies are presented in Table 5.

Table 5 – General characteristics of studies included in the review

First author	Year	Sample size	Gestational age	Country	Index test	Time frame assessed	Target construct
Austin	2007	748	3rd trimester	Australia	BMWS	"General experience"	Worry
Bayrampour	2014	3021	2nd trimester	Canada	STAI (3 six-item short forms)	State (present time); Trait (general feelings)	State/Trait anxiety
Brouwers	2001	197	24 weeks	Netherlands	EPDS-A	Previous week	General anxiety
Carmona Monge	2012	285	Mean 14.1 weeks	Spain	CWS	Present time	Worry during pregnancy
Coates	2016	5551	18 and 32 weeks	UK	EPDS-A	Previous week	General anxiety
Fenaroli	2013	522	27-35 weeks	Italy	W-DEQ	Current expectations about childbirth	Fear of childbirth
Garthus-Niegel	2011	1642	32 weeks	Norway	W-DEQ	Current expectations about childbirth	Fear of childbirth
Gourounti	2012	132	11-14 weeks	Greece	CWS	Present time	Worry during pregnancy
Grant	2008	100	35-39 weeks	Australia	STAI (State and Trait forms)	State (present time); Trait (general feelings)	State/Trait anxiety
Green	2003	1207	1st/2nd/3rd Trimester	UK	CWS	Present time	Worry during pregnancy
Haines	2015	1410	2nd trimester	Australia	W-DEQ	Current expectations about childbirth	Fear of childbirth
Huizink	2004	172	1st/2nd/3rd trimester	Netherlands	PRAQ-R	Present time	Pregnancy-related Anxiety
Huizink	2016	1144	24 and 34 weeks	Finland	PRAQ-R2	Present time	Pregnancy-related anxiety
Johnson	2002	424	3rd trimester	UK	W-DEQ	Current expectations about childbirth	Fear of childbirth
Jomeen	2004	101	Mean 13.57 weeks	UK	HADS-A	Previous week	General anxiety
Jomeen	2005b	129	Mean 13.86 weeks	UK	CWS	Present time	Worry during pregnancy

Jomeen	2005a	101	Mean 13.57 weeks	UK	EPDS-A	Previous week	General anxiety
Karimova	2003	100	12 and 34 weeks	UK and Uzbekistan	HADS-A	Previous week	General anxiety
Levin	1991	266	2nd trimester	USA	PAS	Present time	Pregnancy-related anxiety
Marteau	1992	200	Gestational age not reported	UK	STAI (six-item short form)	Present time	State anxiety
Matthey	2013	132	Mean 14.9 weeks	Australia	EPDS-A HADS-A PRAQ-R	Various time frames	General and Pregnancy-related anxiety
Ohman	2003	200	8-42 weeks	Sweden	CWS	Present time	Worry during pregnancy
Petersen	2009	344	Mean 31.4 weeks	Germany	CWS	Present time	Worry during pregnancy
Simpson	2014	240	1st trimester	Canada	EPDS-A GAD-7	Previous week	General anxiety
Swalm	2010	4706	10-42 weeks	Australia	EPDS-A	Previous week	General anxiety
Tendais	2014	148	1st/2nd/3rd Trimester	Portugal	STAI-S	Present time	State anxiety
Westerneng	2015	6004	Mean 19.8 weeks	Netherlands	PRAQ-R	Present time	Pregnancy-related anxiety
Wijma	1998	196	32 weeks	Sweden	W-DEQ	Current expectations about childbirth	Fear of childbirth
Zhong	2015	946	Mean 9.6 weeks	Peru	GAD-7	Previous two weeks	Generalised anxiety disorder

Note: BMWS = Brief Measure of Worry Severity; CWS = Cambridge Worry Scale; EPDS-A = Edinburgh Postnatal Depression Scale – Anxiety subscale; GAD-7 = Generalised Anxiety Disorder – 7; HADS-A = Hospital Anxiety and Depression Scale – Anxiety subscale; PAS = Pregnancy Anxiety Scale; PRAQ-R and PRAQ-R2 = Pregnancy-Related Anxiety Questionnaire- Revised; STAI = State-Trait Anxiety Inventory; W-DEQ = Wijma Delivery Expectancy/Experience Questionnaire

4.4. Best evidence synthesis

Following an assessment of the methodological quality of all studies, 22 were included in the best evidence synthesis phase of the review. This section discusses the findings from these studies, through an examination of the psychometric properties of each scale and a critical analysis of the content of their items and anxiety domains found to be psychometrically sound for the assessment of antenatal anxiety. The analysis was carried out accordingly to the criteria discussed in detail in the Method section of this paper. For clarity of exposition, a synthesis is presented here separately for each scale, while the final section summarises the key findings of the review.

4.4.1 Edinburgh Postnatal Depression Scale – Anxiety subscale

The Edinburgh Postnatal Depression Scale (EPDS: Cox et al., 1987) is a 10-item self-report scale originally developed to screen for Postpartum Depression. Because of the lack of items specific to the postpartum period, the EPDS has also been validated for use with pregnant women (Murray & Cox, 1990; Green & Murray, 1994). Although the EPDS was developed as a unidimensional measure of depression, it was included in this review due to growing evidence that it contains a separate subscale measuring anxiety rather than depressive symptoms, in both antenatal and postnatal populations (Pop et al., 1992; Ross, Evans, Sellers, & Romach, 2003; Matthey et al., 2013a). Six studies included in this review examined the psychometric properties of the EPDS anxiety subscale in a sample of pregnant women. All studies except one (Simpson et al., 2014) achieved an overall methodological quality rating of good (Brouwers, van Baar, & Pop, 2001; Matthey, Valenti, Souter, & Ross-Hamid, 2013; Swalm, Brooks, Doherty, Nathan, & Jacques, 2010) or excellent (Jomeen & Martin, 2005a; Coates, Ayers, & Visser, 2016) and were thus included in the best evidence synthesis. Four of the five studies examined the factor structure of the EPDS to investigate whether the existence of an anxiety subscale could be confirmed. Brouwers and colleagues (Brouwers et al., 2001) performed exploratory factor analysis (EFA) of EPDS scores in women in their second trimester of pregnancy. The EFA revealed three components within the EPDS, namely two separate depressive (items 1, 2, 8) and anxiety (items 3, 4, 5) symptoms subscales and a third component consisting only of item 10 (*"The thought of harming myself has occurred to me"*). However, this third factor was not included in the final factor solution as the authors argued that a single-item loading could not plausibly

identify a distinct latent factor (Brouwers et al., 2001). A two-factor solution, comprising of separate depression and anxiety subscales, was therefore proposed. The three items of the anxiety subscale (item 3 "*I have blamed myself unnecessarily when things went wrong*", item 4 "*I have been anxious or worried for no good reason*", item 5 "*I have felt scared or panicky for no very good reason*") were the only ones, among the ten EPDS items, with item loadings on a single factor above the pre-defined cut-off of 0.63, ranging from 0.68 (item 3) to 0.73 (item 4). An examination of their content appears to indicate that these questions, all loading highly on a single factor, tap important affective and cognitive components of anxiety (e.g. feeling panicky or worried), although it could be argued that item 3 is more related to depressive symptomatology. Similar findings were reported by Jomeen and Martin (2005a) in a sample of women in their first trimester of pregnancy. EFA resulted in a three-factor solution which included depression and anxiety dimensions, and the same third factor identified by Brouwers and colleagues (2001). The items loading significantly (>0.63, range 0.73-0.85) onto the anxiety subscale were entirely consistent (items 3, 4, 5) with those identified in the previous study (Brouwers et al., 2001). The authors then conducted confirmatory factory analysis (CFA), a more refined data reduction technique than EFA allowing to test predefined factor solutions (Child, 2006), and tested various factor models including the original unidimensional depression model (Cox et al., 1987) as well as both a two- and a three-factor solution identified by Brouwers and colleagues (Brouwers et al., 2001). Results from the CFA revealed once again a clear superiority of the two-factor solution, thus confirming the previous finding that the EPDS both in early and in mid-pregnancy consistently measures two distinct dimensions of depression and anxiety. A further study included in this review (Matthey et al., 2013b) used the three-item EPDS anxiety subscale (EDS-3A) identified in previous studies to examine its criterion and convergent validity in pregnancy when compared to other anxiety measures. The EDS-3A performed better than both the anxiety subscale of the Hospital Anxiety and Depression Scale (HADS-A[82]) and the Pregnancy Related Anxiety Questionnaire-Revised (PRAQ-R: Huizink et al., 2004) in detecting women with an anxiety disorder as determined by DSM diagnostic criteria. Furthermore, the EDS-3A showed a moderately high correlation with the HADS-A (r = 0.68) and a low to moderate correlation with the PRAQ-R (r = 0.23), which may be interpreted as an indication that the three measures tap into different aspects of antenatal anxiety. While a potential limitation of the three studies reported above is their relatively small sample size (n < 200), the existence of an anxiety subscale within the EPDS was further confirmed in two subsequent studies with much larger sample sizes (n > 4000).

Swalm and colleagues (2010) examined the EPDS factor structure in an Australian sample across the three trimesters of pregnancy. A two-factor solution consisting of anxiety and depression components was found once more to be optimal, accounting for 55% of the score variance (anxiety subscale =29.4%; depression subscale =25.4% of the total variance). Moreover, an analysis of individual item loadings confirmed that items 3, 4 and 5 were the only ones with loadings higher than 0.63 on the anxiety subscale (range 0.75-0.78). A recent study (Coates et al., 2017) conducted both EFA and CFA on a large UK population-based sample at two time points (18 and 32 weeks gestation). Although both EFA and CFA indicated a three-factor model as the best factor solution, this was primarily due to the "depression" factor which was split into an anhedonia (items 1 and 2) and a depression (items 7-10) factor. Importantly, this was the only study in which item 3 "*I have blamed myself unnecessarily when things went wrong*" (0.56) did not reach the predefined item loading coefficient of 0.63. In summary, according to the criteria previously discussed to evaluate the strength of evidence in relation to the psychometric properties of reviewed scales, item 3 of the EPDS showed moderate evidence of its psychometric value, while items 4 and 5 demonstrated strong evidence of being psychometrically sound in assessing antenatal anxiety, as their item loadings on the anxiety subscale consistently exceeded the 0.63 cut-off in all reviewed studies.

4.4.2 Hospital Anxiety and Depression Scale – Anxiety subscale

The Hospital Anxiety and Depression Scale (HADS: Zigmond & Snaith, 1983) is a widely popular screening tool (Cosco, Doyle, Ward, & McGee, 2012), originally developed to assess anxiety and depression in non-psychiatric patients. This 14-item measure consists of two subscales (anxiety: HADS-A; depression: HADS-D), both comprising seven items and enquiring about feelings over the past week with four response options (Zigmond & Snaith, 1983). It is particularly important to establish the psychometric properties of the HADS when used in the antenatal period, as a considerable number of studies have used this screening tool to assess anxiety and depression levels in pregnant samples (Rubertsson et al., 2014; Owen et al., 2017). Three studies included in this review examined psychometric aspects of the HADS in a pregnant population (Karimova & Martin, 2003; Jomeen & Martin, 2004; Matthey et al., 2013b). They all achieved a rating of good in relation to their methodological quality. Karimova and Martin (2003) investigated the factor structure of the

HADS in the third trimester of pregnancy by conducting EFA of HADS scores in a sample of nulliparous women, and a post-hoc factor analysis revealed a two-factor solution. Specifically, six of the seven HADS-D items loaded higher on one factor and an equal number of HADS-A items loaded higher on a second factor. However, there was significant overlapping of item loadings on the two subscales, with only four HADS-A items (item 3 "*I get a sort of frightened feeling as if something awful is going to happen*"; item 5 "*Worrying thoughts go through my mind*"; item 9 "*I get a sort of frightened feeling like "butterflies" in the stomach*" and item 13 "*I get sudden feelings of panic*") loading above 0.63 on the anxiety factor. The authors therefore concluded that the 7-item HADS-A and HADS-D subscales do not reliably distinguish between anxiety and depressive symptoms in pregnancy. A further study was conducted by Jomeen and Martin (2004) on a sample of women in early pregnancy. Both EFA and CFA revealed a three-factor solution which confirmed that the HADS in pregnancy is not a bi-dimensional measure of anxiety and depression. However, a comparison of individual item loadings of the HADS anxiety subscale in the two studies was carried out in this review to examine psychometric information for each individual item within the HADS anxiety subscale. This is presented in the next page in Table 6.

Table 6 – Item loading coefficients of the HADS-A subscale in Karimova & Martin (2003) and Jomeen & Martin (2004)

HADS-A items	Karimova & Martin (2003) Anxiety factor (factor 2)	Jomeen & Martin (2004) Anxiety factor (factor 2)
1 I feel tense or wound up	0.18	0.31
3 I get a sort of frightened feeling as if something awful is going to happen	**0.67**	**0.74**
5 Worrying thoughts go through my mind	**0.78**	**0.69**
7 I can sit at ease and feel relaxed	0.33	0.07
9 I get a sort of frightened feeling like "butterflies" in the stomach	**0.65**	0.57
11 I feel restless as if I have to be on the move	0.57	0.36
13 I get sudden feelings of panic	**0.67**	**0.75**

Note: Item loadings in bold indicate loadings of **0.63 or above**.

The observation that three items of the HADS-A (items 3, 5, 13) are the only ones to reach an item loading above 0.63 on the anxiety factor in both studies is of particular importance. Although the two studies reached the conclusion that the seven-item HADS-A as a whole is not a psychometrically sound measure of anxiety in pregnancy, the three HADS-A items identified here showed a consistent pattern across the two studies, with significantly similar loadings on the anxiety factor. These items would therefore appear to have good psychometric value in assessing specific anxiety symptoms in pregnancy. A subsequent study (Matthey et al., 2013b) compared the screening performance of the HADS-A with diagnosis of an anxiety disorder according to DSM criteria. The authors found that high anxiety scores on the HADS-A, defined as the top 15% of scores, had poor concordance (34%) with formal diagnosis of an anxiety disorder. The poor concordance with DSM diagnoses seems to confirm the previous findings indicating that the 7-item HADS anxiety

subscale as a whole is not a reliable screening tool to assess anxiety in pregnancy. However, based on the evidence provided by the two studies discussed above on the factor structure of the HADS, it was concluded that the three identified items represent a shortened version of the HADS-A which, unlike the entire HADS-A, has good evidence of its psychometric properties to measure antenatal anxiety.

4.4.3 State-Trait Anxiety Inventory

The STAI comprises two subscales, each composed of 20 items. It is based on a model of anxiety which distinguishes between state and trait anxiety (Spielberger et al., 1983). State anxiety refers to the situation-specific, transient component of anxiety. Conversely, trait anxiety reflects a relatively stable personality trait, a dispositional anxiety proneness (Green, Kafetsios, Statham & Snowdon, 2003). Response options range from 1 (*not at all*) to 4 (*very much so)* for both the state and trait form, and each scale includes 10 anxiety-present (e.g. "*I am worried"*) and 10 anxiety-absent (e.g. "*I feel secure"*) items. The STAI has been widely validated in the general population (Austin et al. 2007) and is one of the most common measures used in research to assess anxiety in perinatal women (Meades & Ayers, 2011). This review located four studies reporting psychometric properties of the STAI in pregnant populations, one of which (Tendais et al., 2014) scored poor in relation to its methodological quality. Both the state and trait form of the STAI were used in an Australian study by Grant and colleagues (2008) on women in the third trimester of pregnancy. Internal consistency was found to be high for the full version of the scale, with a Cronbach's Alpha (α) of 0.95. A structured diagnostic interview was also used (Mini International Neuropsychiatric Interview [MINI]: Sheehan et al., 1998) to identify women meeting DSM-IV diagnostic criteria for an anxiety disorder. The authors found a cut-off score of 40 to yield the highest accuracy in identifying women with a diagnosed anxiety disorder, with a sensitivity of 81% and a specificity of 80%. However, they also acknowledged the limited generalisability of the findings due to the relatively small sample. The study did not provide any psychometric data at the item level and it was thus not possible to reach conclusions on the psychometric qualities of individual items measuring specific symptoms. A further study (Marteau & Bekker, 1992) tested various shortened versions of the STAI-S form to determine the smallest subset of items that preserved high correlations ($r > 0.90$) with the original, 20-item STAI-S. They found that a six-item version produced scores comparable to the full version

(r >0.94) while retaining a good level of internal consistency ($\alpha = 0.82$). The six items selected were the ones with the highest correlations with the remaining 19 items of the STAI-S (i.e. corrected item-total correlations). Specifically, the authors identified three anxiety-present and three anxiety-absent items, corresponding to the following emotional states: *calm, tense, upset, relaxed, content* and *worried.* This is a significant finding, as it identifies a number of symptoms (i.e. feeling tense, upset or worried) that correlate highly with the 20-item STAI-S total score, providing an initial indication that these anxiety-present symptoms may be considered relatively accurate indicators of problematic anxiety in pregnancy. This was confirmed in a further study by Bayrampour and colleagues (2014) which examined the psychometric properties of three six-item shortened versions of the STAI-S when compared to the full state form. The three short versions are the one discussed above (Marteau & Bekker, 1992) and two other versions developed in non-perinatal populations. The six-item version by Marteau and Bekker (1992) had the highest correlation with the sum score of the full form (r =0.94). Furthermore, confirmatory factory analysis was conducted and the version by Marteau and Bekker (1992) was found once more to consistently have the best values for all fit indexes considered, with the three anxiety-present items (i.e. feeling tense/upset/worried) all found to have coefficient item loadings above 0.63, a further indication of their psychometric soundness. In sum, the three items from the STAI-S short form discussed above were identified in two studies of good methodological quality (Marteau & Bekker, 1992; Bayrampour et al., 2014) as potentially reliable indicators of anxiety symptoms during pregnancy.

4.4.4 GAD-7

The GAD-7 (Spitzer et al., 2006) was developed in 2006 as a brief screening measure for Generalised Anxiety Disorder (GAD). Its original psychometric validation study in a large sample of primary care patients indicated very good screening accuracy in identifying people with a diagnosis of GAD (Spitzer et al., 2006). The scale consists of seven items asking respondents about some of the core GAD symptoms (e.g. excessive or persistent worry, trouble relaxing) experienced in the previous two weeks. As previously discussed, the first two questions of the GAD-7 (GAD-2) have been recently recommended by NICE (2014) as a brief screening measure for anxiety in perinatal women. Only two studies examining the measurement properties of the GAD-7 in a pregnant population were identified by this

review (Simpson et al., 2014; Zhong et al., 2015), and only one (Zhong et al., 2015) achieved a satisfactory rating for its methodological quality. Importantly, this was one of the few included studies which performed assessment of a scale against a gold standard diagnostic interview, the Composite International Diagnostic Interview (CIDI: Kessler & Üstün, 2004), to determine the criterion validity of the scale. In this antenatal sample at a cut-off score of 7 or above, different from the cut-off of 8 identified in the general population, the measure yielded moderately good sensitivity (73%) and specificity (67%) (Zhong et al., 2015). Internal consistency was close to excellent (α= 0.89). Both EFA and CFA were conducted and confirmed the unidimensional structure (e.g. a single factor) of the GAD-7 previously found in the general population (Spitzer et al., 2006). The results of the factor analysis indicated that the seven items loaded on a single factor with item loadings all exceeding 0.63. In order to identify which items provided the most accurate screening performance, I examined the item discrimination parameters, which are based on item response theory and indicate how well individual items differentiate between different levels of the target condition among respondents (Li & Baser, 2012). Two items showed considerably higher discrimination parameter estimates than the remaining ones. These were item 3 *"Worrying too much about different things"* (2.05) and item 2 *"Not being able to stop or control worrying"* (2.04), which clearly tap into the experience of pervasive or persistent worry typical of GAD. All other items exhibited substantially lower discrimination parameter estimates. Considering that this study was of excellent methodological quality, the two identified items have consequently strong evidence of their psychometric value in the antenatal period.

4.4.5 Brief Measure of Worry Severity

A single study (Austin et al., 2007) was located reporting psychometric data of the Brief Measure of Worry Severity (BMWS: Gladstone et al., 2005) in a pregnant sample. Self-report scales assessing the construct of 'worry' were included in this review (see Table 4) as worry is a core clinical feature of Generalised Anxiety Disorder (APA, 2013). A number of studies indicate that GAD is the most common anxiety disorder in pregnancy (for a review see Dennis et al., 2017) and for this reason worry can be hypothesised to be an important dimension of the construct of antenatal anxiety. The BMWS was developed as a unidimensional measure of the functional impact and severity of worry (Gladstone et al.,

2005). It includes 8 items assessing different aspects of worry with four verbally-anchored response options (*not true at all – definitely true)*. Internal consistency was very good (α = 0.89) in this antenatal sample, and the BMWS also showed good convergent validity with the STAI Trait ($r = 0.71$). While psychometric properties of the scale at the item level were not reported, there was evidence that the construct of worry as measured by BMWS is a reliable indicator of antenatal anxiety. First, the BMWS was found to have good construct validity in this sample of pregnant women, as it showed significant correlations with a number of other variables linked to a current episode of anxiety and depression (Austin et al., 2007). Moreover, it was a better predictor of postnatal depression than the STAI-S after controlling for possible confounding factors. As the literature indicates that antenatal anxiety is a strong predictor of PND (Milgrom et al., 2008; Grant et al., 2008), it would appear than the BMWS taps into a core component of antenatal anxiety considering its good predictive validity. Consequently the construct of worry has strong evidence of being psychometrically robust according to the criteria used in this review (i.e. consistent findings in multiple studies of good or excellent methodological quality), as it was also identified as psychometrically sound in other studies in this synthesis as detailed earlier.

4.4.6 Cambridge Worry Scale

The Cambridge Worry Scale (CWS) is a 16-item measure assessing the extent and content of women's worries during pregnancy (Green et al., 2003). The 16 items in the CWS enquire both about worries specific to pregnancy, such as "*The possibility of miscarriage"*, "*The possibility of something being wrong with the baby*" or "*Giving birth",* and more general concerns including "*Money problems"* and "*Your relationship with your family and friends"*. Items are scored on a 6-point Likert-type scale with verbally described anchors ranging from 0 (*Not a worry*) to 5 (*Major worry*). Six studies examining psychometric aspects of the CWS in a pregnant population were included in this review, four of which are considered here. The other two studies were rated as poor (Öhman et al., 2003) or fair (Jomeen & Martin, 2005b) for their methodological quality. Green and colleagues (Green et al., 2003) were the first to investigate the structural validity (i.e. the factor structure) of the CWS. A longitudinal design was used in a large sample ($n = 1207$) of women completing the CWS at gestational weeks 16, 22 and 35. The authors analysed scores at these three time points by means of principal component analysis (PCA), a form of exploratory factor

analysis. The PCA revealed a four-factor structure, consisting of the following factors: 1) *Socio-medical aspects of having a baby* 2) *Socio-economic issues* 3) *Health of mother and baby* 4) *Relationships with partner, family and friends*. This four-factor solution was subsequently replicated in all the other studies examined in this synthesis (Petersen, Paulitsch, Guethlin, Gensichen, & Jahn, 2009; Carmona Monge, Peñacoba-Puente, Marín Morales, & Carretero Abellán, 2012; Gourounti, Lykeridou, Taskou, Kafetsios, & Sandall, 2012). This can be considered robust evidence of factorial invariance of the CWS in different samples and stages of pregnancy. The convergent validity of the CWS was examined by comparing it with STAI state and trait scores (Green et al. 2003; Petersen et al., 2009; Gourounti et al., 2012) and with the anxiety subscale of the Symptom Checklist 90 (CL-90-R: Derogatis, 1977) by Carmona Monge and colleagues (2012). Two of the four CWS subscales were found to have the highest correlations with state anxiety (STAI-S) scores across studies. These were the "*socio-medical*" and the "*health of mother and baby*" factors. For the purpose of this review, the focus is specifically on these two factors, both because of their higher correlations with state anxiety and because the content of items in these subscales appears to reflect worries more closely related to pregnancy. Thus, an examination of individual item loadings for these two factors was carried out. In relation to the "*socio-medical*" subscale, one item ("*Giving birth*") was found to load above the predefined criterion of 0.63 in all four studies, thus demonstrating strong evidence of its psychometric properties in assessing a major worry in pregnancy. Another three items showed moderate strength of evidence as they loaded above 0.63 on the "*socio-medical*" subscale in all studies apart from one. Specifically, "*Internal examinations*" had an item loading coefficient of 0.61 in Gourounti and colleagues (2012), but item loadings above 0.63 in all the other studies; "*Going to hospital*" (0.68-0.79), apart from Gourounti and colleagues (2012) (0.47); and "*Coping with the new baby*" (0.65-0.68), except for the study by Petersen and colleagues (2009), in which its loading was 0.58. An inspection of the second factor examined, "*Health of mother and baby*," indicated two further items with loadings >0.63 in all the studies, namely "*The possibility of miscarriage*", which ranged between 0.75 (Green et al., 2003) and 0.85 (Carmona Monge et al., 2012), and "*The possibility of something being wrong with the baby*" (range 0.65-0.83). The other two items included in this subscale, "*Own health*" and "*Health of someone else close*," consistently loaded below the pre-defined cut-off. In summary, three items of the CWS ("*Giving birth*", "*The possibility of miscarriage*", "*The possibility of something being wrong with the baby*") demonstrated strong evidence of their psychometric value. Three further items ("*Internal examinations*", "*Going to hospital*"

"Coping with the new baby") showed a moderate strength of evidence of their psychometric performance in pregnancy.

4.4.7 Wijma Delivery Expectancy/Experience Questionnaire

The Wijma Delivery Expectancy/Experience Questionnaire (W-DEQ; Wijma et al., 1998) was developed in the late nineties to assess the construct of fear of childbirth. As observed in Chapter 2, within the research literature on pregnancy-related anxiety, fear of childbirth (FoC) or tokophobia has emerged as a central dimension of PrA (Heimstad et al, 2006; Blackmore et al., 2016). The W-DEQ -Version A (Wijma et al., 1998) includes 33 items enquiring about thoughts and feelings relating to the approaching childbirth, with six response options ranging from "*Not at all*" to "*Extremely*". Five studies included in the present review reported psychometric information on the W-DEQ in an antenatal population, and three studies achieved a good or excellent methodological quality rating (Wijma et al., 1998; Johnson & Slade, 2002; Fenaroli & Saita, 2013). In the original development study of the W-DEQ (Wijma et al., 1998), the authors provided good evidence of the face and construct validity of the W-DEQ, with all items formulated based on the clinical experience of the first two authors and incorporating women's input in the wording of items. The W-DEQ showed higher correlations with other anxiety measures than with extraversion or depression measures. However, these correlations were only moderate (STAI-T: $r = 0.54$; S-R Inventory of anxiousness: $r = 0.52$), thus showing a degree of conceptual overlap but also a sufficient level of variance left to indicate that the W-DEQ measures other than anxiety as a dispositional trait (Wijma et al., 1998). At the item level, item-total correlations were ranked and the authors examined the ten items with the highest ranking. Two domains of FoC, "*Negative feelings towards childbirth*" and "*Fear of labour and delivery*", were identified among the items more strongly correlated with the sum score, thus suggesting a stronger relation with the overall construct of FoC. As single items composing the W-DEQ are very specific to a given feeling or cognitive appraisal, it was considered appropriate for this review to focus on domains of FoC rather than individual items. Two other studies included in this synthesis conducted factor analysis of W-DEQ scores and found four distinct dimensions of the construct of FoC as measured by the scale. Johnson and Slade (2002) named the four identified domains *Fear, Lack of positive anticipation, Isolation* and *Riskiness.* The latter two refer to feelings of isolation related to childbirth and to the extent

to which women anticipate risks for the child during delivery. Fenaroli and colleagues (2013) also found a four-factor structure of the W-DEQ, and although the four domains were named with slightly different labels than those used by Johnson and Slade (2002), the four factors were considerably similar and had a high degree of conceptual overlap. In this best evidence synthesis two dimensions of pregnancy-specific anxiety, namely *Fear of labour and delivery* and *Negative feelings towards childbirth* (corresponding to *Lack of positive anticipation* in Fenaroli and colleagues, 2013)*, were thus found to exhibit strong evidence of being psychometrically sound in assessing this specific aspect of antenatal anxiety. A third dimension (*Fear for baby's health*) showed moderate strength of evidence as, although it was identified in two studies, contrasting results were found in another study (Wijma et al., 1998).

4.4.8 Pregnancy Related Anxiety Questionnaire - Revised

This PrA measure is composed of 10 items assessing various manifestations of anxiety related to a current pregnancy. Each item has 5 response options ranging from *"never"* to *"very often"*. Its original version (PRAQ: Van den Bergh, 1991) consisted of 58 items and was developed based on previous anxiety measures. The first study testing the psychometric properties of the PRAQ was carried out by Huizink and colleagues (2004) who initially tested a revised, 34-item version (PRAQ-R: Huizink et al., 2004) of the original PRAQ on 230 nulliparous women. The authors' aim was to examine the factorial structure of the PRAQ-R and test the hypothesis that PrA could be differentiated from general anxiety by comparing STAI and PRAQ-R scores. They found that only between 8 and 27% of the PRAQ-R variance was accounted for by the index of general anxiety at different time-points during pregnancy, with no linear association found between the two measures. This was interpreted as evidence of the distinctiveness of the PrA construct (Huizink et al., 2004) and highlighted once more that measures of general anxiety cannot be accurately used to identify women experiencing fears and worries specific to pregnancy. The authors initially conducted EFA and removed a number of items because of high error variance, resulting in a final version comprising 10 items (PRAQ-R). A subsequent CFA revealed that a solution with three factors provided the best fit to the data, with the three identified factors labelled by the researchers *"Fear of giving birth"* (three items), *"Fear of bearing a physically or mentally handicapped child"* (four items) and *"Concern about one's appearance"* (three items). All

individual items loaded on one of the factors above the cut-off of 0.63, except for two items, *"I am worried about not being able to control myself during labour and fear that I will scream"* and an item related to concerns about one's appearance. Similarly to the approach used for the W-DEQ and discussed above, the whole factors were considered as anxiety domains rather than individual items. Two further studies (Westerneng et al., 2015, Huizink et al., 2016) included here tested the measurement properties of the PRAQ-R, and both replicated the previous finding of a three-factor structure of the PRAQ-R by means of CFA. As the original sample of the 10-item PRAQ-R was exclusively composed of nulliparous women, Westerneng and colleagues (2015) aimed to test the factorial stability of the three-factor solution of the PRAQ-R (Huizink et al., 2004) on a large (n>6000) dataset of both nulliparous and parous women. This involved the deletion of item 8 *"I am anxious about the delivery because I have never experienced one before"*, obviously not suitable for use with women who had already experienced childbirth. CFA confirmed the same three-factor structure of the original 10-item PRAQ-R with good indexes of fit to the data for both nulliparous and parous women. Three factors were also found in a recent study (Huizink et al., 2016) which replaced item 8 of the original PRAQ-R with the more generic *"I am anxious about the delivery"*, in order to preserve a 10-item scale while making it appropriate for all pregnant women irrespective of parity (PRAQ-R2: Huizink et al., 2016). All item loadings were once more above 0.63 (range: 0.70 - 0.93) except for two items, *"I am worried about not being able to control myself during labour and fear that I will scream"*, similarly to Huizink and colleagues (2004), and *"I sometimes think that our child will be in poor health or will be prone to illnesses"*. In summary, across the three studies examined above, eight items from the PRAQ-R were found to consistently have high loadings on one of three factors (i.e. PrA domains). These three PrA domains, namely *"Fear of giving birth"*, *"Fear of bearing a physically or mentally handicapped child"* and *"Concern about one's appearance"*, were all identified in studies of good or excellent methodological quality, thus providing strong evidence of being accurate indicators of pregnancy-specific anxiety.

4.5 Summary of key findings

This review has identified a number of anxiety items and domains from existing self-report scales with demonstrated psychometric value when used to assess symptoms of anxiety in pregnant women. Eight self-report scales were considered in the synthesis of the best available evidence presented above. One further scale located by this review (Pregnancy Anxiety Scale (PAS: Levin, 1991) was not examined at the best evidence stage as the single study reporting its psychometric properties was rated poor for its methodological quality (Levin, 1991). The key findings regarding anxiety items and domains identified as accurate indicators of antenatal anxiety are briefly summarised here. As discussed earlier, they were subsequently used (Chapter 6) to inform the generation of a proportion of candidate items for the assessment of antenatal anxiety. A complete list of all the identified anxiety items and domains is also presented in Appendix 7.

Items assessing excessive, generalised worry were found to be psychometrically sound in the antenatal period in the EPDS, HADS-A, BMWS, GAD-7 and STAI-S. Overall, there was strong evidence of the psychometric robustness of items measuring the domain of worry, with consistent findings in multiple studies of good or excellent quality. Since excessive worry is essentially a cognitive symptom, it could be argued that it is less susceptible to the physical and physiological changes of pregnancy, and it remains thus a good indicator of problematic anxiety in pregnancy as it is in the general population. A second anxiety domain that showed good evidence of its psychometric soundness in pregnant populations concerned items tapping into symptoms of fear or panic. Feelings of fear are another important component of different anxiety disorders, including Panic Disorder, Agoraphobia, Social Anxiety Disorder and Specific Phobia (Craske et al., 2009; APA, 2013). In this review, items assessing the *Fear/Panic* domain were identified as psychometrically sound for use in pregnancy in various scales, including the HADS-A, the EPDS and several PrA scales. Other specific symptoms identified by this review showed moderate evidence of their screening ability in the assessment of antenatal anxiety. These included being excessively self-critical (EPDS, item 3), feeling upset (STAI-S, item 6) and the experience of nervous or motor tension (STAI-S, item 3). While these symptoms may not appear to be specific of anxiety disorders, these findings are in line with the well-established tripartite model of anxiety and depression proposed by Clark and Watson (1991). This model postulates that depressive and anxiety disorders share a common component of general emotional distress, and the

symptoms above can be categorised as manifestations of general distress, which can be present in both depressive and anxiety symptomatology (Clark & Watson, 1991).

In relation to anxiety symptoms specifically related to pregnancy, fear of childbirth was shown to be a good indicator of antenatal anxiety. Specifically, PrA symptoms of fear related to giving birth exhibited strong evidence of their psychometric value in the W-DEQ (several items) and the PRAQ-R (two items related to fear of childbirth). Items assessing persistent worries specifically related to pregnancy also showed good psychometric properties in the CWS, the W-DEQ and the PRAQ-R. The worries with the strongest evidence to support their screening accuracy related to concerns regarding the health or safety of the baby and the possibility of miscarriage. Other worries, including being in hospital and worrying about future parenting showed only moderate evidence of their screening value (see Appendix 7). It may be argued that most women are likely to experience some degree of concern regarding these aspects of pregnancy, but that in women experiencing clinical levels of anxiety these worries may be more intense or persistent (i.e. higher severity or frequency).

In conclusion, the systematic review presented here was conducted with the aim to examine the psychometric properties and content of anxiety scales used in pregnancy and map a set of core anxiety symptoms and domains identified to be psychometrically sound for the assessment of anxiety during pregnancy. Both this systematic review and the qualitative interviews with women with experience of problematic anxiety during pregnancy presented in the following chapter contributed to the initial phase of scale development, as detailed in Chapter 3. Evidence from these two studies was subsequently considered in combination to inform the construct definition of antenatal anxiety and guide the generation of an initial item pool to reflect the proposed construct.

Chapter 5   Women's perspective: qualitative interviews on the experience of anxiety symptoms in pregnancy

5.1 Introduction

The aim of the study presented in this chapter was to explore the experience of problematic anxiety symptoms in pregnant women in order to determine which anxiety symptoms can be used for the accurate identification of antenatal anxiety. This was achieved by conducting semi-structured, qualitative interviews with women with a current or past experience of clinically significant anxiety during pregnancy, who were selected based on predetermined criteria.

Similarly to the systematic review presented in Chapter 4, this study thus contributed to answer Research Questions 1 and 2:

- *What should a construct definition of antenatal anxiety include in order to cover the core domains of significant anxiety symptoms in pregnancy?*

- *Which items are the most appropriate to operationalise the proposed construct of antenatal anxiety into a self-report rating scale?*

While planning the design of this study, I carried out a scoping search of the literature to examine which qualitative studies had previously been conducted to investigate the experience of anxiety during pregnancy. A meta-synthesis of the qualitative research literature on maternal antenatal psychological distress was recently carried out by Staneva and Bogossian (2015). The authors located eight studies, with the majority investigating the experience of antenatal depression. Only two studies were found which focused on antenatal anxiety, and neither examined the full range of anxiety symptoms that can be experienced during pregnancy. Firstly, a Swedish study conducted in 2006 specifically explored the psychological experience of fear of childbirth, using a grounded theory approach to investigate how intense fear related to birth was experienced, as well as women's coping mechanisms and communication strategies (Eriksson, Jansson, & Hamberg, 2006). The second study, which also adopted a grounded theory approach, was carried out in rural

Cambodia (MacLellan, 2010) and aimed to examine the co-occurrence of depressive and anxiety symptoms in the antenatal period, as well as the barriers in accessing midwifery care in this population. While these investigations provided some level of insight into the experience of emotional and psychological distress that may occur during pregnancy, the study presented here specifically aimed to conduct a focused exploration and analysis of the range and relative importance (used with the meaning of impact, significance) of problematic anxiety symptoms that women can experience during pregnancy.

5.2 Method

When considering the most appropriate research method to investigate clinically significant anxiety symptoms experienced by pregnant women, it was clear since the early phases of study design that a qualitative method of inquiry was most suitable. Despite the diverse set of approaches that exist in qualitative research, they all tend to be characterised by an emphasis on the exploration of individual experiences or views in relation to one or more specific topics, and typically provide rich and nuanced accounts of the phenomenon under investigation (Silverman, 2001; Bowling, 2014). Within the realm of qualitative research methods, each with its own epistemological assumptions and methodological techniques, different approaches are used to answer different research questions (Mason, 2009). Given the primary aim of the research documented in this thesis (i.e. the formulation of items for the accurate assessment of antenatal anxiety), the focus of this study was less on the meanings that women attributed to their anxiety symptoms and how these were constructed, and more on the explicit description of these symptoms as experienced and reported by study participants. This was an initial indication that more interpretative approaches such as grounded theory or interpretative phenomenology were likely to be less suited to address the specific research questions of this study (Russell & Gery, 2010). It was considered that individual interviews or focus groups represented the two best candidates for data collection method, since they can both provide information-rich, detailed accounts of a specific object of inquiry by consulting a relatively small number of key informants. Individual, semi-structured interviews, which were analysed using inductive thematic analysis, were eventually favoured over focus groups for several reasons, briefly discussed here. First, it was considered that individual interviews provided a 'safer' environment for the discussion of a personal and potentially sensitive topic (Flick, 2002). Although it has been suggested that focus groups, because of the interactive nature of the technique, may also encourage and facilitate discussion on sensitive or delicate topics (Mason, 2009), other scholars have noted that in focus groups the conversation may be dominated by a small number of individuals, with the risk of loss of information, particularly if sensitive topics are discussed (Patton, 2002). It may also be argued that focus groups are a more efficient way of collecting data from a given number of individuals within a relatively short amount of time. However, in a focus group setting the detailed and nuanced accounts of individual experiences that can be obtained through individual interviews are arguably less likely to emerge. Individual interviews also typically have the benefit of providing more flexibility and choice for study

participants with regard to practical aspects such as the location or date and time of the day, which might in turn facilitate participation in the research (Bowling, 2014).

The following sections discuss a number of methodological considerations and decisions taken in relation to the recruitment of study participants, sampling and sample size, study eligibility criteria, as well as data collection and analysis. The remainder of the chapter presents the findings of this study in relation to the anxiety symptoms and domains that were most commonly reported by women with experience of problematic antenatal anxiety.

5.2.1 Recruitment of study participants

Study participants were recruited through two main routes, specifically the Maternal Mental Health Scotland Change Agents and the Perinatal Mental Health Service in NHS Greater Glasgow & Clyde (NHS GG&C). Maternal Mental Health Scotland (MMHS) is a forum of health professionals and women with experience of perinatal mental illness who work to champion the cause of maternal mental health and support research in this area. The MMHS Change Agents are a group of mothers with lived experience of perinatal mental health problems, who campaign for better services in Scotland and actively challenge the stigma associated with poor mental health during the perinatal period. They had been informed of this study in its early phases and communicated their interest in contributing to the research. The MMHS Change Agents coordinator agreed to facilitate recruitment among the Change Agents according to the recruitment procedure and study eligibility criteria discussed below. The remaining proportion of women who were interviewed were recruited through the Perinatal Mental Health Service in NHS GG&C. In this instance, recruitment was facilitated by a Nurse Consultant in Perinatal Mental Health for NHS GG&C, who is also a member of the Scottish National Managed Clinical Network on perinatal mental health. She coordinated recruitment of pregnant women currently or previously in treatment with the NHS GG&C Perinatal Mental Health Service who had a diagnosed anxiety disorder in the antenatal period.

All potential participants were initially approached either by the MMHS Change Agents coordinator or by the Nurse Consultant, who gauged the interest and suitability of women to be interviewed. This indirect approach to recruitment had the additional benefit of reducing the possibility of a woman feeling obliged to take part in the study (Wilson, Draper. & Ives, 2008). A meeting took place between me and the two recruiters prior to any potential

participants being approached, and study packs for the initial recruitment phase were distributed to them. During the initial approach, all women were given a brief, verbal explanation of the study and what it involved. If a woman expressed interest in taking part in the study, she was given a study pack containing a letter of invitation, the study information sheet, a consent form and a reply slip (see Appendices 8 and 9 for examples), and was asked to read the information provided in her own time. The invitation letter asked women who decided to take part in the study to return the reply slip to the study office (NMHAP Research Unit, University of Stirling) in a pre-paid, addressed envelope. Return of the reply slip allowed me to contact potential participants through their preferred contact method (i.e. phone or email, as indicated in the reply slip). At this stage, women interested in taking part were given the opportunity to ask questions about the study and what it involved and, if they agreed in principle to participate, a suitable date and location for the interview was agreed (further details in 5.2.3). Women could decide to be interviewed in their homes, or at a local NHS or university site (the Mother & Baby Unit within the Leverndale Hospital in Glasgow and Glasgow Caledonian University), where I had access to an interview room. The option of a video interview (via Skype) was also available. Consent to be interviewed was taken on the day of the interview when participants signed the consent form prior to the interview starting.

5.2.2 Study inclusion and exclusion criteria

The target sample consisted of women with a current or past experience of problematic anxiety symptoms in pregnancy. Specifically, the following inclusion and exclusion criteria were applied to determine study eligibility:

Inclusion criteria

- A current or past diagnosis of at least one anxiety disorder during pregnancy, as per ICD-10 criteria[1] (WHO, 1992).

---

[1] These diagnoses were made by consultant psychiatrists or clinical psychologists working for the NHS GG&C Perinatal Mental Health Service, where a proportion of study participants were recruited, as discussed earlier. Women experiencing any of the following anxiety disorders were eligible for inclusion: Agoraphobia, Panic disorder, Social phobias, Specific phobias, Generalised anxiety disorder, Obsessive-compulsive disorder, Posttraumatic stress disorder and Unspecified anxiety disorder.

- A further inclusion criterion was 'self-identified antenatal anxiety'. This criterion was applied independently and not in addition to a formal diagnosis, in order to benefit from the contributions of the lived experiences of MMHS Change Agents, some of whom had experienced significant anxiety during pregnancy but may have not been formally diagnosed. Similarly to well-established diagnostic criteria for anxiety disorders (WHO, 1992; APA, 2013), for the purpose of this study 'self-identified antenatal anxiety' was defined as anxiety symptoms experienced during pregnancy that a woman recognises as problematic and that interfere significantly with her normal routine, occupational functioning or social activities.

Exclusion criteria

- Women who were diagnosed with a severe mental health disorder other than anxiety during pregnancy (e.g. bipolar disorder, psychotic disorder).

- Women with a level of spoken English gauged as not adequate to provide sufficient details about their experience of anxiety in pregnancy.

- Women with no experience of an anxiety disorder or 'self-identified antenatal anxiety' during pregnancy.

In order to determine whether the inclusion criterion of 'self-identified antenatal anxiety' was met by women who did not have a diagnosis of an anxiety disorder during pregnancy, I asked all potential recruits (by phone or email) to rate on a scale from 1 (*not at all*) to 10 (*very much*) how much the definition of 'self-identified antenatal anxiety' presented above reflected their experience of anxiety symptoms during pregnancy, explaining that this was part of the inclusion criteria for the study. The question was slightly adapted depending on whether a woman had a present or past experience of problematic anxiety during pregnancy. In consultation with my supervisory team, it was agreed that only women who reported a self-identified antenatal anxiety score of 7 or above were eligible to be interviewed about their experience.

5.2.3 Sampling and sample size

Due to the nature and the target sample of this qualitative study, a convenience sampling technique was employed. Only participants who were likely to have experienced a past or

current episode of clinically significant anxiety in pregnancy were approached in the initial recruitment phase.

In relation to sample size, the qualitative research literature often cites data saturation as a guiding principle (Glaser & Strauss, 1967; Mason, 2009). Data saturation is considered to occur when no new concepts or themes appear to emerge from the data, and further data collection is thus considered unlikely to provide additional significant insights into the object of inquiry (Francis et al., 2009). It follows that a precise sample size in qualitative research is commonly not determined a priori. An estimated sample size was nonetheless required by the local Research Ethics Committee and the Research & Development service of NHS GG&C (Appendices 1 and 2). It was thus estimated that between 10 and 15 women, depending on data saturation, were going to be interviewed. The choice for this expected sample size was based on a combination of methodological and ethical considerations. First, considering the relatively homogeneous sample (i.e. women with experience of problematic anxiety during pregnancy) and the specific area that was going to be explored (symptoms of anxiety), it was considered that increasing the sample size to more than 15 interviewees was unlikely to result in further key insights into the experience of anxiety in pregnancy (Guest, Bunce, & Johnson., 2006). Furthermore, as previously noted, it had to be acknowledged that the content of these interviews was potentially distressing for participants. Thus, from both an ethical and a methodological perspective, it was considered important to keep to a minimum the number of women interviewed in relation to this sensitive topic while recruiting a sufficient number of participants to obtain an adequately rich and varied data set. The principle of data saturation was nonetheless applied by monitoring the percentage of new themes which were identified at regular intervals during the data collection phase. This process is further detailed in 5.3.

5.2.4 Data collection

Individual, semi-structured interviews were chosen as the data collection method to gather information on women's experiences of antenatal anxiety. The semi-structured format in qualitative interviewing is, by definition, flexible in nature (Bowling, 2014). The interviewer typically makes use of an interview or topic guide, which contains a relatively small number of open-ended questions related to key topic areas that the researcher intends to explore. The guide is, however, used flexibly, for example by altering the order of questions or by following up content areas that were not initially included in the interview guide. It thus

allows the investigation of pre-defined topics as well as the exploration of unanticipated themes (Edwards & Holland, 2013). As can be seen in Appendix 10, the main section of the interview guide consisted of open-ended questions aimed to explore a range of affective, cognitive, behavioural and somatic symptoms experienced during pregnancy by the interviewees. During the interviews, I also used various probes and follow-up questions (e.g. *"Can you tell me a bit more about that?"*) to further explore specific areas of content, as well as empathy statements (e.g. *"That must have been really hard for you"*) when I felt that it was not appropriate to simply move on to the following question.

As previously noted (5.2.1), a choice of location was offered to women who agreed to participate in the study. The majority of interviews were conducted at participants' homes, as further detailed in section 5.3. Before starting the interview, women had a further opportunity to ask questions and agree or decline to be interviewed. Women who confirmed their willingness to participate were asked to complete and sign the consent form on the day of the interview. I aimed to keep the duration of interviews to a maximum of 40-45 minutes. An audio-recording device (Olympus WS-853 recorder) was used to record all the interviews. Once each interview was completed, the recording was securely sent to an approved transcriber, which performed *ad verbatim* transcriptions of all interviews. Confidentiality agreements between the transcriber and the University of Stirling were already in place. Basic socio-demographic and obstetric information was also gathered from all study participants at the end of the interview, and is also presented in section 5.3. While the 'self-identified antenatal anxiety' score was used to determine study eligibility, at the end of each interview I also asked women to rate their 'self-identified antenatal depression' during the pregnancy discussed, using the same 1-10 scale discussed earlier. The rationale for this was to gauge whether self-rated levels of depression indicated a possible comorbidity (i.e. score of 7 or above for both antenatal anxiety and depression) or if participants ascribed their psychological symptoms mainly or entirely to anxiety (i.e. score below 7 for antenatal depression). All interview transcripts were imported into NVivo (version 11), a computer assisted qualitative data analysis software package which was used to support the process of data analysis. Ethical considerations related to interviewing pregnant women, some of whom may have still been experiencing poor mental health at the time of the interview, were discussed in Chapter 3.

A note on reflexivity

In preparation for this qualitative study, I attended two one-day training courses provided by the Social Research Association, specifically 'Designing a qualitative study' and 'Qualitative interviewing'. These courses, which proved particularly useful in practising my interviewing skills before the data collection phase, also made me aware of the important role played by reflexivity in qualitative research. Reflexivity can be defined as the act of critically considering how knowledge in research is generated taking into account the role of the investigator, including her or his biases, beliefs and personal characteristics (Guillemin & Gillam, 2004). Reflexivity contributes to increase the credibility of study findings, through a monitoring process of how the factors listed above can affect the research process, in particular in relation to the data collection and analysis phases (Berger, 2015). With regard to the phenomenon under investigation in this qualitative study, I was aware that I approached the study with some previous knowledge of antenatal anxiety, which I developed while reviewing the literature of interest in the early phases of the PhD. In part I used this knowledge to design the interview guide, by including a list of anxiety domains that I intended to discuss with study participants. At the same time, while conducting the interviews I paid particular attention to any new or unexpected content areas reported by women and tried to ensure that my pre-conceived ideas and assumptions about antenatal anxiety, and pregnancy more generally, did not constitute a barrier to the exploration of unanticipated themes. I was also aware that, as a male interviewer exploring the experience of anxiety symptoms in pregnant women, some interviewees may have felt uncomfortable regarding this aspect. I thus made sure that I informed all potential participants at the initial contact via phone or email of this, so that they could make an informed decision about whether to participate in the study taking also this element into account. Additionally, as the interviews involved the exploration of physical, cognitive, affective and behavioural symptoms occurring during pregnancy, I considered important to become familiar with the natural physiological and emotional changes which typically occur during pregnancy, in order to ask relevant questions and correctly interpret women's accounts. Consequently, I familiarised myself with the topic of interest by consulting specific textbooks and pregnancy websites. Reflexivity was also considered during the phase of data analysis, as detailed briefly in the following section.

5.2.5 Data analysis

Inductive thematic analysis was used to examine and analyse the data, following the well-established and widely used approach for the analysis of qualitative data in psychology proposed by Braun and Clarke (2006). In the initial phase of study design the options of thematic, content and framework analysis were all considered. These approaches to data analysis arguably share more similarities than differences, including their focus on the examination of recurrent themes (or segments of text) and their iterative nature throughout the analytic process (Guest et al., 2012). The six-step approach to thematic analysis proposed by Braun and Clarke (2006), discussed further below, provided a clear but flexible framework to guide the different phases of the analysis and was considered suitable for the specific aim of this study. It could be argued, however, that framework or content analysis would have produced relatively comparable findings.

The process of data analysis partially overlapped with the data collection phase, which is not uncommon in qualitative research and is deemed to inform and enhance the research process (Silverman, 2001). In this study, ongoing analysis during the data collection phase served two main purposes. First, I reviewed the field notes that I took following the conclusion of each interview, as well as the corresponding transcript as soon as available (generally three to five days after each interview), and noted any content areas that were reported and that I had not initially considered to explore. This enabled me to incorporate in subsequent interviews questions around these specific content areas, so that they could be investigated further. Secondly, the cyclical process of data collection and analysis was important to establish if and when data saturation was reached. A brief overview of the six phases of data analysis as recommended by Braun and Clarke (2006) is presented in Table 7.

Table 7 – The six phases of thematic analysis, adapted from Braun and Clarke (2006)

Phase	Key tasks
Familiarisation with the data	*Data transcription, reading transcripts several times, writing down initial ideas*
Generation of initial codes	*Assigning codes to all relevant segments of the data set, clustering of similar data around specific codes*
Identification of themes	*Merging codes into potential themes, collating all data relevant to each individual theme*
Revision of themes	*Reviewing themes for consistency and conceptual meaning. In this phase themes can be merged or subdivided into sub-themes*
Defining themes	*Labelling of themes and formulation of a brief definition; further refinement of themes and sub-themes*
Reporting the findings	*Selection of relevant, meaningful excerpts to represent most relevant themes; final discussion of key themes in relation to the research question*

Here, it is important to briefly discuss a number of key methodological decisions that were made while planning data analysis or in its initial phases. As researchers' judgement in relation to specific methodological choices (e.g. inductive versus deductive approach; when does data saturation occur, etc.) necessarily affects the process of data analysis in qualitative research, these decisions should always be made explicit in order to guarantee the rigour of the analytical process and the credibility of study findings (Mason, 2009)

Inductive vs deductive approach: The analysis was largely inductive, as conceptual categories (i.e. themes) were developed based on the data set rather than established a priori based on a theoretical model and subsequently applied to the text (Flick, 2002). While a predefined coding frame was not used, it is important to note that even more inductive, exploratory approaches to data analysis are not entirely a-theoretical (Patton, 2002), as researchers almost invariably approach the data with some previous knowledge or belief about the topic of interest (see also note on reflexivity).

The coding process: Coding essentially refers to the assignment of category labels (i.e. descriptive units of meaning typically consisting of one or a few words) to all segments of texts within the raw data that contain information potentially relevant to the research question (Flick, 2002). In thematic analysis, coding supports the analytical process by systematically reducing the complexity of the entire data set through the creation of a number of conceptual categories (i.e. themes). In this study, consistent with Braun and Clarke's approach to thematic analysis (2006), preliminary codes were initially assigned to all relevant segments of text in the data set. Once the entire data set was subjected to this process

of initial coding, and all selected extracts collated into a number of distinct codes, the resulting lists of codes were re-examined. At this stage, some extracts were reassigned to a different code, while several of the codes were merged and other collapsed into multiple codes. This process of refinement and further development of conceptual categories occurred throughout the data analysis phase, and resulted in a number of higher-level anxiety domains and more specific themes (i.e. respectively broader and narrower conceptual categories) as detailed in the Findings section.

Data saturation: As noted earlier, the study obtained ethical approval to recruit up to 15 women to be interviewed. The criterion used to determine if data saturation was reached before completing 15 interviews was to stop after three consecutive interviews in which no more than 10% of new codes (i.e. codes not previously emerged) were identified from the data analysis.

Before presenting the findings from this qualitative study, it is also important to briefly describe how the data analysis phase was carried out in practice, based on the approach to thematic analysis recommended by Braun & Clarke (2006). As noted above, data analysis commenced soon after I started collecting data, when I read each transcript several times and started assigning preliminary codes to the data. Anonymised Word versions of the first three transcripts were also read by two of my supervisors (HC and MM), who also assigned preliminary codes to all relevant segments of text. A discussion took place between me and my supervisors following the independent coding of these first three interviews, and a good concordance was found in relation to the extracts identified as relevant for the analysis. However, we considered important to agree on what should constitute a relevant or meaningful segment of text (i.e. a unit of analysis) for the remaining part of the analytical process. Following a discussion based on the preliminary codes we identified, we agreed to focus the data analysis, which I conducted on the remaining nine interview transcripts, on any segment of text describing or appearing to refer to a psychological symptom. This macro-category was labelled "*any signs or symptoms that something is wrong*" as reported by study participants, and allowed a more focused approach to coding for the remaining nine transcripts. At the same time, there was also a recognition that a subset of these symptoms, which we named "*contested symptoms*", might have been due to the normal physiological and physical changes occurring during pregnancy or to a temporary reaction to a particular triggering event (e.g. a scan with uncertain results).

During the initial phases of data analysis, I also discussed with my supervisors the best possible approach for determining the strength of evidence for each of the anxiety symptoms identified through the analysis (i.e. each theme) to be a good indicator of antenatal anxiety. One of the 'counting' techniques indicated by Braun and Clarke to determine the "keyness" of themes (2006, p.82) was used. This approach consisted in calculating the number of interviews that contained at least one instance of a given symptom. Consequently, the relative importance of each single theme was determined by the prevalence of the theme at the data item level (i.e. whether a theme appeared anywhere in each individual interview). Specifically, we agreed to rate the strength of evidence in representing a relevant symptom of antenatal anxiety according to the following criteria:

- Strong evidence: *At least one instance of a given symptom is reported by more than half the sample (>50%, n > 6)*
- Moderate evidence: *At least one instance of a given symptom is reported by more than 25% and up to 50% of the sample (n = 4 - 6).*
- Limited evidence: *At least one instance of a given symptom is reported by more than 10% and up to 25% of the sample (n = 2 − 3).*

Only symptoms with moderate or strong evidence were considered as potentially important indicators of antenatal anxiety, and consequently used to generate a proportion of items for potential inclusion in the antenatal anxiety screening scale. While it can be argued that these numerical criteria are somewhat arbitrary, particularly in view of the relatively small sample, it was considered that they provided a rational and pragmatic way of categorising a relatively large data set and reaching conclusions (i.e. identifying relevant anxiety symptoms) based on a systematic principle of evaluation. Moreover, the 'counting' approach used here was not based on counting all the instances of a given theme wherever they appeared in all the interview transcripts, as more instances of a specific theme in the entire data set do not necessarily indicate that such theme is more important (Patton, 2002; Braun & Clarke, 2006). This system of evaluation was instead based on determining the number of study participants within the entire sample who had experienced a specific symptom. This was considered to provide a reliable indication of the relevance and significance of a given symptom in women experiencing antenatal anxiety.

5.3 Findings

5.3.1 Socio-demographic and obstetric characteristics of study participants

All study data were collected between October 2016 and February 2017. Twelve women were interviewed, with the majority of interviews taking place at participants' homes (66%; $n = 8$). The remaining interviews were carried out at a local NHS facility (25%; $n = 3$) and via Skype ($n = 1$). All women who returned a reply slip had a score of 7 or above in relation to 'self-identified antenatal anxiety' and were thus eligible to take part in the study. This is arguably a consequence of the sampling technique that was used to recruit potential interviewees. A summary of the socio-demographic and obstetric characteristics of study participants is presented in the next page in Table 8.

Table 8 – Socio-demographic and obstetric characteristics of the sample

Age	Current or past experience of antenatal anxiety (gestational week if current OR time since pregnancy)	Pregnancy discussed (total number of pregnancies)	Obstetric complications in previous pregnancies	Education	Self-identified antenatal anxiety score	Self-identified antenatal depression score
29	Past (3 years)	2nd (4)	1st ectopic	Undergrad degree	10	0
43	Current (32 weeks)	2nd (2)		Postgrad degree	8	6
42	Past (3 years)	3rd (3)	Two miscarriages after 2nd pregnancy	Undergrad degree	10	2
34	Past (2 years)	1st and only		Postgrad degree	7	0
35	Past (4 years)	1st (2)	1st was a traumatic delivery, resulting in diagnosed PTSD	Postgrad degree	9	0
29	Past (1 year)	1st and only		Undergrad degree	10	5
38	Current (27 weeks)	2nd (2)	1st was miscarriage	Postgrad degree	7	4
35	Current (35 weeks)	1st and only		Postgrad degree	7	5
32	Current (36 weeks)	3rd (3)		Undergrad degree	8	4
40	Past (4 years)	3rd (3)	2nd was miscarriage	Undergrad degree	8	8
27	Past (recent, less than four months)	3rd (3)	1st and 2nd were miscarriages	Postgrad degree	10	7
36	Past (1 year)	1st		Undergrad degree	8	1

The average age of participants at the time of the interview was 35. However, the average age at the time of the pregnancy discussed in the interview was 33.5 (age range 26-43). Four interviews were conducted with women who were experiencing antenatal anxiety at the time of the study, while the remaining eight interviewees had a past experience of anxiety in pregnancy. In case of multiple pregnancies, at the beginning of each interview I asked study participants to focus on the pregnancy in which they experienced the most problematic anxiety symptoms. The column 'Pregnancy discussed' indicates which pregnancy was discussed in the interview, and the total number of pregnancies for each woman. Notably, half of the sample (n = 6) had a history of obstetric complications or miscarriage in previous pregnancies. Two participants had a 'self-identified antenatal depression score' of 7 or above, indicating elevated levels of both anxiety and depressive symptoms during pregnancy. The average duration of an interview was 38 minutes (27m – 56m).

5.3.2 Data saturation and the analytical process

Data saturation was reached after twelve interviews according to the criterion specified earlier. Interviews were divided into groups of three and the percentage of new codes (i.e. codes not assigned in previous interviews) was calculated at the end of every three interviews. As shown in table 9, only approximately 6% of new codes were identified in the last three interviews.

Table 9 – Proportion of themes identified every three interviews on the total number of themes identified

	Interviews 1-3	*Interviews 4-6*	*Interviews 7-9*	*Interviews 10-12*
Percentage of codes assigned to relevant segments of text [1]	46.7%	73.3%	93.3%	100%
Percentage of <u>new</u> codes assigned to relevant segments of text	46.7%	26.6%	20%	6.7%

1 Percentages in this row were calculated once the twelve interviews were completed.

The analysis was carried out as previously detailed in the Data analysis section. The process of comparison and refinement of codes (i.e. themes) continued until the final stages of the analytical process, when thirty themes were ultimately identified. In the conclusive stages of the analytic process, the analysis moved towards broader conceptual categories, and all the identified themes (i.e. specific anxiety symptoms) were organised into five higher-order anxiety domains, broad conceptual categories which were used to capture the essence of different themes considered to share similar features (Silverman, 2001). The identified anxiety domains were: 1) *Worry and anxious apprehension* 2) *Fear* 3) *Pregnancy-related anxiety* 4) *General distress* and 5) *Anxiety-driven behaviours*. Table 10 illustrates all the identified themes and corresponding anxiety domains identified through the study.

Table 10 – Summary of all identified themes and corresponding anxiety domains

Anxiety domain	Themes identified	No of participants reporting the theme *Strength of evidence:* *Strong n > 6 participants;* *Moderate = 4-6* *Limited = 2-3*
Worry and anxious apprehension	• Excessive worry • Catastrophic thinking • Generalised worry • Feeling tense or on edge • Repetitive thoughts and rumination • Racing thoughts • Worries about dear ones' safety	11 10 10 9 7 6 3
Fear	• Feeling frightened or fearful • Physical symptoms of hyperarousal and somatic activation • Severe or uncontrollable anxiety • Feeling of impending doom • Other physical symptoms	9 9 7 5 3
Pregnancy-related anxiety	• Worried about the baby's health or safety • Fear of childbirth • Worries about future parenting • Worries about possibility of miscarriage • Fear for own health	10 8 6 6 4

General distress	• Feeling upset or distressed • Feeling overwhelmed • Feeling uneasy • Feeling unable to cope • Feeling irritable • Feeling guilty • Being self-critical • Loss of control • Feeling indecisive	10 8 8 6 6 5 5 4 4
Anxiety-driven behaviours	• Avoiding specific places or situations • Reassurance seeking • Withdrawing from people • Safety behaviours	7 5 3 2

The following section presents all the themes which were reported by more than half the sample ($n = 7$ or above), thus indicating strong evidence of their relevance as symptoms of antenatal anxiety. The number of women who reported a specific symptom is indicated in parentheses next to each theme. For reasons of brevity, extracts of themes with moderate or limited strength of evidence are not reported here. Both themes with moderate and strong evidence were subsequently used in the generation of an initial item pool, as detailed in Chapter 6. The discussion below is structured by presenting separately the five higher-level anxiety domains developed in the conclusive phases of the analytical process. Within each domain, the corresponding key themes are illustrated through selected exemplary quotes for each individual theme.

5.3.3 Anxiety domains and key themes

ANXIETY DOMAIN 1: WORRY AND ANXIOUS APPREHENSION

This anxiety domain included themes such as excessive worry, catastrophic thinking, nervous tension, repetitive thoughts and generalised worry that something bad will happen. These symptoms were all reported by the majority of women interviewed and were often characterised by anxious apprehension regarding a variety of topics, as well as excessive or pervasive worry which was generally recognised as disproportionate. While it could be argued that a certain degree of worry during pregnancy can be expected and is presumably a relatively common occurrence, what often characterised the accounts of women with experience of antenatal anxiety was the excessive level or persistence of worrying and apprehension, at times with a considerable negative impact on their quality of life. This is illustrated in the quotes from the key themes related to the anxiety domain of '*Worry and anxious apprehension*' presented below.

Excessive worry (n = 11)

One of the most common symptoms reported by study participants was excessive worry. Women recognised that the level of worry was often disproportionate and caused significant distress. The continuous worrying in certain instances significantly impacted on areas of functioning and well-being, as illustrated in these extracts:

> P5: **"***There wasn't really room for any, you know, for very much happiness to sneak in there because I was just worried all the time"*

> P7: *"Things would build up and then all of a sudden I couldn't, I would just break down and I couldn't think, I couldn't do anything and I would just, you know, I couldn't problem solve so even the simplest things were just too much to cope with and that made me even more anxious and worried"*

The pervasive nature of worry is also evident in this account:

> P11: *"And then eventually I would get a new worry so the old worry would kind of go to the back burner and the new worry would go for a while and then it would just kind of*

move on, so…. yeah, it was like a progression of things that could be worse than the next one"

Catastrophic thinking (n =10)

In numerous instances, continuous and excessive worry resulted in catastrophic thinking (Cox, 1996; Gellatly & Beck, 2016). Of note, both excessive worry and catastrophic thinking are key features of Generalised Anxiety Disorder, which is the most prevalent anxiety disorder during pregnancy (Dennis et al., 2017). Women reported thinking about worst-case scenarios and potentially disastrous consequences that could potentially result from their own actions (or inaction), as in these cases:

P9: "So just really insignificant problems that I wouldn't normally have focused on at all, I would really start to think about them, sort of, blow them all out of proportion and imagine that from one tiny little thing that it was going to get bigger and bigger and something bad was going to happen and it would be all my fault….. the things that you knew were never going to happen but you just thought 'oh well, you know, I'll be the person that this bad thing will happen to"

P11: "So if I changed the cat's litter and I didn't wash my hands thoroughly enough then I'm definitely going to get toxoplasmosis and my baby's going to be born deaf and blind"

At times this catastrophic type of thinking resulted in women being "convinced" that something terrible would happen to them.

P6 "I had back pain when I was pregnant and I was convinced it was a tumour, you know that way I'd totally convinced myself"

Generalised worry (n = 10)

This theme included feelings of generalised worry, distinct from the previous theme for the absence of specific worst-case scenarios. Here, worry was described as more pervasive and not restricted to any particular topic or situation.

P1: "I just think I felt worried a lot of the time and I just constantly had this feeling that something was going to go wrong"

Notably, women generally recognised that an overwhelming sense of worry was different from what other pregnant women experienced:

P7: "I would say with worry it's an overbearing sense of worry maybe, so I think my worry seemed that, it just seemed to take over and it was not necessarily…I would describe that as something that I think differed from me to my other, my friends that appeared to have normal pregnancies, or … have seemed to be able to kind of yes they're worried but they can get on with their life"

While conducting the analysis, I considered merging the three themes presented above (*Excessive worry; Catastrophic thinking; Generalised worry*) into a single '*Worry*' theme. However, as symptoms related to worry and anxious apprehension were among the most commonly reported, an indication of their potential significance as indicators of antenatal anxiety, it was deemed important to preserve distinct features of this component of anxiety by categorising different aspects of worrying into separate themes.

Feeling tense or on edge (n=9)

A considerable proportion of women reported feeling on edge, as shown in these instances:

P2: "Yeah, you feel very much on that sort of edge, so the stress levels or whatever you feel, like, you're more on that level so you can be tipped much quicker, haven't got the same patience"

P5 "Yeah, I mean, being on edge, being very snappy with people because I would... I was very edgy, you know, a loud noise or something I would panic and it would be out of all proportion reaction"

These feelings of apprehensive anticipation were often accompanied by physical symptoms of nervous or muscular tension, as illustrated here:

P9: I would say probably more tension in my hands and my jaw. Lots of teeth clenching as well as that, tension headaches, more towards the eyes rather than the back of the head"

P12: "But I did have a very bad back, obviously that's the carrying the child has had an impact but a lot of it was kind of tension"

Repetitive thoughts and rumination (n = 7)

The key aspects of this theme related to cognitive rumination and repetitive thoughts. A study participant discussing these symptoms described the experience as "paralysing"

P2: ""I tend to get... very stuck in thoughts, so sort of things go over and over again… for me it's almost paralysing, so instead of doing I just do a lot of thinking and a lot of inaction, so it's sort of paralysing in that respect"

The experience of constant rumination appeared to be exhausting for some women:

P6: "Oh yeah, there wasn't a day when there wasn't... there wasn't a time for just... like, my brain was just going constant"

Other themes in the anxiety domain '_Worry and anxious apprehension_' included racing thoughts (_n_ = 6) and worries about dear ones' safety (_n_ = 3), as shown in Table 9.

ANXIETY DOMAIN 2: FEAR

Feelings of fear, often accompanied by physical symptoms of hyperarousal and somatic activation (e.g. racing heart, sweaty hands) are common in phobic anxiety disorders such as Agoraphobia, Panic Disorder and Specific Phobias (APA, 2013). It was thus not surprising to find that this type of symptoms were also frequently reported by women with an experience of antenatal anxiety. The distinction between the '_Fear_' and the '_Worry and anxious apprehension_' anxiety domains is corroborated by the difference, well-documented in the research literature on anxiety disorders, between verbal-subjective symptoms of anxious apprehension typically characterised by anticipation of possible negative events on one hand, and symptoms of fearful mood, often associated with somato-visceral activation in response to a perceived imminent threat, on the other hand (Barlow, 2002: Craske et al., 2009). The extracts in the next page illustrate the key themes related to this anxiety domain.

Feeling frightened or fearful (n = 9)

This theme referred predominantly to the affective, emotional component of symptoms of Panic Disorder or phobic anxiety, which typically manifested in a generalised feeling of fear or terror.

> P12: *"Everyone was like 'oh that's brilliant, you must be so happy' and I was just, like, really frightened the whole time"*

A woman described a sense of terror when asked about the most distressing feelings she experienced during pregnancy:

> P9: *"Definitely sort of terror. And then I felt as if there was no sort of way out what I was feeling and I just imagine that I would feel like this for the rest of my life and this was the way my life was going to be and I couldn't see a time where I wasn't feeling anxious"*

Physical symptoms of hyperarousal and somatic activation (n=9)

I inspected closely physical symptoms reported by study participants. In view of the research literature on antenatal anxiety discussed in previous chapter, suggesting that somatic symptoms might be a confounding factor in the assessment of anxiety symptoms during pregnancy, it was considered crucial to determine if and which physical symptoms women tended to attribute to anxiety rather than to normal changes of pregnancy. This section does not discuss physical symptoms of nervous or muscular tension, which were previously reported in the '*Feeling tense or on edge'* theme.

First, women were generally able to distinguish bodily sensations that they ascribed to pregnancy from other physical symptoms which were attributed to anxiety, as in this extract:

> P11: *"I was getting the physical symptoms so typically if I was having these thoughts I would feel really sick, and I didn't experience any morning sickness or anything so the nausea was anxiety based"*

There were however, instances in which study participants were unsure whether somatic symptoms were pregnancy-related or anxiety-related, as for example in the case of symptoms such as heaviness in the chest, sweaty hands, and stomach issues. These symptoms, each only reported by a small number of women, were included in a separate

category of '*contested symptoms*' (as discussed in 5.3.2), and subsequently excluded from the analysis of potential indicators of antenatal anxiety.

The most common physical symptom of somatic activation that study participants ascribed to anxiety was by far a racing heart or palpitations, as described here:

P1: ""Yeah, just like the, you know, really fast heartbeat and sweaty hands and just that feeling of impending doom"

P3: "palpitations were definitely... I think were definitely due to that. I mean, obviously you get stomach issues anyway, you get heartburn and all these sorts of things, but I didn't actually really have that too bad with the first two pregnancies, but with this one I did, so..."

P7: "Yeah my heart is definitely, not necessarily my breathing but I do notice a racing heart"

There were a number of cases in which palpitations or an excessively fast heartbeat were linked to panic attacks, as illustrated in the next theme.

Severe or uncontrollable anxiety (n = 7)

This theme related to accounts of extremely heightened anxiety, typically perceived as distressing and uncontrollable. It is distinct from the '*Feeling frightened or fearful'* theme as extracts included in this theme essentially describe anxiety symptoms at their peak, as in the case of a panic attack or other acute anxiety symptoms. Five women reported experiencing panic attacks during pregnancy (none with a previous history of panic attacks), as illustrated here:

P3: ""I was having panic attacks just… yeah, you know, that was just normal kinda thing"

P6 : "I wouldn't sleep, would have nightmares... like, my husband would find me at three o'clock in the morning at the bottom of the bed having panic attacks"

In other cases, women did not experience full-blown panic attacks, but found it particularly difficult to control feelings of anxiety, as shown here:

P2: it was a bit like when I moved house I had a sort of total meltdown, was totally anxious, worried about whether I'd made the right decision, everything kind of escalated

really to a point where it was... my thoughts were just really dire, you know, to the point of, you know, 'I don't want this baby anymore, why am I pregnant, this was a stupid idea'"

P11: "So my anxiety had been gradually increasing but it wasn't until… about 32 weeks...that point that it was just off the scale"

As evident in these extracts, the fact that anxiety was perceived as extremely heightened or uncontrollable could result in significant distress. Other themes categorised within the *'Fear'* domain included a feeling of impending doom (*n = 5*) and other physical symptoms (*n = 3*).

ANXIETY DOMAIN 3: PREGNANCY-RELATED ANXIETY

Pregnancy-related anxiety was extensively discussed in Chapter 2. The themes identified in this study relating to concerns and fears specific to a current pregnancy appeared to reflect largely those described in the research literature on pregnancy-related anxiety discussed earlier and the symptoms identified through the systematic review presented in the previous chapter. Fear of childbirth and worries related to the health or safety of the unborn baby were the two key themes of pregnancy-related anxiety experienced by the majority of women who participated in the study. Persistent or intense worries around the possibility of miscarriage were reported by half of women in the sample.

<u>*Worried about the baby's health or safety*</u> *(n =10)*

This was the most common worry reported by women within the anxiety domain of pregnancy-related anxiety. Once more, it can be reasonable to expect that worries around the health or safety of the unborn baby are considerably common during pregnancy. However, as in previous instances regarding general anxiety symptoms, the problematic aspect appeared to be related to the intensity or frequency of the symptoms. A study participant who was asked which aspects of anxiety during pregnancy was the most distressful for her commented:

P11: "Thoughts that the baby wasn't going to be well were really distressing. The typical anxiety symptoms that I had were seeking reassurance, so I was saying to my husband 'oh I've not felt kicking in a couple of hours' and obviously he would have to kinda talk me down and say 'drink some water, lie on your side' do all that typical stuff, but yeah I was really conscious, probably more so of watching myself for movements and really focused in

on looking for movements, and obviously sometimes the more you focus on something the more it doesn't happen."

Some women reported persistent feelings of worry and tension which were characterised by intense anxiety, as in this instance:

P3: ""you know, I literally felt like I was spending my day just holding my breath, you know, and every day was like 'right, we've got through another day with this baby still alive and kicking' kinda thing, you know, that was the stage that I got to"

For others, the feeling was perhaps less intense but still described as "really uncomfortable":

P12: ""It was something I couldn't put my finger on, I just felt really uncomfortable but I did have worries about my daughter, I was sure that something was going to be wrong with her, that she was going to have a condition of some sort or wouldn't reach term"

Fear of childbirth (N = 8)

As previously discussed in other chapters, fear of childbirth experienced over the course of pregnancy is one of the most commonly reported symptoms of pregnancy-related anxiety. For the analytic purpose of this study, it was important to identify characteristics (e.g. intensity, frequency, specific focus) of fear of childbirth that could be used to distinguish it from the arguably much more common but less problematic worries that can be experienced regarding labour and delivery.

Some women reported a specific fear that something bad would happen to the baby during childbirth, which understandably was a considerable source of distress:

P7: "I think of all the things that could happen during the birth the baby could, you know, so... I just can't seem to not think of the bad things and I can't seem to picture, these bad thoughts I can seem to or images if you like, I can't seem to picture a happy one."

A woman with a previous experience of a particularly traumatic childbirth (which resulted in diagnosed Posttraumatic Stress Disorder), described how she felt "petrified" in the current pregnancy about the prospect of giving birth again.

P5: "I am petrified about actually giving birth this time and I'm a lot less anxious about, you know, just the day to day pregnant woman"

Other themes included in the '*Pregnancy-related anxiety*' domain were worries around future parenting (*n = 6*), worry about the possibility of miscarriage (*n = 6*) and fear for own health (*n = 4*)

ANXIETY DOMAIN 4: GENERAL DISTRESS

There is a vast research literature documenting the relatively common occurrence of comorbid anxiety and depressive symptomatology, both in the general population and during the antenatal period (Sanderson, Beck & Beck, 1990; Lamers et al., 2011). Most of the symptoms reported by women in this study were clearly related to anxiety symptomatology, as detailed in the previous sections. There were, however, several other symptoms reported by women such as feeling guilty, upset, overwhelmed, or a general sense of being unable to cope, which could not be categorically ascribed to symptoms of an anxiety disorder (or pregnancy-related anxiety). These symptoms of general distress were included as one of the domains potentially indicative of antenatal anxiety, based on the well-established tripartite model of anxiety and depression by Clark and Watson briefly discussed in Chapter 4 (1991). In their influential model the authors proposed that while depression and anxiety are characterised by distinct features (i.e. anhedonia and absence of positive affect in depression; anxious apprehension and physiological tension in anxiety disorders), they also share a non-specific component of *General distress*, which can include symptoms such as feeling uneasy, irritable, guilty or overwhelmed. The rationale for including symptoms of general distress here (if they were reported by a sufficient number of women) is based on two considerations. In the first place, as noted above coexistent depression and anxiety is common in the antenatal period. It can thus be expected that some women experiencing problematic antenatal anxiety will also suffer from symptoms of general distress. Secondly, the tripartite view of anxiety and depression implies that an accurate assessment of either depressive or anxious symptomatology requires the consideration of both the unique and the common elements of the syndromes (Clark & Watson, 1991). The key themes related to the '*General distress*' domain which were reported by more than half of women are illustrated in the next pages.

Feeling upset or distressed (n = 10)

This theme was initially named *'Feeling distressed'* in the phase of preliminary coding. However, at a closer inspection of the extracts included in this code it appeared clear that the term '*upset'* was also a useful descriptor of the content of this theme. The following extract exemplify this theme:

P7: "I was hysterically crying and I'm not actually someone who cries a lot… So I found that very distressing as well, crying a lot, that upset me even more which didn't help, which spiralled more."

In some instances, study participants felt distressed in relation to specific situations:

P11: "Social situations, my cousin's baby shower was really difficult because there was loads of people there and because I was pregnant too people were making comparisons and I found that comparisons between us, like, 'oh my god you're so much bigger than her', that was really quite stressful. People that have never before or never would before comment on your body and how you look."

Another woman described becoming very upset and feeling that her reactions were excessive:

P11: "Yeah, I would get quite hysterical, excessive crying and I know that crying obviously when your hormones are changing and things like that happens, but it was excessive."

Feeling overwhelmed (n = 8)

A sense of feeling generally overwhelmed was often reported by women with experience of antenatal anxiety. This extract illustrates how even a simple choice such as deciding which friend to call could lead a study participant to feel overwhelmed:

P2: "it's like 'well what am I going to do and who do I ring?' I have got lots of friends but that in itself becomes an anxiety of 'who do I ring, who should I go and see, should I go and do that or should I go and do that?' so it's just a bombardment of all these thoughts"

A sense of mental exhaustion was also evident in these two accounts:

P6: "But, like, you don't, everybody just expects you to be so happy and 'oh you're pregnant and it's wonderful' and I just hated every single second of it, just cause of the way it made me feel, I mean, I was fine with him, I wasn't sick, I wasn't... didn't feel generally unwell, it was more a kinda mental... like, an emotional tiredness, mental tiredness"

P10: "Yeah, I had loads, loads of different feelings, you know, fear, anxiety, panic, being overwhelmed, like, tiredness cause I was so shattered because obviously your mind's going into overdrive. There was just so much going on it was horrible."

Feeling uneasy (n = 8)

A considerable proportion of study participants also reported a persistence sense of unease, as in this example:

P9: "I mean, even at a time where I wasn't worrying about something I would have this general sense of unease because I was waiting for the next thing"

Remarkably similar words were used by another woman to describe how she generally felt:

P1: "I also had just a general feeling of unease"

Other themes included in the domain of *'General distress'* included feeling unable to cope (*n = 6*), feeling irritable or snappy (*n = 6*), feeling guilty (*n = 5*), being self-critical (*n = 5*), loss of control (*n = 4*) and feeling indecisive (*n = 4*).

ANXIETY DOMAIN 5: ANXIETY-DRIVEN BEHAVIOURS

The final anxiety domain presented here comprised a range of behaviours that women used in the hope of reducing or keeping under control anxiety levels. However, only one theme was reported by more than half of the women in the sample, and consequently only extracts from this theme are presented here. Other anxiety-driven behaviours reported by women are listed in conclusion of this section.

Avoiding specific places or situations (n = 7)

Behavioural avoidance is a common characteristic of many anxiety disorders (e.g. Panic Disorder, Social Phobia). Social situations appeared to be particularly anxiety-provoking for some of the study participants and at times entirely avoided, as illustrated in these two extracts:

P8: "you know, I stopped... I also stopped, like, my social circle and for me, like, you know, I'm always the person who's involved with the parent council at schools, I'm the one who volunteers for things, d'you know, so it was all of that kinda went as well, so it was a real sense of kinda identity loss in that way"

P6: "I kinda stopped going out to like shopping centres and into town, like, into busy places, I couldn't... I just couldn't cope with people round me"

In other instances, women avoided specific places as they anticipated that they could trigger distressing levels of anxiety. This is exemplified by this account:

P6: "I couldn't even go to the shops and stuff, like, even to do a supermarket shop, I used to have to leave cause I would get so overwhelmed, everything was just... totally, like, going to the shops, that's something you know... you nip in and out all the time don't you, don't even think about it, I couldn't even do a full shop"

Other anxiety-driven behaviours included themes such as reassurance seeking ($n = 5$), withdrawing from people ($n = 3$) and use of safety behaviours ($n = 2$)

5.4 Conclusion

The interviews described in this chapter explored the experience of problematic anxiety symptoms during pregnancy, as well as several factors useful to differentiate between normal experiences of anxiety and worries in pregnancy as opposed to elevated levels of anxiety which would indicate the need for specific support. A range of anxiety symptoms, as discussed in detail above, were identified as potentially important indicators of antenatal anxiety. A considerable overlap was found among the affective, cognitive, behavioural and somatic content areas of the construct of antenatal anxiety identified through the systematic review presented earlier and the qualitative interviews discussed here. Virtually all symptoms and anxiety domains identified in the systematic review could be categorised within one of the five higher-order anxiety domains which were used to describe the findings of the qualitative interviews. Evidence from these two studies was subsequently combined, using predefined criteria to evaluate the strength of evidence for each content area to be an important domain of the target construct. The following chapter presents a conceptual and an operational definition of the construct of antenatal anxiety, based on the findings from these two studies, as well as the process of generation of an initial item pool to reflect the proposed construct. As noted in Chapter 3, this phase of item generation occurred simultaneously with a number of decisions that were taken in relation to the format of individual items, as well as the overall scale. This process is also detailed in the next chapter.

Chapter 6 – Construct definition, generation of an item pool and preliminary item reduction

6.1 Introduction

This chapter describes Phase 1 and the initial stages of Phase 2 of the research, as outlined in Chapter 3 and in Figure 3 below. Phase 1 was informed by the studies presented in the previous two chapters and consisted of developing a construct definition of antenatal anxiety (Section 6.2) and generating an initial pool of candidate items to reflect this target construct (Section 6.3). In the first half of this chapter, the outcomes of Phase 1 are presented. During the process of item generation, a number of decisions were also made in relation to the format of measurement of the scale. These important considerations, which directly informed item writing, are also discussed in section 6.3. In the second part of the chapter, the initial stages of Phase 2 are presented. These consisted of preliminary item refinement based on inputs from women with experience of perinatal mental health problems (MMHS Change Agents) and experts in the field (Section 6.3.3); and the stage of initial item reduction via a Delphi study with experts in the area of perinatal mental health (Section 6.4).

Figure 3 – A reminder of the phases and stages of the research

PHASE 1 Scale development	• Construct definition of antenatal anxiety • Generation of initial item pool • Decisions on format of measurement
PHASE 2 Scale refinement	• Consultations with key informants • Item reduction and refinement • Preliminary psychometric testing and further item reduction
PHASE 3 Psychometric validation	• Psychometric validation of the screening accuracy (criterion validity), internal consistency, construct validity and factor structure of the final version of the scale

6.2 Conceptual and operational definitions of the construct of antenatal anxiety

In this section, a conceptual, formal definition of the construct of antenatal anxiety is initially provided. A more detailed, operational definition, which was instrumental to the generation of an initial item pool, is subsequently also presented. This latter operational definition, in which the different content areas of the construct are delineated, should be considered as a specification of the conceptual definition rather than a distinct description of the construct. As pointed out earlier in this thesis, a thorough review of the literature related to the construct of interest and interviews with the target population are two commonly recommended sources of evidence that can be used to inform construct definition and item generation (Turner et al., 2007; DeVellis, 2012). In particular, theoretical and empirical evidence from these sources can provide the basis for articulating the conceptual boundaries of the construct, clarifying its key content domains, and ultimately contribute to enhance different forms of scale validity, such as face, content, and construct validity (Clark & Watson, 1995; Netemeyer et al., 2003; Kline, 2005; DeVellis, 2012). In this research, evidence from both the systematic review of the psychometric literature on antenatal anxiety and the qualitative interviews with women with experience of problematic anxiety symptoms during pregnancy was used to inform the conceptual and operational definitions of the construct of antenatal anxiety presented here.

As noted in conclusion of the previous chapter, a considerable overlap was found between problematic anxiety symptoms identified through these two studies. In particular, a range of symptoms which could be categorised into five broad symptom domains emerged as relevant indicators of antenatal anxiety. This is further discussed in 6.3 in relation to the generation of the initial item pool. Here, based on evidence from the research literature on antenatal anxiety (as per Chapter 2, 4) and on interviews with the target population (as per Chapter 5), the following conceptual definition of the construct of antenatal anxiety was proposed:

Conceptual definition of antenatal anxiety

Antenatal anxiety can be defined as the experience of clinically significant symptoms[2] of anxiety in pregnant women. The term clinically significant is used here to indicate that the symptoms are sufficiently problematic to:

A) be perceived as distressing

and/or

B) have a negative impact on at least one area of individual functioning (e.g. daily routine, social relationships, occupational functioning).

Relevant symptom domains of the construct of antenatal anxiety are: *Worry and anxious apprehension*, *Fear*, *Pregnancy-related anxiety*, *General distress*, and *Anxiety-driven behaviours*. Antenatal anxiety can manifest as the experience of symptoms in one or more of these domains. A further specification is that, in order to qualify as antenatal anxiety:

C) symptoms must be experienced for a sufficiently prolonged period of time (i.e. not limited to a temporary reaction to a specific event or situation).

This conceptual definition of the construct of antenatal anxiety has two key implications. Firstly, the criteria A, B, and C in this definition are instrumental in operating a distinction between pregnant women who may be experiencing occasional, non-problematic worries and anxieties from pregnant women experiencing distressing and persistent symptoms of anxiety. Secondly, evidence from the qualitative interviews presented in Chapter 5 indicated that the co-occurrence of symptoms in two or more of the symptom domains listed above is not uncommon in pregnant women experiencing problematic anxiety symptoms. There were, however, also frequent instances in which symptoms specific to a single domain (e.g. *Pregnancy-related anxiety* or *Worry and anxious apprehension*) were largely predominant and nonetheless caused significant distress. Consistent with this observation, in the definition proposed, even only one of these symptom domains is sufficient to be categorised as antenatal anxiety (on condition that either or both criteria A and B, as well as criterion C, are met).

---

[2] As specified in the operational definition provided below

Operational definition of antenatal anxiety

As previously discussed in the Method chapter, in scale development the phase of construct definition is followed by the operationalisation of the construct, which essentially corresponds to the translation of the construct definition into an initial list of items that comprehensively reflect all facets of the target construct (Kline, 2005). A more thorough, detailed definition of antenatal anxiety, which specified a range of possible symptoms of the construct, was thus required in order to inform the generation of a pool of candidate items for its assessment. An operational definition of the construct was therefore proposed. This includes the formal definition presented above, and additionally a list of all the key facets (i.e. symptoms) of the construct identified as potentially accurate indicators of antenatal anxiety in the two studies presented earlier in the thesis. Specifically, the following symptoms were identified through the systematic review and the qualitative interviews as having moderate or strong evidence of being relevant indicators of antenatal anxiety in at least one of the two studies:

- Excessive or generalised worry
- Nervous or motor tension, feeling on edge
- Repetitive thoughts and rumination
- Feelings of panic or intense fear
- Uncontrollable or severe anxiety, potentially accompanied by symptoms of somatic activation (in particular a racing heart)
- Specific symptoms of general distress (e.g. feeling overwhelmed, feeling unable to cope)
- Specific symptoms of pregnancy-related anxiety (i.e. fear of childbirth, excessive worry about the baby's health or the possibility of miscarriage, anxious apprehension about future parenting, fear for own health).
- Behavioural avoidance of specific places or situations

Perhaps not surprisingly, most of the symptoms listed in this operational definition as key indicators of antenatal anxiety, based on the evidence discussed in Chapter 4 and 5, correspond to symptoms of anxiety disorders as described in the general population. This is entirely consistent with the general evidence, well-documented in the research literature, that anxiety disorders as they present in the general population are also relatively prevalent

during pregnancy (Heron et al., 2004; Marchesi et al., 2016; Dennis et al., 2017*).* This operational definition, however, also introduces some key distinctive features of the construct of antenatal anxiety. First, it includes symptoms of pregnancy-related anxiety, thus implying that anxiety scales developed for the general population cannot be reliably used for a comprehensive assessment of antenatal anxiety, as they do not include symptoms that are specific to pregnancy. Another relevant characteristic of this operational definition is the absence of physical symptoms, with the exception of a racing heart. This is also in line with the findings presented earlier, which appeared to confirm the limitations of using somatic symptoms in the assessment of antenatal anxiety, as discussed in Chapter 2.

Once an operational definition of the target construct is formulated, the subsequent stage in scale development consists in the development of an initial pool of candidate items that comprehensively reflect all facets of the construct (DeVellis, 2012). The following section discusses the process of generation of an initial item pool for the assessment of the construct of antenatal anxiety. This discussion is preceded by an overview of a number of important considerations in relation to the desired format and structure of the final scale, as well as to the process of item writing.

6.3 Formulation of an initial item pool

6.3.1 Issues and considerations related to item writing and scale format

Before proceeding to present and discuss the stage of generation of an initial pool of candidate items for the assessment of antenatal anxiety, it is important to discuss briefly a number of issues that scale developers need to consider at this stage, in particular in relation to item writing and the desired format and length of the scale. Decisions concerning these aspects of scale development are critical as they directly guide and inform the process of item generation and can also have a considerable impact on several forms of scale validity. A relatively large body of literature has examined each of these issues in great detail (Netemeyer et al., 2003; Streiner & Norman, 2008; Furr, 2011) and they have all been carefully considered as part of the work discussed in this thesis. Here, only a brief overview of these aspects of scale development is presented in order to provide a rationale for the decisions made in relation to the development of the scale documented in this thesis.

Scaling response and response format: Issues related to these aspects of scale development include those related to the choice of the scaling response (e.g. Likert scale, visual analogue scale), as well as to other characteristics of the scale such as its focus of assessment (e.g. intensity, frequency, severity) and the type and number of response options for scale items. In relation to scaling response, for the reasons discussed earlier in the thesis, the Likert scale was chosen as the format for the antenatal anxiety scale developed in this research. Likert scales, also commonly known as rating scales, consists of the sum of responses on a number of Likert items. A Likert item (hereafter referred to as scale item, or simply item) is typically a declarative statement related to the target construct. This is followed by a number of response options or levels. Respondents are asked to assign a quantitative value to one of the response options to indicate the degree of endorsement or agreement with the statement (Netemeyer et al., 2003; DeVellis, 2012). An important aspect to consider in Likert scales is the focus of assessment. This refers essentially to the issue of what a respondent is asked about (Rose & Devine, 2014). Both the instructions given to respondents (e.g. *'Please indicate how much…'*) and the type of response categories (e.g. *'Not at all'; …; 'Very much'*) dictate the focus of the assessment. In rating scales measuring specifically psychological symptoms, individuals can be asked about their experience of symptoms in a number of ways. Common response formats are those focused on frequency (e.g. *never* to *always*), extent or degree (e.g. *not at all* to v*ery much*) which can be used to measure either severity

or impact, and similarity or agreement (e.g. *most like me* to *least like me).* Of importance is also the issue of the number of response options given to respondents. Likert-type scales typically vary between three and ten options (Furr, 2011). One might argue that a desirable quality of a psychological scale is its variability, and accordingly that generally a higher number of response options would be preferable to capture individual differences. However, the ability of respondents to discriminate in a meaningful way between more than six or seven categories is questionable (Abell et al., 2009) and many Likert scales used in clinical practice or research have between four and seven response levels. It has been suggested that an upper practical limit for the usefulness of a scale can be reasonably set at seven response categories, and there is evidence that five response options might represent the optimal trade-off between scale precision and accuracy on one hand, and practicality and cognitive burden placed on respondents on the other hand (Streiner & Norman, 2008). For the scale developed as part of this research, a five-point Likert scale, with the focus of assessment on frequency of symptoms based on a temporal reference (i.e. *'Never'* to *'Always'*) was chosen as the format of the scale. The decisions made in relation both to the focus of assessment and to the number of response levels aspects were guided by the PROMIS anxiety item bank, as documented in detail in the following section on sources of item writing.

Sources of item writing:

The sources that scale developers use for writing items are of considerable importance, and to maximise face, content and construct validity they can be generated from a range of sources, which should ideally include the relevant literature, intended respondents and expert opinion (Rattray & Jones, 2007; DeVellis, 2012). A number of scholars recommend considering existing scales or item banks as a valuable source for the generation of an initial item pool (Kline, 2005; Streiner & Norman, 2008). In the early phases of the PhD, I became aware of the PROMIS (Patient-Reported Outcomes Measurement Information System) anxiety item bank (Pilkonis et al., 2011). PROMIS is a research initiative by the US National Institutes of Health (NIH) designed to improve self-report scales in healthcare and health research using state-of the-art psychometric approaches (Riley, Pilkonis & Cella, 2011). A range of item banks have been developed for the measurement of both physical and mental health (e.g. fatigue, pain intensity, depression) and designed to be applicable to a range of populations. All item banks were constructed based on extensive development work

including reviews of existing scales, qualitative assessment by intended respondents and experts, cognitive interviewing and repeated psychometric evaluation of items (Riley et al., 2011). These item banks, which are freely available for partial or full use (as long as scale items are not altered), have been described as arguably the most advanced attempt to date in the application of psychometric methods to health-related assessment (DeVellis, 2012; Smith & Jensen, 2019).

The specific PROMIS Anxiety item bank consists of 29 Likert items enquiring about a range of symptoms of anxiety disorders (Pilkonis et al., 2011). Examples of scale items are *'I Felt terrified'* or *'I had difficulty calming down'*. The possibility of using the PROMIS anxiety item bank as one of the sources for the initial item pool was discussed with my supervisors. The item bank was examined and, considering also the robust evidence-base which informed its development, it was determined that several PROMIS anxiety items could be used for all the non-specific symptoms of antenatal anxiety identified as relevant indicators of the construct, if they were adequately represented by one of the PROMIS anxiety items. PROMIS items eventually formed approximately one quarter of the initial pool of candidate items for the assessment of antenatal anxiety. For all the other symptoms which were not appropriately represented in the item bank, items were written *de novo*. Various authors have indicated that the wording of items should reflect, as far as possible, the terms used by individuals from the target population (Kline, 2005; Streiner & Norman, 2008). Consequently, these items were generated based on the findings of the psychometric systematic review and, when appropriate, on the wording used by interviewed women and phrased to be consistent with the included PROMIS items. Expert opinion from health professionals working in the area of perinatal mental health was also sought as an additional source of item generation (6.3.3, 6.4). For the scale developed in this research, as noted earlier, the choice of the focus of assessment and number of response levels were also based on the PROMIS Anxiety item bank, primarily to maintain consistency with the PROMIS items included in the item pool. It was thus decided to use the PROMIS scaling response and number and type of response options, which are also based on rigorous development work, as the format for the antenatal anxiety scale. Specifically, a 5-point Likert-type scale measuring frequency of symptoms over the past week, with the response options being *'Never'*, *'Rarely'*, *'Sometimes'*, *'Often'* and *'Always'*, was chosen as the scale format for the initial item pool.

Clarity and comprehensibility of scale items

Clarity and comprehensibility are essential rather than desirable qualities of item wording. Some authors have argued that the clarity of items should be considered a basic psychometric property in scale development and validation (DeVellis, 2012). The issue of clarity and comprehensibility of scale items can be broken down into a few different considerations, including: Are items worded in a way that makes them easily comprehensible to intended respondents? Are they as unambiguous as possible (i.e. likely to be understood in the same way by all individuals)? Is the structure, syntax and grammar of items sufficiently simple so that minimal cognitive burden is placed on respondents?

The PROMIS anxiety items included in the initial item pool were generated consistently with best practice in item generation. All the further items written *de novo* for inclusion in the item pool were also generated taking into account widely accepted principles of item writing (Streiner & Norman, 2008; Abell et al., 2009). For instance, it was considered that the scale needed to be relevant across a range of levels of literacy and education (McHugh, Rasmussen, & Otto, 2011). For this reason items were kept as short as possible, both in relation to overall item length and to the length of individual words. They were also formulated using simple syntax and grammar, jargon-free language, and with the aim of being directly relevant to the target population, as this issue of face validity is known to increase respondents' motivation to answer accurately (Mokkink et al., 2010a). It is also considered best practice to avoid specific types of items such as double negatives and double-barrelled items (i.e. items containing more than one central idea), as they have been shown to compromise item clarity (Bowling, 2014). Similarly, it was important to avoid qualifiers, as well as vague or ambiguous terms. Once an initial item pool was generated, the wording of all newly devised items was reviewed by my supervisory team as well as by other key informants, as further discussed later in the chapter.

Scale length and clinical utility

As noted in Chapter 1, since the early stages of this research one of the objectives was to develop a screening scale for the assessment of antenatal anxiety that was potentially feasible to implement in routine antenatal care. With this objective in mind, the full NICE guideline on antenatal and postnatal mental health (2014) was consulted to examine which type of screening scales for perinatal mental health problems were considered appropriate for

potential implementation in maternity care settings. The Guideline Development group limited their review "to instruments likely to be used in UK clinical practice that is, 'brief instruments', defined as those which are less than 12 items" (NICE 2014, p.84). This was the primary reason for aiming to develop a scale which contained no more than 11 items, in order to make it relevant and potentially applicable to the context of UK clinical practice. Both in research and clinical settings, shorter scales are also generally preferable as they tend to have shorter administration time and place less cognitive demand on respondents, thus providing a reasonable compromise between measurement accuracy and projected burden of use (Abell et al., 2009; Rose & Devine, 2014).

The next section discusses the process of generation of a pool of candidate items and provides a list of all the 52 items that were included in this initial item pool.

6.3.2 Formulation of a pool of candidate items

As DeVellis (2012) notes, ultimately the psychometric properties of a scale are directly determined by the items that constitute it. The fundamental importance of devising items that reflect the scale purpose and are clear and relevant to the target population cannot thus be underestimated. It has been noted several times in this thesis that scale items in the initial item pool should comprehensively reflect all facets of the construct of interest (Clark & Watson, 2003; Furr, 2011). Further procedures, which can include the use of expert opinion and preliminary psychometric analyses, can subsequently be used to reduce the number of items. Consequently, although the final target was to produce a scale shorter than 12 items, the initial set of candidate items included considerably more items.

The findings from the psychometric literature on antenatal anxiety and the qualitative interviews with women with experience of the target construct were combined in order to represent in the initial item pool the entire range of anxiety symptoms identified as potentially important indicators of antenatal anxiety. All symptoms with moderate or strong evidence for their importance as indicators of antenatal anxiety in at least one of the studies were considered. An evaluation of the strength of evidence for all identified symptoms was presented in conclusion of the two chapters (Appendix 7, section 5.3.2). This evidence was considered when determining the relative importance, and proportional representation, of different symptoms and symptom domains of the construct in the initial item pool. While a

precise numerical rule was not applied, some basic principles to determine the proportional representation of different symptoms in the item pool were used, as follows:

- Symptoms with strong evidence in both studies were represented in the item pool by multiple items
- Symptoms with strong evidence in one of the two studies, or moderate evidence in both studies, were represented by one or more items
- Symptoms with moderate evidence only from one of the two studies were discussed with the supervisory team and a decision was made on whether to include one item to represent it.

A total of 52 items were included in the item pool. This initial pool of candidate items included 15 items from the PROMIS anxiety item bank, which were considered to accurately reflect specific facets of the construct of antenatal anxiety and be applicable to a pregnant population, and a further 37 items which were generated in order to comprehensively represent all facets and symptom domains of the target construct. These new items were written according to the principles of item writing discussed above and worded to be consistent with PROMIS items. The list of items is presented (Table 11) according to the five core symptom domains discussed in the definition of the construct of antenatal anxiety discussed earlier. All PROMIS items, mainly included in the *'Worry and anxious apprehension'* and '*Fear'* symptom domains, are indicated by a (P).

Table 11 – Initial pool of candidate items for the assessment of antenatal anxiety

Worry and anxious apprehension	Fear	General distress
I felt tense	*I felt fearful (P)*	*I felt distressed*
I felt on edge	*I felt scared*	*I felt upset (P)*
I had difficulty calming down (P)	*I felt frightened (P)*	*I felt overwhelmed*
I felt something awful would happen (P)	*I felt panicky for no good reason (P)*	*I felt uneasy (P)*
I felt worried (P)	*I had sudden feelings of panic (P)*	*I was much more irritable than usual*
My worries overwhelmed me (P)	*I had sudden feelings of panic (P)*	*I felt unable to cope*
I worried more than usual	*I felt really anxious*	*I felt like I needed help for my anxiety*
Many situations made me worry (P)	*I had a racing or pounding heart*	*I felt like I was losing control*
I found it hard to stop worrying	*I had a feeling of impending doom*	*I felt guilty*
I expected the worst to happen		*I was harsh with myself*
I had repeated thoughts		*I felt I had lost my confidence*
I had racing thoughts		*I felt indecisive*
I found it hard to focus on anything other than my anxiety (P)		

Pregnancy- related anxiety	**Anxiety-driven behaviours**
I felt scared about giving birth	*I avoided certain places or situations*
I worried about the birth	*My anxiety stopped me from doing things*
I was afraid of the pains of contractions or delivery	*I avoided people*
I worried about losing my baby	
I worried about my health	*I felt the need to keep checking if my baby was moving*

I had negative thoughts about childbirth	I sought reassurance from friends, family or health professionals
I worried something will go wrong with the birth	
I worried about my baby's health	
I felt scared that I couldn't feel my baby moving	
I worried that something may be wrong with my baby	
I was anxious about not being able to cope with the new baby	
I worried about my own appearance	
I felt detached from pregnancy and the baby	

Consistently with the recommendations that a pool of candidate items should be over-inclusive, the item pool was characterised by a certain degree of item redundancy, particularly for symptoms found to be central in the target construct, such as problematic worry. It also included items only marginally related to the construct of interest. As previously mentioned, in later stages of scale development the contribution and inputs of key informants such as experts in the field or intended respondents (which may be asked to consider factors such as items' content, wording, clarity and acceptability), as well as pilot psychometric testing, can all be used to reduce the number of items in order to produce a final, shorter scale with robust psychometric properties. The initial stage of item revision and refinement is presented in the next section.

6.3.3 Preliminary revision of the item pool based on key informants

Before the process of initial item reduction based on expert opinion, which is detailed in the following section (6.4), the initial pool of candidate items for the assessment of antenatal anxiety was subject to a preliminary revision based on inputs from the target population and experts. For this purpose, three MMHS Change Agents were invited to review all the items included in the initial item pool and provide feedback on their wording, clarity and acceptability. Specifically, they were asked, with two separate questions, to rate how clear and acceptable each item was on a scale from 1 to 10 (e.g. 1 = *not clear at all* to 10 = *perfectly*

clear"). They were also asked to suggest an alternative wording for items that were not considered sufficiently clear. Feedback and suggestions for modifications were gathered via email. The vast majority of items received a rating of 9 or 10 in relation to their clarity and acceptability. The wording of four items was modified based on their feedback:

- *"I had racing thoughts"* was changed to *"My mind was racing"*

- *"I had repeated thoughts"* was changed to *"Thoughts got stuck in my head"*

- *"I felt really anxious"* was changed to *"I felt extremely anxious"*

- *"I felt I had lost my confidence"* was changed to *"I did not feel like myself"*

In a subsequent phase, the same three Change Agents also contributed to the design of the final version of the scale, by providing feedback on consecutive versions of the scale.

In order to maximise the chances that the item pool comprehensively reflected the range of anxiety symptoms that can be experienced during pregnancy, two clinicians with specific expertise in perinatal mental health (a Clinical Psychologist and a Nurse Consultant in Perinatal Mental Health) were also invited to review the item pool. They were specifically asked whether they considered any content areas to be missing from the initial list of items, and in case to suggest further items for inclusion in the initial item pool. Three further items were suggested:

- *"I have felt so anxious that I had thoughts about terminating the pregnancy"*
- *"I worried whether having a baby is the right thing for me at this time in my life"*
- *"I felt that my anxiety made me act impulsively"*

Consequently, based on their input, a total of 55 items were included in the initial item pool.

6.4 Using expert opinion to reduce the initial item pool: a Delphi study

The first part of this chapter documented the process of definition of the construct of antenatal anxiety and the subsequent generation of a large item pool based on the evidence from the studies reported earlier (Chapters 4 and 5) in the thesis. The inputs and feedback of key informants, aimed to improve the clarity and acceptability of items, as well as to ensure that all key facets of the target construct were adequately represented, was also briefly discussed. In this second part of the chapter, the process of initial reduction of the number of items based on expert opinion is presented. As discussed in the Method chapter, this stage is often required in scale development to reduce the original item pool to a scale of reasonable length before preliminary psychometric testing can be carried out on a sample of individuals from the intended population of respondents (Abell et al., 2009; Furr, 2011). The key objective of this phase was thus to discard items considered to be less relevant for the assessment of the target construct, as well as problematic items. In scale development, the involvement of individuals with specific expertise in the area of the target construct to support this process of item reduction is recommended by a number of authors (Netemeyer et al., 2003; Streiner & Norman, 2008; DeVellis, 2012). In particular, it has been suggested that expert opinion can significantly enhance the face, content and construct validity of the final scale (Clark & Watson, 2003; Simms, 2008). In the development of a scale aimed to assess a psychological construct, health professionals with clinical knowledge of the target condition are clearly well-placed to evaluate the relevance and appropriateness of items in assessing the condition of interest (Streiner & Norman, 2008). Their views in scale development can thus be used to inform the selection of items that are judged to be sufficiently relevant to the measurement of the target construct, and discard those that are considered problematic (e.g. unclear or ambiguous), redundant or that simply are not deemed to tap into an important aspect of the target construct (Abell et al., 2009).

In this research, expert opinion was sought to support this initial phase of item reduction. Specifically, a range of health professionals working in the area of perinatal mental health in Scotland participated in a Delphi study, and were asked to rate all items in the initial item pool in relation to their importance as indicators of antenatal anxiety. The Delphi technique is one of a range of consensus methods that can be used to consult experts and systematically gather their views. This technique is discussed in more detail in the following section.

6.4.1 Method

The Delphi method is one of the most popular consensus methods used to gather the opinion of experts and establish an adequate level of consensus in relation to a specific topic under investigation (Hsu & Sandford, 2007). The Delphi technique has a number of possible variations, also known as modified Delphi studies. However, it generally involves a group of individuals with expertise in a subject area taking part (often remotely) in two or more rounds of questionnaires in order to provide their expert opinion in relation to one or more topics of interest (Iqbal & Pipon-Young, 2009). Questionnaires typically take the form of Likert-type scales listing a number of study items. Depending on the focus and objectives of the Delphi, study items presented to experts can be as varied as research priorities, issues to be addressed in a specific area of healthcare or a pool of candidate items in scale development (Waggoner et al., 2016). Subject experts (also known as Delphi panellists or simply panellists) are asked to give their views on the topic of interest, in an anonymous form, by rating each of the study items according to a given criterion, often related to the items' importance, relevance, or some other parameter of significance. After the first round, experts are provided at each subsequent round with individualised feedback comprising a statistical summary of the ratings of all panel members in an anonymous form, as well as a reminder of their own ratings (Hsu & Sandford, 2007). At this stage, panellists are given the opportunity to reconsider and change their initial responses. The iterative component of the Delphi (i.e. more than one round) thus provides a means for consensus building through anonymous feedback on the collective opinion of the group (Gill et al., 2013). Consensus is generally achieved when a predefined level of agreement is reached by a sufficient number of experts (Hsu & Sandford, 2007). For the specific aim of item reduction, which was the primary objective of this phase, the Delphi method appeared to be particularly suitable among other consensus methods. Its features of anonymity (face-to-face contact is not required, thus minimising the influence of dominant individuals in group dynamics), iteration through statistical feedback aimed to build consensus of opinion, and the possibility of conducting a Delphi study via an online platform (e-Delphi), all contributed to the choice of this technique for the purpose of initial item reduction based on expert opinion.

As noted above, there are a number of variations to the Delphi method. The classical Delphi process generally makes use of an open-ended questionnaire in the initial round to facilitate the generation of ideas. In this case, three rounds of Delphi are typically considered optimal,

with the second and third round conducted using structured questionnaires generated based on the information gathered in the initial round (Gill et al, 2013). However, it is a common modification of the Delphi technique to start directly with a structured questionnaire, when this is based on previous research such as an extensive review of the literature, and plan to conduct two or three rounds in total (Alexander, 2004; Iqbal & Pipon-Young, 2009). A third round may be considered if the level of consensus is not deemed sufficient following conclusion of the second round (Gill et al., 2013). In this research, both the systematic review and the qualitative interviews that informed the generation of the initial item pool were considered to provide a robust theoretical base for the structured questionnaire in round one. It was thus decided to conduct a multi-round e-Delphi and start with the item pool generated through the two studies presented earlier as the initial set of study items in round one. An advantage of conducting, when appropriate, a two-round Delphi also relates to the potential issue of attrition of Delphi panellists. It has, in fact, been noted that panellists' response rates can quickly be compromised when more than two rounds are conducted (Hsu & Sandford, 2007).

In conclusion of this section, the aim of this e-Delphi was to reduce the number of items in the initial item pool by selecting only those who achieved a sufficient level of consensus among expert panellists regarding their importance as indicators of antenatal anxiety. Other specific methodological choices made in preparation for this study, such as those related to the type and optimal number of experts to be recruited, the procedure for data collection and the criteria chosen to determine an acceptable level of consensus are discussed in the following sections.

6.4.2 Sampling and recruitment of Delphi panellists

The selection of an expert panel for a Delphi study is a task that needs to be considered with attention, as the credibility of the findings obtained is directly dependent on the relevance of the knowledge and experience of individuals participating in the Delphi process. In health research, it has been indicated that an expert can be "any person with experience and knowledge of a particular topic" (Cantrill, Sibbald, & Buetow, 1996, p.69). While the inputs of women with experience of antenatal anxiety were considered in a previous stage of scale development (Chapter 5), this phase of item reduction focused on the contribution of individuals with clinical expertise in the area of perinatal mental health.

A convenience sampling technique was thus adopted to recruit health professionals with specific expertise in the area of perinatal mental health to take part in this e-Delphi. The collaboration established with members of Maternal Mental Health Scotland in the early stages of the PhD proved to be, once again, particularly beneficial to the research. Experts taking part in the e-Delphi were, in fact, recruited through the mailing list of MMHS. The mailing list contained email addresses of 38 individuals, including a range of professions working in relevant roles in the area of perinatal mental health in Scotland (e.g. psychiatrists; clinical psychologists, mental health nurses; health improvement officers). I was introduced to the Secretary of MMHS by the MMHS Change Agents' coordinator. The Secretary agreed to forward an invitation email for the e-Delphi to all health professionals included in the list. While they were all approached as potential recruits through this introductory email, it was also considered important to limit participation to individuals with significant expertise in the area of interest. It was thus decided to consider eligible only individuals with at least two years of experience in a professional role in the area of perinatal mental health. Subsequently, the following inclusion criteria were used to recruit expert panellists for this e-Delphi:

- Individuals subscribed to the mailing list of Maternal Mental Health Scotland in a professional capacity
- At least two years of experience in a clinical or other relevant role in the area of perinatal mental health

With regard to the optimal number of experts taking part in a Delphi, Linstone and colleagues (2002) suggest that between 10 and 50 experts can be generally considered sufficient. It has also been noted that when there is a relative homogeneity with regard to the range of expertise among panellists and the topic under investigation, a total of 10 to 15 experts can be considered adequate (Hsu & Sandford, 2007). In this study, it was considered that the mailing list of MMHS with its approximately 40 health professionals was an appropriate sampling frame, even when taking into account that a proportion of them may have not been eligible and that others would not complete the entire Delphi process.

Recruitment took place over the course of three weeks. In an introductory email forwarded by the Secretary of MMHS to all professionals in the mailing list, potential participants were given information about the general aim of the research and specifically about the eDelphi. Attached to this introductory email were a leaflet briefly introducing the e-Delphi (Appendix 11) and a summary of the study. The invitation email made clear that recruitment was limited

to professionals with at least two years of experience working in the area of perinatal mental health. All potential recruits were invited to contact me for any questions or clarifications.

6.4.3 Data collection and analysis

Data collection

The use of web-based platforms to conduct Delphi studies (e-Delphi) has become increasingly common in research studies (Boulkedid et al., 2011; Gill et al., 2013*rs*). DelphiManager, a web-based system developed by the COMET initiative (http://www.comet-initiative.org/delphimanager/) at the University of Liverpool and specifically devised to facilitate the set up and management of Delphi surveys, was used in this study. The study was launched by circulating an email to all potential recruits with a link to the registration page of DelphiManager. This page was set up to collect basic information on potential participants (job role, professional background, expertise in perinatal mental health in years) and to determine inclusion in the study. Those who indicated less than two years of clinical or other relevant experience in the area of perinatal mental health received an automatic message in which they were reminded of the inclusion criteria and thanked for their interest in the study. Completion and submission of the registration form was considered to indicate consent to take part in the e-Delphi. All experts who met the inclusion criteria and completed the registration form were automatically sent a secure, individualised link to take part in the first round of eDelphi and were informed that they had two weeks to complete this initial round.

The individualised link allowed experts to access the homepage of the eDelphi (Appendix 12). This included a brief summary of the study and description of the Delphi process, as well as specific instructions to complete the first round. As it can be seen in the second page of the appendix, information provided to panellists also included a list of points to consider when completing the eDelphi. The eDelphi was set up so that the entire list of items in the initial item pool was presented to experts in each round, asking them to score each item on a Likert scale. Specifically, Delphi panellists were asked to "*rate each item according to how much you consider it to be an important indicator of problematic anxiety in pregnant women*". The Likert scale was set on a scale ranging from 1 to 9 divided into three response categories, with 1-3 indicating '*limited importance*', 4-6 indicating that the item was '*important but not essential*', and 7-9 to indicate items considered '*essential*'.

The use of the DelphiManager platform greatly facilitated the data collection process. The system has an inbuilt functionality to calculate the distribution of scores for a particular round. The score distribution of the whole group, alongside a reminder of their own score, is thus automatically displayed to each panellists at every subsequent round. In the second round experts, having been shown the distribution of scores and a reminder of their own score for the previous round for each item, were asked to reflect on the group opinion and their own ratings, and rescore all the items using the same instructions of the initial round.. Although, as documented earlier in the chapter, the inputs of two experts in relation to any item missing from the initial item pool had already been sought prior to this e-Delphi, it was also considered important to ask Delphi panellists whether they considered any key symptoms of the target construct to be missing, and in case suggest one or more items that would adequately reflect it. Although the first round was based on a structured questionnaire (i.e. the initial item pool), panellists were thus also invited through a free text box (only for the first round) to suggest any additional item referring to a specific indicator of antenatal anxiety that they considered to be missing.

Eventually only two rounds of eDelphi were required, according to the pre-defined criteria discussed in the following section. Following the calculation of scores for the second round, all participants were automatically emailed a link to a "Thank you" page on DelphiManager, and were given the option to indicate whether they wanted to be sent a summary of the research following its completion. Confidentiality was preserved at all times.

Data analysis

Descriptive statistics were used to examine information on the general composition of the expert panel, including the panellists' job role, professional background and expertise in years. With regard to the ratings that experts assigned to all study items, the mean, median and mode scores for all items both for round 1 and round 2 were calculated. The selection of items for further psychometric testing, however, was based solely on the results of round 2. Further, relevant statistics that were also calculated in round 2 were the number and proportion of panellists who scored an item in the 7-9 range, the category indicating items considered *'essential'*. A free text box available to experts in the first round was also examined for any items suggested for inclusion in the item pool.

As discussed earlier, expert opinion in the specific case of the development of psychological scales can be used to obtain a quantifiable rating of the relevance of each item to the assessment of the condition of interest, and consequently inform the selection of items that are judged to be adequately relevant for its measurement, according to pre-defined criteria for determining consensus. This was the primary aim of this study. Several authors point out that in the literature on the Delphi method there is no consensus with regard to how to establish when consensus is achieved (Waggoner et al., 2016). While this may appear slightly ironic, it implies that researchers using this technique are required to decide from the outset of a study the rules for determining when consensus is reached and which statistics should be considered relevant (e.g. central tendency, dispersion, ranges) to determine it. An additional point to consider is that these choices will primarily depend on the focus and aim of the study, and in particular on whether the main objective is to establish consensus in a positivist sense (as opposed to a lack of consensus) or the extent and nature of consensus (Linstone & Turoff, 2002).

The main statistics used in Delphi studies to present and analyse the collective judgment of a group of experts are those of central tendency (mode, mean, media) or variation (ranges and standard deviations) (Hasson, Keeney, & McKenna; 2000). The proportion of ratings falling within a given range has also been suggested as a workable alternative to determine whether a certain level of consensus is achieved (Hsu & Sandford, 2007; Gill et al., 2013). Consequently, results can be presented in different ways depending on the statistical summaries of interest and the approach chosen to establish adequate consensus.

The criteria utilised in this study were determined in consultation with a colleague with specific expertise in the Delphi method, and primarily based on considerations related to the objectives of the study itself. First, the initial reduction of the item pool presented here aimed to obtain a shorter list of items for preliminary psychometric testing on a sample of intended respondents. However, it was considered important to retain a sufficient number of items for this phase of further preliminary testing (Kline, 2005; DeVellis, 2012). This implied that the level of consensus required for item selection to the next phase needed to be strict enough to ensure the credibility and scientific robustness of the process, while sufficiently broad to allow items with only adequate evidence of importance to be tested further on a sample from the target population.

Specifically, the following two criteria were used to determine whether an item was considered sufficiently important by the Delphi panellists as an indicator of antenatal anxiety:

- An average rating of at least 6.50.
- More than 50% of panellists rating the item in the *'essential'*, 7-9 category.

It was considered that these criteria ensured that items selected for pilot testing were considered essential by a substantial proportion of experts (i.e. $> 50\%$) while also guaranteeing that the panel as a whole judged the item to be sufficiently important (average score across all panellists ≥ 6.50). These criteria were applied to experts' ratings following completion of round two, and once established that a sufficient number of items achieved the pre-defined level of consensus to be selected for the next phase of psychometric testing, the eDelphi was closed.

6.5 Results: reaching consensus on a preliminary version of the scale

The two rounds of this e-Delphi were completed over slightly less than five weeks. A total of 26 health professionals agreed to take part in the first round and were eligible to participate, from an initial sampling frame of 38 individuals. However, only 22 of the 26 panellists who started the first round completed it by rating all items. The remaining 4 panellists were sent two reminders during the course of two weeks, and although the majority (n = 3) had started rating the items none of them completed them all. The incomplete ratings of these 4 panellists were not included in the summary statistics of the first round, and it was decided in consultation with my supervisors not to invite them to the second round. As a result, only the 22 panellists who rated all the items were considered for the second round. Round two was launched two days after the conclusion of the first round, and the remaining experts were sent a further individualised for the completion of the second round. This final round, which also took place over the course of two weeks, was eventually completed by 16 panellists. As discussed earlier, while all study data was analysed, only the ratings of these experts who completed both rounds were considered to determine whether consensus was achieved following two rounds of eDelphi.

The experts who took part in this study included health practitioners from a range of disciplines (e.g. psychiatry, clinical psychology, midwifery, mental health nursing), all with clinical knowledge or other significant expertise in the area in perinatal mental health. This is further detailed in Table 12 in the following page, which presents some key characteristics of the 16 professionals with expertise in perinatal mental health who completed both rounds of eDelphi. The 'Job title' and 'Professional background' categories were entered in free text boxes, while 'Years of experience in perinatal mental health' had five possible options; 0-1; 2-3; 4-6; 7-9 and 10 or more. The 0-1 option is not represented among the 16 panellists who completed both rounds, as they were all eligible to participate in the study.

Table 12 – Key characteristics of panellists participating in round two of the e-Delphi

Panellist ID	Job title	Professional background	Years of experience in perinatal mental health
2	Clinical nurse specialist	*Mental health nursing*	4-6
3	Consultant clinical psychologist	*Clinical psychology*	10 or more
5	Nurse Consultant Perinatal Mental Health	*Mental health nursing*	10 or more
9	Senior Midwife	*Midwifery*	10 or more
10	Community Psychiatric Nurse	*Mental health nursing*	2-3
11	Midwife/National Officer	*Midwifery*	4-6
13	Consultant Clinical Psychologist	*Clinical psychology*	10 or more
14	Community Psychiatric Nurse	*Mental health nursing*	4-6
15	Consultant Clinical Psychologist	*Clinical psychology*	2-3
17	GP	*Medical Doctor*	10 or more
18	Charge Nurse	*Mental health nursing*	10 or more
19	Specialist Midwife in Perinatal Mental Health	*Midwifery*	7-9
20	Consultant Psychiatrist	*Medical Doctor*	10 or more
22	Counsellor	*Counselling*	4-6
23	Advanced Nurse Practitioner Perinatal Mental Health	*Mental health nursing*	7-9
25	Consultant clinical psychologist	*Clinical psychology*	4-6

As discussed earlier, the instructions for panellists in the first round also asked them to suggest items to cover key areas of the construct of interest that they considered to be missing. In the course of the first round, four additional items were suggested by four panellists. These were: *"I did not feel worthy of being a mother"*; *"I felt that my anxiety made me argue with loved ones"*; *"I felt so anxious that I had thoughts of ending my life"* and *"I needed someone to support me with my anxiety"*. The fact that only four items were suggested for inclusion in the second round would appear to suggest that the initial, 55-item pool of candidate items was sufficiently comprehensive to represent at least the vast majority of key symptoms and indicators of the construct of antenatal anxiety in the views of experts taking part in the eDelphi. Consequently, while 55 items were presented to panellists in the first round, the four additional items suggested by experts during this round were included in the second and final round bringing the total to 59 items.

For reasons of brevity, only the 30 items which achieved an adequate level of consensus according to the criteria discussed above are presented here (Table 13), with their mean scores in round two, and the proportion of panellists who scored each item in the 7-9 category. Items suggested by experts during round one are highlighted as (NEW ITEM). These 30 items were selected for preliminary psychometric testing.

Table 13 – Mean scores of top 30 items in round two (*n* = 16) and number of panellists rating the item in the 7-9 range

Scale item	Mean score in round 2 (Range 1-9)	No. and % of panellists rating the item as *'Essential'* (*n* = 16)
"I felt so anxious that I had thoughts of ending my life" (NEW ITEM)	8.37	15 (94%)
"I found it hard to focus on anything other than my anxiety"	8.00	13 (81%)
"I have felt so anxious that I had thoughts about terminating pregnancy"	8.00	15 (94%)
"I could not control my anxiety"	7.94	16 (100%)
"My worries overwhelmed me"	7.87	15 (94%)
"I felt detached from pregnancy and the baby"	7.81	15 (94%)
"I felt that my anxiety made me act impulsively"	7.56	15 (94%)
"I felt extremely anxious"	7.50	16 (100%)
"I had a feeling of impending doom"	7.50	16 (100%)
"I had a racing or pounding heart"	7.37	11 (69%)
"I felt something awful would happen"	7.25	15 (94%)
"I felt unable to cope"	7.25	15 (94%)
"I felt like I was losing control"	7.25	16 (100%)
"I did not feel worthy of being a mother" (NEW ITEM)	7.25	13 (81%)
"I felt panicky for no good reason"	7.12	15 (94%)
"I had sudden feelings of panic"	7.12	16 (100%)
"I felt overwhelmed"	7.06	11 (69%)
"My mind was racing"	7.06	13 (81%)
"I avoided people"	7.06	14 (87%)
"My anxiety stopped me from doing things"	7.00	11 (69%)
"Thoughts got stuck in my head"	7.00	12 (75%)
"I felt like I needed help for my anxiety"	6.87	13 (81%)
"I worried that something may be wrong with my baby	6.81	13 (81%)
"I needed someone to support me with my anxiety" (NEW ITEM)	6.81	10 (62%)

"I have felt scared about giving birth"	6.75	12 (75%)
"I have had negative thoughts about childbirth"	6.69	11 (69%)
"I did not feel like myself"	6.62	10 (62%)
"I worried about losing my baby"	6.56	11 (69%)
"I worried more than usual"	6.50	9 (56%)
"I was much more irritable than usual"	6.50	9 (56%)

6.6 Conclusion

As a result of this phase of the research, the 30 items reaching an adequate level of expert consensus in relation to their importance as indicators of antenatal anxiety were selected to be included in a preliminary version of the scale. The 30 items selected through the eDelphi are extensively commented in the next chapter, which documents the process of pilot psychometric testing of this preliminary version of the scale on a population of intended respondents.

Following the Delphi study presented in this chapter, in consultation with my supervisors it was also decided to name the scale Stirling Antenatal Anxiety Scale (SAAS). The word 'antenatal' was preferred to 'pregnancy' to highlight that the scale aimed to assess symptoms of anxiety in the antenatal period, but was not limited to symptoms of pregnancy-related anxiety. In the following chapters, the scale is thus at times referred to as 30-item SAAS (i.e. the preliminary version of the scale tested in Chapter 7), and 10-item SAAS or simply SAAS for the final version of the scale discussed in Chapter 8 and 9.

Chapter 7 Pilot study: preliminary psychometric testing and further item reduction

7.1 Introduction

This chapter presents the first of two psychometric studies that were conducted in conclusion of the research. Here the final stage of Phase 2, consisting of further item reduction through psychometric testing is discussed, while the following chapter presents the preliminary psychometric validation of the final version of the SAAS. A total of 236 pregnant women were recruited to take part in these two cross-sectional surveys. This chapter specifically documents the stage of pilot psychometric testing (Pilot study) which was carried out by administering the 30-item version of the SAAS resulting from the eDelphi to a sample of 62 pregnant women, with the aim of reducing it to a shorter, psychometrically robust version of the scale. The final psychometric validation study presented in the subsequent chapter shared many similarities with the pilot study presented here in relation to setting, sampling, recruitment and part of the data collection procedures. When this was the case, this is indicated in the following sections.

In scale development, before the phase of initial psychometric validation of the measure, it is important to ensure that sufficient pilot work is conducted so that only items with the most robust psychometric performance are included in the final version of the scale (Rattray & Jones, 2007; Streiner & Norman, 2008). Item analysis is central in this stage of psychometric testing. Preliminary psychometric testing through item analysis is specifically aimed to refine the scale by discarding items which do not significantly contribute to improve its psychometric properties, while retaining those that are psychometrically robust and appear to adequately capture the target construct (DeVellis, 2012). Considerations related to item content, clarity and acceptability to the target population should also inform this phase (Streiner & Norman, 2008). Item reduction in pilot testing is thus based on a dynamic and iterative process of examination and comparison of each item's psychometric performance and content (Worthington & Whittaker, 2006). In this instance, the main challenge was to produce a scale which needed to be at the same time relatively short ($\leq$ 11 items), psychometrically robust and able to preserve a range of key facets of the target construct of antenatal anxiety. The process of discarding items is mainly quantitative in nature, being

predominantly based on the analysis of statistical parameters of each individual item (e.g. item-total, inter-item correlations, response distributions), and the relevance of these and other statistical indicators is discussed in detail in the Data analysis section of this chapter. As noted above, in relation to the more qualitative aspects of item analysis, this pilot study also served the purpose of identifying any potential issues in relation to item clarity and acceptability to the target population.

7.2 Study aim and objectives

The aim of this pilot study was to reduce the 30-item pool of the SAAS through psychometric testing on the target population and produce a shorter, psychometrically robust version of the scale.

There were also several secondary aims, which included:

1) To assess the ease of completion, clarity of instructions and acceptability of this preliminary version of the SAAS in the target population
2) To identify any items that, based on respondents' feedback, were unclear or were not considered acceptable for use in routine antenatal care
3) To evaluate the appropriateness of the scale's response options and time frame assessed.

7.3 Method

This study was conducted using a cross-sectional survey design. In cross-sectional surveys, all observations are made at one single point in time. It is a particularly popular method in the social and psychological sciences, as it can be used to gather data from relatively large samples in a time- and cost-efficient way (Bowling, 2014). In relation to the data collection method, an online survey was initially considered. However, because of the nature of the sample and of the topic under investigation, a postal survey was preferred. Women were initially approached and given study booklets by midwives, and this enabled midwives to apply the study inclusion and exclusion criteria and determine study eligibility for all women approached as potential study participants. The initial contact between potential recruits and midwives also ensured that women were given sufficient information regarding the study (De Vaus, 2014). As noted above, for the specific purpose of this study, item analysis was central throughout the phase of item reduction. The following sections discuss sampling and sample size, the recruitment procedure and the stages of data collection and analysis of the pilot study. In cases in which the study methods are the same used for the validation study presented in the following chapter, this will be indicated.

7.3.1 Sample and recruitment procedure

All women taking part in both the pilot and validation studies were initially approached by midwives from hospital and community antenatal clinics in Glasgow, according to the method and procedure presented in the following sections. These include the study setting, sampling and sample size, recruitment procedure, ethical considerations, and data collection and analysis. Both the study setting and the recruitment procedure were the same for the pilot and the validation study.

Setting

Both studies were planned in consultation with the Chief Midwife and the Lead Midwife Community and Outpatient Services for NHS Greater Glasgow & Clyde (GG&C). In an initial meeting arranged by my first supervisor (HC), the aims and recruitment targets for the two studies were discussed and the Princess Royal Maternity was recommended as an appropriate site because of its capacity, with over 6500 births each year. Because of the

relatively challenging recruitment targets (pilot study $n = 50$; validation study $n = 200$, as further detailed below), and following further discussions with my supervisors and with maternity care managers in NHS GG&C, it was decided to include also the Queen Elizabeth University Hospital as a second recruitment site. Both sites had the additional benefit of providing a number of hospital and community antenatal clinics in different parts of the city. The inclusion of a second site thus allowed to reach a considerably larger number of women as potential recruits, as well as to increase the representation of participants from different areas of Glasgow, a city characterised by significant health and socio-economic differences.

Sampling and sample size

The sampling technique initially planned for both the pilot and the validation study was convenience sampling, with the intention to recruit an equal number of women representing the three trimesters of pregnancy. However, the vast majority of women who took part in the two studies were recruited in their second or third trimester of pregnancy, as further detailed below. While the reasons for this are likely to be varied, discussions with midwives suggested that introducing the study to women during the antenatal booking appointment, the only one typically occurring during the first trimester (i.e. before the end of the 12th gestational week), was particularly challenging because of the amount of other areas to cover in this initial appointment. Once it became clear that recruiting women in the first trimester was likely to be unfeasible, midwives were asked to consider as potential recruits all women attending the clinics and meeting the inclusion criteria, until a target quota was reached for each site.

In relation to sample size for this pilot study, as noted above the recruitment target was set to 50 study participants. A combination of methodological and practical reasons informed this choice. The aim of this pilot study was to conduct item analysis, which is focused primarily on the psychometric performance of individual scale items. For this specific purpose, some authors have indicated 30-40 subjects as sufficient to allow the calculation of parameter estimates at the item level (Mooney & Duval, 1993). However, others have shown that it is only above 50 participants that the impact of sample size on item statistics become minimal (Johanson & Brooks, 2010). A practical criterion of sufficient information with minimum use of health professionals and study participants' time was also adopted.

Inclusion and exclusion criteria

The following criteria were used in both the pilot and the validation study to determine study eligibility:

Inclusion criteria

- Women who at the time of recruitment are pregnant between 6 and 38 gestational weeks

- At least 18 years of age

- Receiving routine prenatal care

- Level of English sufficient to understand and complete questionnaires in lay language. This was gauged by midwives recruiting participants

- Able to provide written informed consent to take part in the study

Exclusion criteria

- Major medical or obstetrical complication of pregnancy, as defined by the clinical judgement of the midwife providing antenatal care

- Severe cognitive impairment

- Current severe mental health disorder (any psychotic illness or bipolar disorder)

Recruitment procedure

Similarly to the qualitative interviews, ethical and management approval was sought and obtained from the South East Scotland Research Ethics Committee 02, and additionally from the Research & Development service of the NHS Greater Glasgow & Clyde (GG&C) Health Board (Appendices 1 and 2) for both the pilot and validation study. Thanks to the Chief Midwife for NHS GG&C, I was introduced to the approximately ten Senior Charge Midwives (SCMs) working in maternity care in the Glasgow area, each coordinating a team of 10-20 midwives. Both in preparation for the two studies and during the recruitment phase, I attended a number of their bi-monthly meetings, which provided the opportunity to plan collaboratively part of the recruitment procedure, as well as to monitor recruitment and address any issues which were identified. All midwives involved in recruitment were given study booklets and provided with information about the study by their SCM. Additionally,

information sheets with instructions on how to recruit study participants were also provided to all midwives, both for the pilot and the validation study (see Appendix 13 for an example). During recruitment for the pilot study, midwives were given a total of 300 study booklets over the course of seven weeks. Recruitment was interrupted once the required number of questionnaires were returned by post to the study office (NMAHP Research Unit, University of Stirling).

In the initial recruitment phase, women attending antenatal clinics and meeting the inclusion criteria were verbally informed of the study by a midwife part of their direct healthcare team. If a woman showed interest in taking part in the research, her midwife provided her with a study pack. The study pack contained an invitation letter, the information sheet, a consent form, a study questionnaire, and a pre-paid/pre-addressed envelope. Interested participants were instructed to post the questionnaire and consent form to the study office using the pre-paid, addressed envelope provided. The study documents were all reviewed by three of the MMHS Change Agents before ethical approval was sought, and modifications to the information sheet were made based on their feedback

Ethical considerations

Considerations related to obtaining informed consent from study participants, minimising any burdens and risks and safeguarding the confidentiality of sensitive information were previously discussed in 3.4.1, 3.4.2 and 3.4.3. Here some further ethical considerations are discussed. As noted above, all women taking part in either the pilot or the validation study were approached during routine antenatal clinics by midwives who were part of their direct healthcare team. Midwives approaching all potential recruits gave women a brief explanation of what the study involved, including what they would be consenting to. The information sheet given to all women taking part in both studies (see Appendix 3 for an example) included information and contact details for accessing support in case they felt distressed as a result of completing the self-report scales. Advice in relation to health professionals that could be contacted in order to discuss further any potential issues and in case access relevant support, and an out of hours contact number were also provided.

The procedures that were followed to safeguard the confidentiality of sensitive information gathered from study participants in the two studies are also presented here:

- Identifiable information contained on paper was returned by post in the form of consent forms and reply slips with participants' names and contact details. After a unique code was assigned to each participant, this identifiable information was separated from other study data (i.e. the completed scales and demographic/obstetric information) and kept in a locked filing cabinet.

- Information on the outcome of the diagnostic interview conducted in the validation study was electronically transferred from the researcher using only unique participant codes.

- Once interviews were completed, reply slips with participants' names and contact details were securely disposed of, consistently with Principle 5 of the Data Protection Act (1998).

7.3.2 Data collection

62 questionnaires and consent forms were returned between August and October 2017, thus slightly exceeding the recruitment target ($n = 50$). As 300 study booklets were given to SCMs, the response rate was 20.6%. While this was a relatively low response rate, it is not uncommon in studies using postal surveys (Aday & Cornelius, 2006; Sahlqvist et al., 2011). Moreover, this estimate may not be completely accurate as it was not possible to gather precise data on the number of study booklets that were handed out in total. This was mainly due to the fact that midwives carried study booklets with them in a number of different antenatal clinics, and it was thus unfeasible to keep track of the number of study booklets distributed by each midwife.

The study booklet included an invitation letter, information sheet, consent form and a questionnaire (see invitation letter and questionnaire in Appendices 14 and 15). The main section of the questionnaire, as shown in Appendix 15, consisted of the 30-item version of the SAAS. In relation to other data gathered for this pilot study, it was decided not to ask participants for detailed socio-demographic information not strictly required for the aims of the study. However, three questions were asked to collect information on participants' age, current gestational week and parity (i.e. whether this was the first pregnancy or not). At this final stage of scale development, it was also considered important to gather participants' feedback in relation to specific aspects of the scale. Three further questions were thus asked to evaluate the scale's overall ease of completion, clarity of instructions and acceptability

for use in maternity care. Specifically, women were asked to indicate: *"How easy was the questionnaire to complete?", "How clear were the instructions?"* and *"Would you find it acceptable to complete a questionnaire like this as part of routine antenatal care?"* (Appendix 15). Participants could answer these three questions on a scale from 1 to 10, with 1 representing the worst rating. Finally, a free textbox was also included for any comments that women had regarding the scale itself or specific items within the scale. In particular, participants were asked to comment on items that they found to be unclear.

7.3.3 Data analysis

All data from the questionnaires were initially entered onto a Microsoft Excel Spreadsheet. A simple codebook was developed in order to list all variables, assign names to each variable, specify how each of the variables was measured (e.g. ordinal, interval), and indicate what each numeric code for a given variable represented (Brace, Kemp & Snelgar, 2013). The 30 items included in the scale were considered at the interval level of measurement (Streiner & Norman, 2008). Other variables included the three pieces of information on a respondent's age, current week of pregnancy and reproductive history, as well as the ratings related the scale's ease of completion, clarity of instructions and acceptability for use in antenatal care. Data were subsequently transferred onto SPSS software (version 23). In relation to missing data, it was decided not to include in the data analysis questionnaires in which more than one item response was missing. If only one item within the 30-item scale was not scored by a respondent, the median score for all other items for that respondent was used to replace the missing value (Bland, 2000). Initially, descriptive statistics were calculated to summarise the respondents' characteristics. These included means, standard deviations and ranges. Means and frequency distributions for the responses to the three questions regarding the participants' general feedback on the questionnaire were also calculated. All relevant data were subsequently summarised and tabulated, and are presented in the Results section.

Despite the literature on scale development does not provide precise criteria to inform the selection of items to be included or discarded (DeVellis, 2012), the item reduction phase was based on guidelines and recommendations provided in a number of psychometric textbooks (Kline, 2000; Costello & Osbourne, 2005; Streiner & Norman, 2008; Abell et al., 2009; Furr, 2011). A constructive approach (Abell et al., 2009) was used in this phase of item reduction

to discard items based on several criteria. The section below describes the statistical analyses that were conducted and the criteria used to inform the selection of items.

- Mean score, standard deviation and range for each item

These descriptive parameters were calculated and the mean score for each item is reported in a summary table. While the mean score of an item in a scale assessing frequency of symptoms can provide initial information regarding how common a symptom measured by a specific item is in a given sample, this parameter was secondary in determining which items to include or exclude from the final version of the SAAS. As indicated below, there were other indexes more relevant to this purpose.

- Response distributions

Frequency analyses were conducted to evaluate response distributions at the item level. Response distributions indicate the proportion of respondents who endorse each response category for each individual item, and are thus also known as endorsement frequencies. As noted earlier, participants scored each item of the 30-item SAAS on a 5-point Likert scale ranging from *"1 = Never"* to *"5 = Always"*. The analysis of endorsement frequencies for each item enables the identification of items showing floor or ceiling effect (DeVellis, 2012). These effects refer to cases in which the lower or higher response option is selected by a remarkably large proportion of participants, thus indicating a lack of discriminative power (Turner et al., 2007). The criterion used in this analysis was to discard items for which a single response option was endorsed by > 90% of respondents (Turner et al., 2007; Streiner & Norman 2008). Endorsement frequencies of participants scoring at the minimum or maximum for each item were thus examined. Furthermore, response distributions for specific items were also inspected in cases in which two or more items assessing a similar facet of the target construct had other comparable psychometric properties. In these instances, response distributions contributed to inform the choice regarding which item to retain.

- Corrected item-total correlations

Item-total correlation refers to the degree of correlation of a given item with the total scale score (Coolican, 2009). A corrected item-total correlation is the correlation of an item with the total score omitting that item, as its inclusion would lead to an artificially inflated score (Hunsley & Mash, 2008). A high corrected item-total correlation is a particularly desirable attribute in an item, as it indicates that the item is measuring the same underlying construct assessed by the full scale (DeVellis, 2012). Conversely, it can be assumed that items with low corrected item-total correlations are not measuring adequately the target construct, and may well be measuring a different construct (Kline, 2000). Various authors seem to converge on the recommendation that items with a corrected item-total correlation < 0.30 should be discarded (Clark & Watson, 1995; Kline, 2000; Abell et al., 2009). This criterion was therefore adopted in this study.

Items with a high or moderately high corrected item-total correlation (i.e. > 0.70) were primary candidates for inclusion in the shorter version of the SAAS. However, it was also important to consider that it would be methodologically incorrect to simply select the top items ranked according to their corrected item-total correlations. This procedure would not generate the scale that best measures the latent construct, as it would increase internal consistency at the expenses of breadth of the construct measured (Netemeyer et al., 2003; Furr, 2011). This is primarily because the few items with the strongest correlations with the total score are very likely to be highly redundant. Current thinking thus indicates that a psychometrically robust scale should be mainly composed of items with moderate to moderately high correlations with the total score (Streiner & Norman, 2008; DeVellis, 2012). Pearson product-moment correlation coefficients were used to calculate corrected item-total correlations (Pallant, 2013). For reasons of brevity, for the remaining part of this chapter corrected item-total correlations will be referred to simply as item-total correlations.

- Inter-item correlations

An inter-item correlation is the correlation between two items within a scale. An inter-item correlation matrix was generated in SPSS and inspected to examine this form of correlation and identify items with a particularly high inter-item correlation, likely to indicate item redundancy (De Vaus, 2014). It has been suggested that items with an inter-item correlation

above 0.80 are very likely to be redundant and asking in essence the same or a very similar question (Kline, 2000). However, other authors have indicated that also marginally lower inter-item correlations may indicate a repetition of content between two items (Streiner & Norman, 2008). The criterion used in this study was to inspect all inter-item correlations equal or above 0.75 for suspect redundancy. Inter-item correlations were used to aid the selection of a single item when two or more showed other similar psychometric properties or had a clear overlap in content. In relation to inter-item correlations for the final scale, the main guiding principle was to generate a scale in which items showed correlations between 0.2 and 0.8 (Netemeyer et al., 2003). This criterion was chosen based on the recommendation that a range of inter-item correlations are required to generate a scale that preserves the breadth of the target construct assessed (Furr, 2011). Similarly to item-total correlations, Pearson product-moment correlation coefficients were used as indexes of inter-item correlations.

- Internal consistency

At this preliminary stage of psychometric testing, this was the only parameter examined at the scale level. Factor analysis was not feasible because of the relatively small sample size (Pallant, 2013), and it was only conducted on the final version of the SAAS for the psychometric validation study. Internal consistency is an important aspect related to the reliability of a scale, which was discussed in Chapter 3. The importance of producing a scale with high internal consistency relies on the observation that a high value of internal consistency can be assumed to be an indication that all scale items are measuring the same latent construct (Tavakol & Dennick, 2011; DeVellis, 2012). High internal consistency is thus a particularly desirable characteristic of a scale, and the aim of this pilot study was to produce a shorter, internally consistent version of the SAAS.

Internal consistency is commonly determined by calculating Cronbach's Alpha (Cronbach, 1951), a widely used measure of scale reliability expressed as a number between 0 and 1, with 1 indicating perfect internal consistency. The value of α is dependent on the magnitude of the inter-relatedness between scale items, and on the overall number of scale items, with longer scales typically producing higher values of alpha (DeVellis, 2012). Although there is no agreement on the value of Cronbach's alpha that can be considered to indicate adequate scale reliability, the most commonly cited in the research literature is that indicated by

Nunnally of 0.70 (Nunnally, 1978). While this value is generally considered to be acceptable for scales used in research studies which focus their observations at the group level (e.g. comparing group means), other scholars have suggested higher alpha values for scales to be used in clinical settings to make decisions about individuals. Abell and colleagues (2009), for example, in the case of clinical applications indicate that alpha should be above 0.80, and other authors have suggested values closer to 0.90 (Nunnally & Bernstein, 1994; Hunsley & Mash, 2008; DeVellis, 2012). One of the goals of this pilot study was consequently to produce a final, shorter version of the scale with Cronbach's alpha approximating 0.90.

- Sub-analyses on the six pregnancy-related anxiety items.

Further analyses were also separately conducted considering only the six pregnancy-related anxiety items included in the scale (items 5, 10, 13, 16, 24, 26) as a separate scale. As discussed in previous chapters, pregnancy-related anxiety has been indicated by some authors to be a distinct syndrome from general antenatal anxiety (Blackmore et al., 2016). However, it remains unclear whether this is the case (Bar-Shai et al., 2014). A different approach is to consider pregnancy-related anxiety as one dimension of the general construct of antenatal anxiety, rather than a distinct construct, as implied in the construct definition of antenatal anxiety proposed in the previous chapter. This sub-analysis served two specific purposes. Firstly, at least an acceptable level of internal consistency with a Cronbach's alpha value > 0.70 (Nunnally, 1978) was hypothesised, as it would indicate that these items were significantly interrelated and indeed measuring a unitary dimension related to specific anxiety and worries about pregnancy. Secondly, of interest were also the inter-item and item-total correlations among these six items. If one or more item were found to have only low or moderate correlations with other items in this shortened scale, this would provide further indications that the item does not reflect the theoretical core of the construct measured by other pregnancy-related anxiety items. Consequently, item-total and inter-item correlations, and the contribution of each item to the internal consistency of this PrA scale were all calculated and examined.

The following section reports the findings from this pilot study, which informed the selection of the 10 items included in the final version of the SAAS.

7.4 Results and selection of scale items for validation study

7.4.1 Sample characteristics

Recruitment took place over the course of seven weeks, between August and October 2017. 62 pregnant women in total participated in the pilot study by returning the questionnaire and the consent form. As 300 study booklets were distributed, the response rate was 20.6%. As discussed earlier, convenience sampling was adopted as it became clear that recruiting women in the first trimester of pregnancy was highly challenging. As a result, the sample showed a fairly similar representation of women in their second (42 %, $n = 27$) and third (58 %, $n = 35$) trimester of pregnancy (range 15-42), as illustrated in Table 14. Equally, women at their first pregnancy and women who had previously experienced pregnancy were represented in relatively similar proportions, with respectively 40% of nullipara and 60% of women at their second or subsequent pregnancy. This is consistent with maternity statistics both at the UK and at the Scottish level (respectively 42% and 43% of all live births were first births in 2017 (Office for National Statistics [ONS], 2017; NHS Information Services Division [ISD], 2017). The mean age of women in the sample was 32.1 years (range 21-40), slightly higher than the average age (30.1) at which women in Scotland give birth (NHS Information Services Division [ISD], 2017).

Table 14 – Mean age (SD) and gestational week of study participants

Sample characteristics	Mean (SD)	Range
Age	32.1 (4.6)	21 - 40
Weeks pregnant	29.3 (7.7)	15 - 42
Of which:	Frequency ($n = 62$)	%
15 – 21 weeks	11	18
22 – 28 weeks	16	26
29 - 35 weeks	16	26
36 or more weeks	19	30

Percentages are rounded to the nearest integer

Feedback on the 30-item SAAS

Feedback from respondents in relation to the scale's ease of completion, clarity of instructions and acceptability for use in antenatal care was also sought. The three specific questions are reported in conclusion of the Data collection section above. Participants could answer these three questions on a scale from 1 to 10, with 1 representing the worst rating (e.g. in relation to ease of completion, 1 = *'not easy at all'* and 10 = *'extremely easy'*). The mean, mode and median score for the three questions is presented in Table 15.

Table 15 – Mean, median, mode and standard deviation for questions on feedback on the 30-item SAAS

	Easy to complete 1(No) - 10(Yes)	Clear instructions 1(No) - 10(Yes)	Acceptable 1(No) -10(Yes)
N	62	62	62
Mean	9.66	9.84	9.60
Median	10	10	10
Mode	10	10	10
SD	.82	.51	1.15

As can be seen in the table, all descriptive parameters indicated that this 30-item version of the SAAS was considered extremely easy to complete, with clear instructions and highly acceptable to study participants. This is also illustrated by the number of women scoring 9 or 10 to each of the questions (not reported in Table 15), respectively 57 for both ease of completion and acceptability (92%) and 60 for clarity of instructions (97%).

With regard to missing data, only one participant missed an item score (item 4). As per the data analysis plan, this was replaced with the median value of the remaining items.

7.4.2 Item analysis and internal consistency

The item reduction process which was used to generate the final version of the SAAS occurred in two stages, according to the criteria discussed earlier. In the first stage, item statistics were examined in order to discard the ones that clearly did not contribute to improve the psychometric qualities of the scale. This process was based primarily on the statistical parameters discussed in the Data analysis section. A comparative examination of both item content and psychometric properties was, however, necessary when the content of two or more items was considered to have a high degree of semantic or conceptual overlap. When this was the case, only one item was retained assessing a specific anxiety facet, in order to avoid redundancy in the final scale and maximise content and construct validity (Streiner & Norman, 2008; Abell et al., 2009). The second stage consisted of further examination and comparison among all items that were considered to have sound psychometric properties in the first stage. As noted earlier, in this subsequent stage, while psychometric properties continued to play an important role, other aspects including item clarity, length and the individual contribution of each item to a short measure that retained construct relevance and a sufficiently broad scope were also considered in order to inform the selection of items for the final version of the SAAS. In cases in which decisions regarding the inclusion or exclusion of an item from the final version of the scale were based solely on statistical analyses, these will be reported and commented. Additionally, when the selection of an item derived also from considerations related to its qualitative aspects (e.g. clarity, acceptability), this will be narratively discussed and motivated.

Initially, mean scores, standard deviations and ranges for each item were examined. The mean score for each item is presented below in Table 16, which also reports the item-total correlation for each item. Endorsement frequencies for all items were also calculated and inspected. These, for reasons of brevity, are not presented in a table but discussed narratively, when appropriate, throughout this section.

Table 16 – Mean scores and item-total correlations of the 30-item SAAS

Item number / Item	Mean score (range 1 - 5)	Item-total correlation (30-item SAAS)
1. *I worried more than usual*	2.56	.66
2. *My anxiety stopped me from doing things*	1.56	.71
3. *I had sudden feelings of panic*	1.79	.60
4. *I had a racing or pounding hearth*	1.79	.58
5. *I felt detached from pregnancy and the baby*	1.50	.44
6. *My mind was racing*	2.55	.62
7. *I was much more irritable than usual*	2.77	.44
8. *I felt panicky for no good reason*	1.79	.70
9. *I did not feel like myself*	2.08	.70
10. *I have felt scared about giving birth*	2.66	.42
11. *I felt that my anxiety made me act impulsively*	1.39	.68
12. *I found it hard to focus on anything else than my anxiety*	1.40	.75
13. *I worried about losing my baby*	2.45	.64
14. *I felt unable to cope*	1.47	.75
15. *I felt something awful would happen*	1.81	.61
16. *I worried that something may be wrong with my baby*	2.35	.63
17. *I felt like I needed help with my anxiety*	1.42	.86
18. *Thoughts got stuck in my head*	1.92	.79
19. *I avoided people*	1.71	.77
20. *I have felt so anxious that I had thoughts about terminating my pregnancy*	1.00	/

21. *I could not control my anxiety*	1.29	.84
22. *I felt so anxious that I had thoughts of ending my life*	1.02	/
23. *I felt extremely anxious*	1.35	.81
24. *I have had negative thoughts about childbirth*	2.00	.41
25. *My worries overwhelmed me*	1.66	.77
26. *I did not feel worthy of being a mother*	1.37	.73
27. *I felt like I was losing control*	1.48	.72
28. *I had a feeling of impending doom*	1.31	.80
29. *I felt overwhelmed*	2.16	.75
30. *I needed someone to support me with my anxiety*	1.60	.74

All item-total correlations are expressed as Pearson's *r*.

At an initial inspection of mean scores and endorsement frequencies for all items, it was evident that item 20 *"I have felt so anxious that I had thoughts about terminating my pregnancy"* and Item 22 *"I felt so anxious that I had thoughts of ending my life"* had to be excluded from further analyses, as they were both characterised by a strong floor effect. Item 20 showed no variance, with all sixty-two participants endorsing option *"1=never"*, while item 22 had extremely low variance, with only one out of sixty-two participants scoring *"2=rarely"* and the remaining sixty-one scoring *"1=never"*. As previously noted, items with particularly strong floor or ceiling effect ($\geq 90\%$) should be discarded, as they lack discriminative power among respondents. Of note, neither of these items were formulated based on the systematic review or the qualitative interviews with women who had experienced antenatal anxiety discussed in previous chapters. Conversely, they were suggested for inclusion in the item pool during the Delphi study by health professionals with expertise in perinatal mental health. A possible explanation is that some of the health professionals taking part in the eDelphi may be used to work often with women with severe symptoms for whom these questions may have relevance. However, to screen for problematic anxiety symptoms in the general population of pregnant women they clearly did not appear to have sufficient discriminatory power. When an item shows no or extremely

low variance among respondents, it is recommended not to include it in analyses of item-total correlations or internal consistency reliability, as it will clearly impact negatively on both parameters (Streiner & Norman, 2008). Item 20 and 22 were accordingly excluded before conducting further analyses at the item and at the scale level.

At the scale level, Cronbach's alpha as index of internal consistency was thus calculated for the resulting 28-item scale (after removing items 20 and 22), and was found to be $\alpha = 0.96$. This particularly high level of internal consistency for this preliminary version of the scale provided an initial confirmation that the vast majority of items in the scale were homogeneous and thus likely to be measuring the same latent construct (Furr, 2011). However, some further considerations are required when interpreting such a high Cronbach's alpha value. While, as noted earlier, it has been indicated that scales designed for clinical applications should have values of Cronbach's alpha close to 0.90 (Hunsley & Mash, 2008; DeVellis, 2012), it has also been suggested that values significantly exceeding $\alpha = 0.90$ are likely to result from scales which are either excessively long or that are characterised by a high degree of item redundancy (Brace, Kemp & Snelgar, 2013). This is arguably the case for this version of the scale and, as discussed in previous chapters, a degree of redundancy in the initial item pool was planned and thus expected. Such a high value of Cronbach's Alpha (0.96) for the 28-item version of the scale, however, provided an excellent basis to aim to a shorter, version of the scale that retained a value of internal consistency recommended for use in clinical settings (approximately $\alpha = 0.90$).

At the item level, continuing with the analysis of endorsement frequencies of the 28 remaining items, five items were scored *"3 = Sometimes"*, *"4 = Often"* or *"5 = Always"* by at least 50% of respondents. Specifically, these were item 1 *"I worried more than usual"* (54.8%), item 6 *"My mind was racing"* (51.6%), item 7 *"I was much more irritable than usual"* (61.8%), item 10 *"I have felt scared about giving birth"*(62.9%) and item 13 *"I worried about losing my baby"* (50%). Correspondingly, these were also the five items with the highest mean scores. Considering their endorsement frequencies, it can be reasonably assumed that these items tap into relatively common experiences of pregnancy and cannot thus be considered a reflection of problematic anxiety symptoms or poor mental health more in general. These items were consequently strong candidates to be discarded, but they were not at this stage as it was also important to examine their item-total and inter-item correlations.

With regard to item-total correlations, notably the item with the highest item-total correlation was item 17 *"I felt like I needed help for my anxiety"*, with a correlation of 0.86 with the total score. This indicated that, among all items, this was the one with the best performance in predicting the total score. This, in turn, provided good preliminary evidence that the 28 items as a whole were indeed measuring a latent construct related to problematic anxiety symptoms, considering that the item enquires about the perceived need for support around anxiety symptoms. The average item-total correlation for all items was moderately high (0.67). The inter-item correlation matrix (composed of 729 correlation values) was also visually inspected. For the vast majority, inter-item correlations showed values comprised between 0.30 and 0.70, well within the range specified earlier (0.20-0.80) (Netemeyer et al., 2003). A number of inter-item correlations between two or more items were specifically examined while comparing the psychometric performance of multiple items. When these correlation coefficients, as well as other relevant item statistics (i.e. item-total correlations, response distributions) were used to inform the decision to retain or discard an item, these are reported and discussed narratively in the following section.

7.4.3 Item selection process

The section below documents the iterative process of examination and revision of items which resulted in the exclusion of a number of items, based on the criteria and item selection strategy previously presented. As noted above, following the exclusion of item 20 and 22 in the initial phase of analysis, the process of discarding items and selecting those for the final version of the SAAS occurred in two stages. During the first stage, 17 items were discarded and 13, including three pregnancy-related anxiety items, were retained for further examination, as detailed below. In the second stage, the 10 items included in the final version of the SAAS were selected based on considerations related to their psychometric qualities and the contribution of each item to a short measure that retained construct relevance and a sufficiently broad scope. In the section below, the first stage of the item selection process is initially presented. The analysis is discussed separately for each item, with items presented in the same order in which they were presented to respondents completing the scale. The only exception is for the six pregnancy-related anxiety items. The analysis of these six items is presented separately at the end of this section, although conducted similarly to the other items. The full wording of a scale item is only reported when it was considered important in

order to comment on its content. For reasons of brevity, in other instances items will be referred to only using their number (e.g. item 1) indicating their position in the scale. When this is the case, please refer to Table 16 for the full item wording.

Item 1 – *"I worried more than usual"*

Item 1 did not exhibit a particularly high correlation with the total score, being only the 18th highest item-total correlation out of the 28 items (0.65). As previously noted, an inspection of endorsement frequencies showed that item 1 was also one of the five items with more than 50% of respondents (54.6%) endorsing *"3 = sometimes"* or more frequently as response option. Although this item was included in the initial item pool, based on the systematic review and qualitative interviews, as a potential indicator of Generalised Anxiety Disorder in which excessive worrying is a key feature (APA, 2013), its endorsement frequencies would appear to indicate that the item taps into a relatively common experience of pregnancy. Other items in the scale, such as item 18 *"Thoughts got stuck in my head"*, and item 25 *"My worries overwhelmed me"* seemed to assess a similar but more distressing cognitive process related to excessive or repetitive worrying. These two items also had higher item-total correlations (0.79 and 0.76 respectively) and more negatively skewed endorsement frequencies (i.e. *"Never"* and *"Rarely"* were chosen more frequently), expected and desirable qualities for items assessing psychological symptoms. For these reasons, the two alternative items were considered superior and item 1 was discarded in this first stage of analysis.

Item 2 – *"My anxiety stopped me from doing things"*

This item showed the 15th highest item-total correlation (0.71). Its distribution of endorsement frequencies indicated that 14.6% of women scored *"3 = sometimes"* or higher on this item. This item is clearly an indicator of a behavioural component of anxiety, related to avoidance of specific places or situations and more generally to a change of behaviours that may negatively impact on a person's daily functioning, a common feature of several anxiety disorders, including Agoraphobia, Panic Disorder, Social Anxiety Disorder and Specific Phobia (APA, 2013). The only other items within the 28-item scale capturing a behavioural component of anxiety are item 11 *"I felt that my anxiety made me act*

impulsively" and item 19 *"I avoided people".* However, the content of these items is rather specific and does not appear to significantly overlap with the content of item 2. Furthermore, these two items had inter-item correlations with item 2 within the desired range (i.e. $r \geq .20$ and $<.80$) and consequently did not show redundancy with item 2. Based on the above considerations, item 2 was retained at this stage.

Item 3 – *I had sudden feelings of panic*

Item 3 was found to have one of the lowest item-total correlations (0.60, 23rd higher). Simply based on this relatively low correlation with the total score, item 3 was considered a candidate for deletion. Furthermore, an alternative item (item 8 *"I felt panicky for no good reason"*) whose content appeared to measure a very similar facet of anxiety related to feelings of panic showed comparable endorsement frequencies but a significantly higher item-total correlation (0.71, 14th highest). Item 3 was thus discarded at this stage.

Item 4 – *I had a racing or pounding heart*

Item 4 was the only item within the 28-item SAAS referring to a physical symptom of anxiety. It showed one of the weakest item-total correlations (0.58, 24th highest) and only low or moderate inter-item correlations with the vast majority of other items within the scale (0.13-0.66) with inter-item correlations above 0.60 with only two other items in the scale, item 6 and item 28. Item 4 was consequently not considered for inclusion in a shorter version of the SAAS and discarded at the initial stage.

Item 6 – *My mind was racing*

This item also showed a relatively weak item-total correlation (0.62, 21st highest). It was also found to have the 4th highest mean score (2.55) and was one of the five items for which at least 50% of respondents scored *"3=sometimes"* or above (51.6%). Similarly to what was discussed for item 1 in relation to the poor evidence of discriminative ability because of its response distribution, item 6 was discarded at this stage.

Item 7 – I *was much more irritable than usual*

Item 7 also exhibited one of the lowest item-total correlations (0.44, 25th highest) and the lowest among items assessing general symptoms of anxiety as opposed to pregnancy-related anxiety items. It was also one of the items for which respondents endorsed a response option of 3 or above most frequently (61.8%) and had the highest mean score among all items (2.77). These observations clearly highlight the lack of discriminative power of this item. Considering in particular the distribution of endorsement frequencies, it can be reasonably assumed that item 7 tap into a rather common experience of pregnancy and cannot thus be considered a reliable indicator of poor mental health, or in particular problematic levels of anxiety. This item was thus discarded.

Item 8 – I *felt panicky for no good reason*

This item had the 14th highest item-total correlation (0.70) and a mean score of 1.79. As previously mentioned, the content of this item seems to overlap considerably with item 3 *"I had sudden feelings of panic"*, which was discarded because of inadequate item parameters. Item 8 would appear to enquire about a very similar facet of anxiety when compared with item 3. However, item 8 had a higher item-total correlation and the way it is worded makes it clear that the feelings of panic are not related to an objective risks (WHO, 1992). For this reasons, and for its moderately high item-total correlation, item 8 was retained for potential inclusion in the SAAS.

Item 9 – I *did not feel like myself*

Item 9 also showed a moderately high item-total correlation (0.70, 16th highest). Inter-item correlations were all in the moderate or moderately high range (0.30 – 0.80). At an initial examination of its content, this item would appear to enquire about a symptom of general distress rather than specific anxiety symptomatology. It was, however, included in the initial item pool as it emerged to be a common symptom among women with experience of antenatal anxiety who were interviewed as part of the process of scale development, and was also subsequently indicated as relevant by experts taking part in the Delphi study. The content of other items in the scale was inspected to examine whether any other items were similar in content to item 9. Only item 27 *"I felt like I was losing control"* had arguably a

degree of conceptual similarity with item 9, with a slightly higher item-total correlation (0.72). However, because item 9 had sufficiently robust psychometric properties and appeared to assess a relatively unique facet of the target construct, at this stage it was decided to retain it for further consideration, before making a final decision on its inclusion in the SAAS.

Item 11 – *I felt that my anxiety made me act impulsively*

This item was found to have a moderate item-total correlation (0.67, 17th highest). It also had one of the lowest mean scores (1.39, 5th lowest) and an inspection of its endorsement frequencies revealed that the vast majority of participants (93.8%) to the pilot study endorsed option *"1 = never"* or *"2 = rarely"* for this item. Although these responses may be expected in a scale assessing psychological symptoms, the proportion of women scoring one of the two response options was extremely high. Only 4 out of 62 participants scored higher than *"2 = rarely"* for this item. Simply based on these distribution of responses, and considering its only moderate item-total correlation, item 11 did not show sufficient discriminative ability in this phase of preliminary psychometric testing and was thus discarded.

Item 12 – *I found it hard to focus on anything other than my anxiety*

This item showed a good item-total correlation (0.75, 8th highest) and an average score of 1.40. Two items were found to correlate strongly (i.e. at least 0.75) with item 12. These were item 17 *"I felt like I needed help for my anxiety"* (0.75), and item 23 *"I felt extremely anxious"* (0.80). These three items would seem in fact to capture an analogous component of anxiety related to particularly elevated anxiety feelings and a perceived inability to manage these feelings. Two main reasons led to the exclusion of this item. Firstly, the two items (17 and 23) correlating strongly with item 12 had both higher item-total correlations, specifically the first and third highest correlations with the total score among all items. Consequently, these two items showed a closer association with the latent construct than item 12. Considering also the similar content of items 12, 17 and 23, including more than one of these items in the final, shorter version of the SAAS would almost inevitably lead to increased redundancy at the expenses of breadth of construct and content validity (Clark & Watson, 1991; DeVellis, 2012). In second place, item 12 is significantly longer than most

other items considered at this stage, being composed of twelve words. In scale construction, shorter items are almost always preferable as item brevity is associated with increased clarity for respondents. Item 17 and 23 have both higher item-total correlations than item 12 while also containing fewer words. Item 12 was therefore discarded at this stage of the analysis.

Item 14 – *I felt unable to cope*

Item 14 exhibited a good item-total correlation (0.75, 9th highest). Its endorsement frequencies were consistent with estimated prevalence for antenatal anxiety (15%), with 16.1% of respondents (n=10) scoring 3 or above. Inter-item correlations for this item were also all in the acceptable range. The item which is closer in content to item 14 is arguably item 29 *"I felt overwhelmed"*. These two items had essentially the same item-total correlation, 0.750 and 0.752 respectively. Accordingly, at this stage item 14 was retained as it showed a satisfactory psychometric performance and appeared to measure a component of general distress not evaluated by other items. Further examination and comparison with all other items retained at this stage was required in order to make a final decision.

Item 15 – *I felt something awful would happen*

Item 15 was found to have one of the lowest item-total correlations (0.61, 21st highest). Due to its comparatively weak correlation with the total score this item was a candidate for deletion. Another item with similar content, item 28 *"I had a feeling of impending doom"* showed a significantly higher item-total correlation (0.80, 4th highest). Considering the relatively low item-total correlation of item 15, as well as the presence of a similar item with better psychometric qualities, item 15 was excluded from further analyses.

Item 17 – *I felt like I needed help for my anxiety*

Item 17 was the item with the highest item-total correlation (0.86). As commented earlier, it is significant that the item with the highest correlation with the total score enquired about the perceived need for support around the respondent's anxiety symptoms, considering that the aim of the full scale was in fact to identify women experiencing problematic anxiety. Because of its excellent item-total correlation, this item was obviously considered as a

candidate for inclusion in the shortened, final version of the SAAS. However, inter-item correlations with the remaining 27 items were also inspected to investigate whether item 17 showed particularly high correlations with any other item, a possible indication of item redundancy (Streiner & Norman, 2008). Two items were found to correlate strongly with item 17, namely item 21 *"I could not control my anxiety"* (0.88) and item 23 *"I felt extremely anxious"* (.80). Notably, these two items also exhibited the second and third highest item-total correlations. Despite suspect redundancy, at this stage item 17 was retained for further consideration based on its excellent item-total correlation.

Of note, an inspection of the inter-item correlation matrix also revealed that three further items showed high correlation (i.e. equal or above 0.75) with item 17. Specifically, these were item 12 *"I found it hard to focus on anything other than my anxiety"* (0.75), previously discarded, item 27 *"I felt like I was losing control"* (0.76) and item 28 *"I had a feeling of impending doom"* (0.76). A closer examination of their content appeared to indicate that these items also shared common features with item 17, 21 and 23, all referring to feelings of lack of control over particularly distressing levels of anxiety. As discussed in Chapter 6, this component of antenatal anxiety concerned with problematic anxiety symptoms that are perceived as distressing and interfere with a person's daily functioning is an essential aspect of the target construct. Consequently, it was important to represent this feature of antenatal anxiety in the final version of the SAAS. However, including more than one of the items discussed above would have almost inevitably led to item redundancy, while narrowing the scope of the construct being assessed (Kline, 2000; Streiner & Norman, 2008). Specifically, among these five items, based on psychometric considerations it was deemed appropriate to select one of the three items with the highest item-total correlations (item 17, 21 or 23), as further discussed in this and the following section.

Item 18 – *Thoughts got stuck in my head*

Item 18 had the 5th highest correlation with the total score (.79). This item would appear to assess the cognitive process of rumination, a common symptom of Obsessive Compulsive Disorder related to persistent and recurrent thoughts (APA, 2013). Repetitive thoughts in the form of worries are also one of the key features of Generalised Anxiety Disorder (APA, 2013). Inter-item correlations were examined for this item and showed that the vast majority of correlations with other items were between 0.20 and 0.75, indicating that this item tapped

into a relatively unique facet of the target construct. Item 18 also had a moderately high correlation with item 17 *"I felt like I needed help for my anxiety"* (0.69), which provided further evidence that this item showed good discriminative ability without being redundant. It was thus decided to retain item 18 for potential inclusion in the scale based on its robust item statistics and relative uniqueness of content.

Item 19 – *I avoided people*

This item also showed one of the strongest item-total correlations (0.77, 7th highest) and a mean score of 1.71. As noted above, this was one of three items within this preliminary version of the SAAS assessing a behavioural symptom of anxiety. The other two were item 2 *"My anxiety stopped me from doing things",* which was retained for further consideration, and item 11 "*I felt that my anxiety made me act impulsively"* which was discarded. Item 19 and item 2 had a high inter-item correlation (0.74), although not at the level for which redundancy can be assumed (i.e. > 0.8). Item 19 is also a key diagnostic criterion of Social Anxiety Disorder (WHO, 1992; APA, 2013), in which avoidance of social situations is often marked. At this stage, it was thus considered appropriate to retain it for further analysis.

Item 20 – *I have felt so anxious that I had thoughts about terminating my pregnancy*

As previously discussed, this item was excluded at the initial stage of analysis as it showed no variance, with all 62 participants endorsing option *"1=never".*

Item 21 – *I could not control my anxiety*

Item 21 exhibited a particularly high item-total correlation (0.84, 2nd highest). The specific component of anxiety evaluated by this item was discussed in the section on item 17 *"I felt like I needed help for my anxiety".* It was noted that both items shared similarities in content, related to distressing and uncontrollable feelings of anxiety. As they showed respectively the first and second highest item-total correlation in this pilot study, at this stage they were both retained in the pool of items for potential inclusion in the final version of the scale.

Item 22 – *I felt so anxious that I had thoughts of ending my life*

As documented earlier, this item was also excluded at the initial stage of analysis because of extremely low variance. Only one out of sixty-two participants scored *"2=rarely"* and the remaining sixty-one scored *"1=never"*.

Item 23 – *I felt extremely anxious*

This item had the 3rd highest correlation with the total score (0.81). However, as previously noted, it also showed considerably high inter-item correlations with item 17 *"I felt like I needed help for my anxiety"* and item 21 *"I could not control my anxiety"*, respectively 0.80 and 0.87. Because of a high risk of redundancy among these three items, item 23 was discarded since it was the one among the three with the lowest item-total correlation.

Item 25 – *My worries overwhelmed me*

Item 25 was discussed in previous sections because of its apparent, partial overlap in content with item 1 *"I worried more than usual"* which was discarded, item 14 *"I felt unable to cope",* and item 18 "*Thoughts got stuck in my head* ", which were both retained at this stage of the analysis. Item 25 had the 6th highest item-total correlation (0.77) and 19.3% of participants scored *"3=sometimes"* or above for this item. This item seems to capture a symptom very common in Generalised Anxiety Disorder, an excessive and uncontrollable level of worrying. It is also important to note that various studies have indicated that GAD is the most prevalent anxiety disorder in pregnancy (Grant et al., 2008; Marchesi et al., 2016; Dennis et al., 2017). Another item, item 29 "*I felt overwhelmed"* enquires about a similar feeling without including the cognitive component of worrying. However, item 29 showed a marginally lower item-total correlation (0.75) than item 25. At this stage, because of its good psychometric performance, distinctiveness of content and centrality in a prevalent anxiety disorder such as GAD, it was considered appropriate to retain item 25 for further consideration.

Item 27 – *I felt like I was losing control*

Item 27 showed the 13th highest item-total correlation (0.72). As reported in Item 17, this item was discarded because of its high correlations with both item 17 (0.77) and 21 (0.75), coupled with a markedly lower item-total correlations compared to these two items.

Item 28 – *I had a feeling of impending doom*

Similarly to item 27, this item was discarded for reasons discussed in the paragraph for item 17. Specifically, although this item had the 4th highest item-total correlation (0.80), it correlated highly with both item 17 (0.76) and item 21 (0.78). Furthermore, this item had the second lowest mean score (1.31) among all items included in this analysis. A particularly low mean score may indicate issues with item response distribution (Furr, 2011). An examination of endorsement frequencies showed in fact that only 4 out of 62 respondents (6.4%) scored *"3=sometimes"* or above for this item. This would appear to be indicative of poor discriminative performance among respondents and provided a further reason for the exclusion of this item.

Item 29 – *I felt overwhelmed*

Item 29 showed a relatively high item-total correlation (0.75, 10th). Its mean score was 2.16, with more than one third of respondents (35.5%) endorsing the response *"3=sometimes"* or above for this item. Arguably the item closest in content to item 29 was item 14 "*I felt unable to cope*", which also showed an identical item-total correlation (0.75). Inter-item correlation for these two items was high (0.71), but not at the extent that would indicate that one of the two items was certainly redundant. For this reason, item 29 was retained at this stage of the analysis.

Item 30 – *I needed someone to support me with my anxiety*

This final item was found to have the 11th highest correlation with the total score (0.74). While this item parameter was adequate, when considering both the item-total correlation and its response distribution, item 17 and item 21 showed once again a superior psychometric performance. Their inter-item correlations with this item are high (0.73 for item 17 and 0.72

for item 21) and only marginally do not reach the criterion for suspected redundancy of > 0.75. Also in consideration of the remarkably better performance of items 17 and 21 in relation to their item-total correlations compared with item 30, this item was discarded.

Sub-analyses on the six pregnancy-related anxiety items

Initial inspection of the item-total and inter-item correlations for the six pregnancy-related anxiety items (item 5, 10, 13, 16, 24, 26) was carried out similarly to all the other items discussed above. As discussed in the Data analysis section, however, the six PrA items were also analysed as a separate subscale, in order to examine its internal consistency and determine which items provided the strongest contribution in relation to their psychometric performance in measuring pregnancy-related anxiety symptoms. In relation to the internal consistency specific to the six-item subscale, DeVellis (2012) notes that, when a distinct construct is considered possible within an item pool (i.e. more than one factor), the internal consistency of the items composing the suspected subscale should be examined. When the six PrA items were treated as a separate subscale, its internal consistency was found to be α = 0.77. This provided an initial confirmation that these items taken together assessed a unitary dimension.

An examination of item-total correlations of the six items with the 28-item scale and within this six-item PrA subscale was also considered important in determining the relative contribution of each of the items in assessing the dimension of pregnancy related anxiety. Table 17 presents the six anxiety items and their item-total correlations, both when considered as part of the full, 28-item scale and as a separate subscale.

Table 17 – Corrected item-total correlations for the six pregnancy-related anxiety items (28-item scale and 6-item scale)

Item number	Item content	Corrected item-total correlation (28-item scale)	Corrected item-total correlation (6-item PrA scale)
5	*"I felt detached from pregnancy and the baby"*	0.44	0.35
10	*"I have felt scared about giving birth"*	0.42	0.54
13	*"I worried about losing my baby"*	0.64	0.59
16	*"I worried that something may be wrong with my baby"*	0.63	0.61
24	*"I have had negative thoughts about childbirth"*	0.41	0.48
26	*"I did not feel worthy of being a mother"*	0.73	0.52

All correlations are reported as Pearson's *r*.

Following an examination of the item-total correlations reported in Table 17 and of other item parameters such as endorsement frequencies and inter-item correlations among the six PrA items, three of the six PrA were retained for further consideration. This decision was informed both by considerations based on the construct relevance of these PrA items and on their psychometric performance, as briefly discussed below.

At an initial stage of analysis of the six-item PrA subscale, item 5 *"I felt detached from pregnancy and the baby"* was discarded because of its particularly poor psychometric parameters. It exhibited the lowest item-total correlation within the PrA subscale, and it was the only item that increased the internal consistency of the six-item PrA subscale when not considered. Moreover, inter-item correlations with all other PrA items were all surprisingly

low (all below 0.25) except than with item 26 (0.46). Conversely, also in the initial phase of analysis of subscale it was decided to retain item 26 *"I did not feel worthy of being a mother"* for further consideration. This item showed moderately good item-total correlation both within the PrA subscale and when considered as part of the full scale. Moreover, item 26 had the highest item-total correlation with the full scale among the six PrA items. Its endorsement frequencies were also consistent with a measure assessing problematic symptoms, with 14.5% of respondents scoring *"3 = sometimes"* or higher for this item.

The remaining four PrA items measured two relatively distinct facets of pregnancy-related anxiety. Item 10 *"I have felt scared about giving birth"* and 24 *"I have had negative thoughts about childbirth"* refer to fears and negative thoughts about childbirth, while item 13 *"I worried about losing my baby"* and 16 *"I worried that something may be wrong with my baby"* assess specific worries about the unborn baby. In support of this distinction, item 10 had the highest inter-item correlation with item 24 among all items (0.59) and item 13 showed a similar pattern with item 16 (0.69). While these items appear to tap into two core facets of pregnancy-related anxiety, in order to avoid redundancy in the final version of the SAAS, it was considered appropriate to discard one item from each of these two specific components of PrA. Between item 10 and 24, item 24 was preferred for various reasons. These items had comparable item-total correlations. However, an inspection of their endorsement frequencies revealed that item 10 was scored *"3 = sometimes"*, *"4 = often"* or *"5 = always"* by 63% of women taking part in this pilot study. Similarly to other items previously discussed, it can be argued that this distribution of responses suggest that some level of fear around childbirth is considerably common among pregnant women, and that the way in which the question is formulated does not yield sufficient discriminatory power. Item 24, for comparison, was scored 3 or higher only by 24 % of respondents, suggesting a superior discriminative accuracy than item 10. Only item 24 was thus retained at this stage.

Finally, item 13 *"I worried about losing my baby"* and 16 *"I worried that something may be wrong with my baby"* were also examined and compared. Similarly to the two previous items, item 13 and 16 showed very similar item-total correlation. However, they also showed comparable endorsement frequencies. An examination of their content suggested that item 13 focused specifically on worries regarding the possibility of miscarriage, while item 16 appeared to assess a broader range of negative thoughts related to the health of the baby. Primarily for this reason, item 16 was retained for further consideration while item 10 was discarded.

Comments provided by study participants

A free text box was provided at the end of the questionnaire, and study participants were asked to indicate whether they found any of the questions unclear or to provide any other comments they had about the scale. Before presenting the stage of the selection of 10 items for the final version of the scale, it is important to discuss the comments that women provided, as they also led to a modification in the time frame assessed by the scale, which in this version asked about symptoms experienced in the past week. Below a selection of relevant quotes from study participants is reported.

P03 - *"If the questions were about the last e.g. 2/3 weeks the answers may have been different. The last 7 days I've felt less anxious than previously"*

P20 – *"I feel this questionnaire is too long to be routinely included in antenatal care – the midwife appointments are already quite "full", could it be reduced to ~ 15 questions?"*

P32 – *"– It is difficult to focus on the past 7 days of pregnancy rather than the past few weeks"*

P40 – *"My anxiety is related to a previous miscarriage"*

A number of other study participants commented positively on the scale, such as P12: *"Questions all clear and straightforward"* or P34: *"My feelings (anxiety) don't last + I can control/realise this is likely normal. I think it would be good to complete a questionnaire like this to recognise if I was not coping. I don't think I have completed any questionnaire about my mental health through pregnancy + this would be a good addition to support offered in case I needed help."*

A participant commented specifically on two items, item 20 *"I have felt so anxious that I had thoughts about terminating my pregnancy"* and item 22 *"I felt so anxious that I had thoughts of ending my life"*. The comment noted that: *"Qs 20 and 22 I thought were very confronting and I was much happier responding to say Q23. Do all expectant mothers need*

to be asked Q20 and 22 or are these more relevant to those noted as having heightened anxiety?" (P27). As documented earlier, these two questions were both discarded in the initial stage of the item selection process because of zero or extremely low variance.

As a result of the comments provided by women, it was considered appropriate to change the time frame assessed by the scale from 'the past 7 days' to 'the past 14 days' in its final version. Based on women's considerations, and on the observation that anxiety levels during pregnancy might significantly vary as a response to specific situations (e.g. a problematic scan), a time frame of two weeks was considered more suitable to assess problematic anxiety.

7.5 Final selection of scale items for inclusion in the 10-item SAAS

As indicated earlier, following the stage presented above in which items were discarded based on a number of considerations related to their psychometric performance, content and contribution to the assessment of the target construct, a total of 13 items were retained for further consideration. Three of the 13 items focused on pregnancy-related anxiety symptoms, while the remaining 10 items assessed general symptoms of anxiety. The rationale for aiming at a scale containing no more than 11 items was discussed in detail in Chapter 6, where it was noted that NICE recommends that exclusively short scales should be used to screen for perinatal mental health problems (i.e. less than 12 items). A 10-item scale was considered optimal, as it included a sufficient number of items to represent a range of facets of the target construct (Cox et al., 1987). In order to select the 10 items for the final version of the SAAS, the psychometric performance and content of the 13 items retained at this stage were all further examined and discussed with my supervisory team.

First, it was considered essential to include in the 10-item SAAS at least a small number of items specific to pregnancy-related anxiety. The importance of the component of pregnancy-related anxiety in the construct of antenatal anxiety was documented in the systematic review of the psychometric literature on antenatal anxiety and further supported by the qualitative interviews presented in Chapter 5. It was consequently considered theoretically and empirically important to include items tapping into this dimension of antenatal anxiety in the final version of the SAAS. Some authors have observed that a minimum of three items are typically required to form a separate dimension or factor within a scale with sufficient structural validity (Kline, 2000; Swalm et al., 2010). As a dimension within a scale cannot typically be composed by less than three items, and in consideration of the relevance of the dimension of pregnancy-related anxiety as documented in the research literature (Huizink et al., 2004; Blackmore et al., 2016) and summarised in this thesis, it was thus decided to retain the three PrA items identified above as those with the strongest psychometric performance. In relation to the remaining 10 items assessing symptoms of anxiety not specific to pregnancy, the following three items were excluded in this final stage of item reduction:

- o Item 9 *"I did not feel like myself":*

As noted earlier, this item appears to assess general distress rather than specific anxiety symptomatology. Among the other nine items, the only other items which appeared to measure a more generic component of psychological distress were item 14 *"I felt unable to*

cope” and item 29 *“I felt overwhelmed”*. Item 29 is discussed below. In relation to item 14, this item showed a significantly higher item-total correlation than item 9 and, as previously discussed, distribution responses more consistent with problematic anxiety symptoms. Item 9 was thus discarded in this final stage favour of item 14, which showed superior item parameters and was consequently retained as a possible indicator of general distress.

- o Item 17 *“I felt like I needed help for my anxiety”:*

Despite having the highest item-total correlation (0.86) among the 28 items, this item was discarded at this final stage of analysis. Item 21, *“I could not control my anxiety“*, with the second highest item-total correlation (0.84) was preferred for two main reasons. As noted earlier, the two items also showed a very strong inter-item correlation (0.88), essentially indicating that a person scoring high on one of the two items was also very likely to score high on the other item. Based on previous psychometric considerations with regard to item redundancy, it was thus deemed appropriate to include only one of the two in the final version of the scale. The content of item 17, one of the items from the PROMIS anxiety item bank, clearly indicated a need for support around problematic symptoms. However, in consideration of the well-documented stigma associated with perinatal mental health problems (Buist et al., 2015), the wording of this item might have led a proportion of women not to answer openly for fear of disclosing their need for support. On the other hand, item 21 *“I could not control my anxiety“*, while having only a marginally lower item-total correlation, was considered to be less direct. Moreover, item 21 comprised less words than item 17, a desirable feature in relation to item clarity and comprehensibility. Item 17 was thus discarded in this final stage of item reduction.

- o *Item 29 “I felt overwhelmed”:*

Item 29, as reported earlier, had the 10th highest item-total correlation (0.75), identical to item 14 *“I felt unable to cope”*. It was considered appropriate to include only one of these two items in the final version of the SAAS as a possible indicator of general distress. A decision between Item 14 and item 29 was guided primarily by the observation that another of the remaining items, item 25 *“My worries overwhelmed me”,* partially overlapped in content with item 29. While feeling overwhelmed was a shared feature of these two items, item 25 was considered to be more specific to anxiety symptomatology, with the inclusion

of the component of worry, and was thus preferred to item 29. As item 29 was discarded in this final stage, it was consequently decided to retain item 14 to represent the component of general distress in the final version of the SAAS.

In conclusion of this chapter, Figure 4 in the next page presents the 10 remaining items which were included in the final version of the SAAS. The scale is shown in the format in which it was presented to study participants in the psychometric validation study discussed in the next, final experimental chapter.

Figure 4 - 10-item, final version of the Stirling Antenatal Anxiety Scale (SAAS)

STIRLING ANTENATAL ANXIETY SCALE

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

Chapter 8   Preliminary psychometric validation of the SAAS

8.1 Introduction

As discussed in the previous chapter, once a final, 10-item version of the SAAS was produced, this was psychometrically tested on a larger sample of pregnant women to evaluate a range of psychometric properties. The primary aim of the psychometric validation study presented in this chapter was to evaluate the screening accuracy of the final version of the SAAS by validating it against a structured diagnostic interview for anxiety disorders. One hundred and seventy-four women completed the SAAS and other self-report scales to assess depression and anxiety symptoms. Additionally, a subsample of 37 women were also assessed using a brief diagnostic interview, which served as the 'gold standard' in this psychometric validation study. This procedure is documented in detail in the following sections. The internal consistency, convergent and discriminant validity, and factor structure of the SAAS were also evaluated in this preliminary psychometric validation of the scale. The word 'preliminary' is used here to indicate that psychometric testing and validation of a scale is an ongoing process, in which evidence of scale reliability and validity needs to be supported by a number of studies testing the scale in a range of samples selected from the intended population of respondents (Abell, 2009; DeVellis, 2012). Consequently, while making any final claims regarding the psychometric properties and screening accuracy of the SAAS would be premature, the psychometric validation study presented here provided preliminary evidence of its psychometric performance in a relatively large sample from the target population.

Two additional scales were also completed by women as part of this validation study. These were the GAD-7 (Spitzer et al., 2006) and the Edinburgh Postnatal Depression Scale [EPDS] (Cox et al., 1987). The GAD-7 was included for two main reasons. The first, and most important, was discussed earlier in the thesis (see section 1.2.2). The GAD-7 (and the ultra-brief screener GAD-2, consisting of the initial two questions of the GAD-7) are the NICE-recommended measures to screen for anxiety during the perinatal period (NICE, 2014). It was thus considered important to evaluate and compare the screening accuracy of both the GAD-2 and the GAD-7 in a sample of pregnant women, particularly in light of the fact that there is very limited evidence supporting their screening accuracy for problematic anxiety

in perinatal populations (Zhong et al., 2015). A secondary reason was that the GAD-7 was used to evaluate the convergent validity of the SAAS. The EPDS, a well-established measure of perinatal depression also previously discussed in the systematic review chapter, was used to assess the discriminant validity of the SAAS.

8.2 Study aims

The primary and secondary aims of this preliminary psychometric validation of the SAAS are presented here.

Primary aim:
To validate and compare the SAAS, GAD-2 and GAD-7 against a gold standard structured diagnostic interview for anxiety disorders, by evaluating their screening accuracy and determining their optimal cut-off scores in a population of pregnant women.

Secondary aims:

- To evaluate the internal consistency of the SAAS and the GAD-7
- To assess the convergent and discriminant validity of the SAAS, GAD-7 and EPDS
- To examine the factor structure of the SAAS
- To determine the screening accuracy of the SAAS, GAD-2 and GAD-7 in identifying women experiencing an anxiety disorder and/or pregnancy-related anxiety symptoms
- To evaluate the acceptability and ease of completion of the SAAS in a sample of the target population

8.3 Study design

The COSMIN criteria for evaluating whether a psychometric validation study meets the standards for good methodological quality (Terwee, 2007; Mokkink et al., 2010b) were used to inform the study design and ensure that the study met the criteria for excellent methodological quality in the evaluation of a range of psychometric properties. These criteria were previously introduced in the systematic review chapter to assess the methodological quality of included studies (4.2.3). Specifically, the Consensus-Based Standards for selection of Health Measurement Instruments (COSMIN) checklist (Mokkink et al., 2010b) was used in the validation study for this purpose. The four 'boxes' of the COSMIN checklist that were used to provide evidence of the methodological quality of the study were box B *"Reliability"*, box E *"Structural validity"*, box F *"Hypotheses testing"*, and box H *"Criterion validity".* Each box includes a variable number of items (between 5 and 18) that are used to assess the quality of study design in relation to a specific psychometric property (Terwee et al., 2012). As in the COSMIN checklist a quality score for each box is determined by taking the lowest score of any item in a box (i.e. 'worse score counts' approach), the study was designed and conducted so that it scored excellent for methodological quality for all the items in each of the four boxes evaluated.

The recruitment and data collection phases of this validation study followed the same procedure of the pilot study, with a cross-sectional postal survey method used to collect data from study participants. However, this was also a psychometric study of screening accuracy and there were thus also some substantial differences. In studies of diagnostic or screening test accuracy, the result of the test or scale of interest is compared against the reference standard, which can be defined as the best available method for establishing the presence or absence of the target condition (Furr, 2011). For the identification of psychological symptoms, the reference or gold standard is considered to be a structured clinical interview based on well-established diagnostic criteria (Gibson et al., 2009; APA, 2013). The most common study design for studies of screening accuracy involves systematic comparisons of the results of the index test (i.e. the scale) and those of the reference standard in the same subjects, in order to determine the screening or diagnostic accuracy of the scale under investigation (Centre for Reviews and Dissemination [CRD], 2008). The main difference in study design between the pilot and the validation study was thus that the validation study required a subsample of women who completed the scales to be additionally assessed with a structured diagnostic interview. Primarily because of the diagnostic interviews, a number

of aspects related to sample size, data collection and analysis differed from the pilot study and are thus discussed below in relevant sections. However, the setting, sampling method, eligibility criteria, recruitment procedure and part of the data collection process were identical in the two studies. Please refer to the corresponding sections of the pilot study in Chapter 7 for these aspects of study design.

8.3.1 Sampling, recruitment and data collection

As noted above, the setting, study eligibility criteria, recruitment procedure and part of the data collection process were the same as for the pilot study. A further meeting took place with eight Senior Charge Midwives (SCMs) in February 2018, two weeks before starting recruitment for this validation study. In this meeting 300 study booklets were distributed among four different teams. The results of the pilot study with regard to recruitment were discussed. It was also clarified that the recruitment procedure for the validation study was identical, with the exception that in this study midwives, when presenting the study to pregnant women, were required to include information about the possibility of also being invited to take part in a telephone clinical interview. SCMs were asked to share this information with all midwives participating in recruitment. Furthermore, copies of the information sheet *"Information for midwives: How to recruit women into the study"* (Appendix 13), which included a brief script to explain the study that midwives could use to describe the study to women, were also given to SCMs. A number of further study booklets were distributed in three subsequent meetings with SCMs over the course of the following five months, as further documented later. The sections below discuss considerations related to sampling and sample size, which differed from the pilot study, as well as the phase of data collection with regard to the structured diagnostic interviews conducted as part of this psychometric validation study.

Sampling and calculation of sample size

As was discussed for the pilot study, the initial intention was to recruit a similar number of women representing the three trimesters of pregnancy (i.e. 60-70 women per trimester, target $n = 200$). However, over the course of the pilot study, it became clear that it was not feasible to recruit an equal representation of women across the trimesters in the sample. A

convenience sampling technique was adopted. The target sample size for this psychometric validation study ($n = 200$, with a subsample of $n = 60$ assessed via a structured clinical interview) was determined in consultation with a statistician while planning the phases of data collection and analysis, as further detailed below. A sufficiently large sample size was required for the purpose of evaluating the screening accuracy of the SAAS, GAD-2 and GAD-7, as certain requirements need to be met in studies of test accuracy (CRD, 2008; Furr, 2011). A formula proposed by Buderer (1996), specifically devised to calculate sample size requirements in studies evaluating the sensitivity and specificity of a screening or diagnostic test, was used for the calculation of the target sample size. This method of determining sample size requirements takes into account a clinically acceptable precision for the estimates of sensitivity and specificity, expected values of sensitivity and specificity and the estimated prevalence of the condition (anxiety disorders) in the target sample (pregnant women). The screening accuracy of the scales could only be calculated on the subsample of women who were assessed both by completing the scale and with the diagnostic interview. It was calculated that a subsample of 60 women drawn from an initial sample of 200 study participants was sufficiently large, based on an expected prevalence of 50% (i.e. 30 women in the subsample expected to have the target condition, as detailed below), expected optimum sensitivity and specificity of 90%, a maximum acceptable width of the confidence interval of 10% (i.e. confidence intervals for sensitivity and specificity from 85% to 95%), and a two-sided significance level of 5% (Buderer, 1996). In order to achieve a prevalence of the target condition of 50% in the subsample (i.e. 30 women with an anxiety disorder), it was expected that a target sample size of 200 women would have included approximately 30 women with an anxiety disorder, based on estimated prevalence for anxiety disorders in pregnancy of approximately 15% (Heron et al., 2004; Dennis et al., 2017). It was consequently planned to purposively select, for the subsample of 60 women, the 30 women with scores in the highest range of the GAD-7 (i.e. 7 or above: Zhong et al., 2015) and a further 30 women to represent a range of low and moderate scores on the scale (i.e. 15 with scores ranging from 0 to 3 and 15 in the 4-6 score range). GAD-7 scores were used to select women to be invited for the diagnostic interview as it was, at this stage, the only validated anxiety scale that women completed as part of the study. The decision to maximise the presence of study participants with the target condition in this subsample drew on procedures commonly used in psychometric validation studies (Cox et al., 1987, Sackett & Haynes, 2002). The sampling method used to select the subsample also offered the additional advantage of minimising the number of women that were assessed with a structured clinical

interview, while allowing a statistically valid calculation of sensitivity and specificity by using the formula discussed above.

Measures

Women who participated in the validation study were asked to complete three self-report scales, specifically the 10-item SAAS, the GAD-7 and the EPDS. The item set, response options and scale format of the SAAS were extensively discussed earlier in this thesis, and are thus not repeated here. The rationale for choosing the GAD-2/7 and EPDS as additional scales was also provided earlier. Their strengths and limitations were discussed in detail in the systematic review in Chapter 4, and subsequently here only a brief reminder of their key features is provided. All items of the SAAS, GAD-2/7 and EPDS can be seen in Appendix 16.

The GAD-7 includes seven items related to symptoms of Generalised Anxiety Disorder (Kroenke et al., 2007). Respondents are asked to indicate how frequently they "have been bothered" by each of the symptoms over the past two weeks, with all items scored on a 4-point Likert scale (0 = *not at all*, 1 = *several days*, 2 = *more than half the days*, 3 = *nearly every day*). Scores range from 0 to 21, with higher scores indicating more severe symptoms. The first two questions of the scale are known as GAD-2 and, as previously noted, NICE guidelines recommend midwives consider asking these two questions to screen for anxiety in perinatal women (NICE, 2014). If a woman scores 3 or above on the GAD-2, NICE recommend that the GAD-7 should be used for further assessment. To our knowledge, the only study in which the screening accuracy of the GAD-7 was examined in pregnant women was conducted by Zhong and colleagues (2015). The authors found that an optimal cut-off score of seven or above, notably different from cut-offs of eight or ten identified in the general population, yielded good sensitivity (73%) and moderate specificity (67%). This cut-off score ($\geq$ 7) was thus used in the selection of a proportion of the subsample for the diagnostic interview.

The EPDS is a widely used and well-validated 10-item scale for the assessment of depression in the perinatal period (Cox et al., 1987; Murray & Cox, 1994). Although the scale appears to contain a 3-item anxiety subscale, the EPDS has very good sensitivity and specificity in identifying women experiencing perinatal depression (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.

The M.I.N.I. International Neuropsychiatric Interview (M.I.N.I.) 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). An average administration of the M.I.N.I is estimated to take approximately 15 minutes. The anxiety modules of the MINI PLUS version 5.0 (Sheehan et al., 1998), 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. All diagnostic interviews were administered by a researcher, as further detailed below.

Other information collected by all women taking part in the validation study (as shown in Appendix 16) included their age, gestational week, score on a single pregnancy-related anxiety item, parity status, history of obstetric complications, ethnicity, educational level and marital status. The ease of completion and acceptability of the SAAS to women completing the scale were also evaluated through two questions, as detailed in the Results section.

Recruitment and data collection

Recruitment took place between February and July 2018, with the last study questionnaires returned to the study office in August. The recruitment procedure was equivalent to that of the pilot study. Women meeting the inclusion criteria were verbally informed about the study by midwives while attending routine antenatal clinics in the Glasgow area. The study pack given to potential participants contained documents similar to those used in the pilot study, including an invitation letter, information sheet, study questionnaire (consisting this time of the SAAS, GAD-7, EPDS, and demographic and obstetric questions) and consent form. However, all the study documents were modified to reflect the different study design of the validation study (for examples see Appendices 3, 16 and 17). In addition, only for the psychometric validation study, the study booklet also contained a reply slip which was used to contact the subsample of women who were selected for the clinical diagnostic interviews. All women who participated in the study returned completed study questionnaires, consent

forms and reply slips by post. Once study booklets were returned to the study office, a proportion of women were selected to be invited for the diagnostic interview, according to the procedure discussed earlier to represent a range of GAD-7 scores. All interviews were telephone-administered and conducted by a researcher from the NMAHP-RU at the University of Stirling, who received training to administer the M.I.N.I and was blind to the scores of the scales. Contact details of women selected for the interviews were securely passed to the researcher, who contacted study participants by email or phone, depending on what they indicated in the reply slip, for the purpose of arranging a telephone interview. If a woman was still willing to take part in the study at this stage, a mutually suitable date and time was agreed. An attempt was made to conduct all interviews within four weeks from completion of the scales. In all cases in which this was not possible, the SAAS and the GAD-7 were re-administered by the researcher at the end of the interview, in order to allow a clinically meaningful comparison between scale scores and the diagnostic interview. In the majority of cases in which the interview was completed within four weeks, the researcher conducting the diagnostic interviews was blind to all the original scale scores. Results of the clinical interviews were then securely transferred to the author. Only unique ID numbers were used for this purpose.

Further ethical considerations

With regard to ethical aspects specific to the psychometric validation study, the information sheet and consent form given to all women taking part in this final study made it clear that a proportion of women taking part in the research would also be contacted and invited to attend a brief clinical interview. Although women had already given consent to take part in this interview by signing and returning the consent form for the validation study, on the day agreed for the interview the researcher who phoned all study participants confirmed consent verbally before commencing the interview. The researcher was additionally instructed to contact me, with the participant's permission, if a woman expressed the need for psychological support during the interview or disclosed information indicating risk to herself or others. In this eventuality, a variety of referral options to appropriate pathways of care were available by contacting Elaine Clark (NHS GG&C Perinatal Mental Health service), a study collaborator who was able to discuss with participants a support plan if required.

8.3.2 Plan for data analysis

A data analysis plan for the psychometric validation study was developed in consultation with a statistician. The procedure for data entry was identical to the pilot study, with all study data (item scores for the three scales, respondents' demographic and obstetric information, and two questions on the ease of completion and acceptability of the SAAS) initially entered onto a Microsoft Excel spreadsheet and subsequently transferred to SPSS software (version 23). A codebook was developed which listed all variables, specifying their level of measurement and indicating what each numeric code for a given variable represented (Pallant, 2013). Two dichotomous variables indicated whether women were part of the subsample who attended the diagnostic interviews, and whether they were identified as experiencing an anxiety disorder or not. Nominal variables (e.g. parity, ethnicity) were examined and are summarised narratively and through frequency distributions in the Results section. In relation to scale items, there is a long-standing debate on whether items in a Likert scale should be treated at the ordinal or interval level. However, for the purpose of both item analysis and the calculation of a range of psychometric properties in scale development and validation, it is a common and accepted practice to treat Likert items at the interval level (Spector, 1992; Bowling, 2009; DeVellis, 2012). In both the pilot and the validation studies, all scale items were thus treated as interval variables ('*Scale*' in SPSS).

Data cleaning and missing data

Before conducting any analysis on the data, a number of quality and accuracy checks on the dataset were planned and conducted in SPSS, to check for any clear mistakes in data entry, inconsistencies and out-of-range values. For categorical variables, responses were checked for unusual or impossible values. For continuous variables, frequency distributions and box plots were visually inspected to identify any mistakes in data entry and outliers. As many statistical techniques are sensitive to outliers (i.e. cases with values considerably above or below most of the other cases), these were examined through histograms and examination of the tail distributions, boxplots, and range checks. Some authors advocate removing all extreme outliers from the analyses (Williams, 2010), while others suggest that genuine outliers (i.e. not clear mistakes in data entry) should be retained or changed to less extreme values (Tabachnick& Fidell, 2007). In this study, all outliers were inspected but no evident

mistakes in data entry were found and all outliers were consequently considered in the analysis.

With regard to missing data, for all nominal and ordinal variables such as ethnicity and educational level, when data were missing this was simply reported in the Results section. In relation to item scores, a rule suggested by Wilson & MacLean (2011) was applied. If in a scale more than 20% of responses were missing, this was excluded from the analysis. If 20% or less of responses were missing, the missing score or scores were substituted with the median score for that participant on all other items in the scale, and documented in the Results as per recommended practice (Mokkink et al., 2010b).

Sample characteristics and respondents' feedback on SAAS

Descriptive statistics were used to calculate and summarise respondents' demographic and obstetric characteristics, including measures of central tendency and dispersion, and response distributions. Continuous variables such as age and weeks of pregnancy were summarised and reported using means, standard deviations and ranges. Categorical variables such as ethnicity were inspected by examining frequencies of response distributions, as reported in the Results. 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.

Primary analyses

Consistently with the study aims, a number of analyses were conducted to examine different psychometric properties of the SAAS and the GAD-2/7 (the EPDS was only used to assess convergent validity). The primary aim of this psychometric validation study was to assess the screening accuracy (i.e. criterion validity) of the SAAS, GAD-2 and GAD-7 in identifying women experiencing an anxiety disorder, as determined by M.I.N.I diagnoses. Additionally, the screening accuracy of the SAAS and GAD-2/7 in identifying women experiencing an anxiety disorder and/or pregnancy-related anxiety was also examined. Other important measurement properties assessed in this psychometric validation study included the internal consistency of the SAAS and the GAD-7, the convergent and discriminant validity of the three scales, and the factor structure of the SAAS, as detailed below.

Screening accuracy

As briefly noted earlier, in studies of screening or diagnostic accuracy of psychological scales, the screening performance of a scale is tested by administering both the clinical "gold standard" and the scale to the same subjects recruited from the intended population of respondents. The reference standard typically produces a dichotomous outcome (e.g. the subject either has or does not have the condition of interest). The outcome of the scale, on the other hand, is a continuous variable (i.e. total scale score), and thus a range of possible cut-off scores for the scale are considered. Specifically, numbers of true positives (TP: those who have the condition as determined by the reference standard, and test positive on the scale at a given cut-off), true negatives (TN: those who do not have the condition and test negative), false positives (FP: those who do not have the condition and test positive) and false negatives (FN: those who have the disease and test negative) are typically calculated for a range of cut-off scores (Clark & Watson, 2003, Abell, 2009). For the purpose of calculating the parameters above, a number of possible cut-off scores for the SAAS, GAD-2 and GAD-7 considered. Subsequently, 2x2 contingency tables were populated (CRD, 2008). These tables are used to describe the relationship between the outcome of the reference standard and the results of a scale at a given cut-off score, in terms of proportion of TP, TN, FP and FN, as illustrated in Figure 5.

Figure 5 – 2x2 contingency table (adapted from CRD, 2008)

	Reference standard	
Scale	Positive	Negative
Positive *(at a given cut-off score)*	True Positives	False Positives
Negative *(at a given cut-off score)*	False Negatives	True Negatives

A number of 2x2 contingency tables were generated, allowing the calculation of the following parameters used to determine the screening accuracy of the SAAS, GAD-2 and GAD-7 in identifying women experiencing an anxiety disorder. Sensitivity and specificity were the two main outcome measures used to evaluate the screening accuracy of the SAAS, GAD-2 and GAD-7. The positive and negative predictive values, other similarly informative indexes, were also calculated and reported. Definitions of these parameters of screening accuracy and the procedure used to calculate these indexes using 2x2 contingency tables are provided here:

Sensitivity: The proportion of people with the target condition (as diagnosed with the reference standard) who have a positive test result (i.e. score above a given cut-off score). Sensitivity is also known as true positive rate.

Specificity: The proportion of people without the target condition who have a negative test result. Specificity is also known as true negative rate.

Positive predictive value (PPV): The probability of having the target condition given a positive test result.

Negative predictive value (PPV): The probability of not having the target condition given a negative test result.

Sensitivity, specificity, positive (PPV) and negative (PNV) predictive values were calculated through these simple formulas (CRD, 2008):

$$\text{Sensitivity} = \frac{TP}{TP + FN} \qquad \text{Specificity} = \frac{TN}{TN + FP} \qquad PPV = \frac{TP}{TP + FP} \qquad NPV = \frac{TN}{TN + FN}$$

Values of sensitivity, specificity, PPV and NPV are commonly multiplied by 100 to give percentages (Kline, 2005). 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 values of sensitivity, specificity, PPV and NPV were also

calculated, using the formula for the standard error of a proportion as indicated by Harper and Reeves (1999).

The Area Under the ROC curve (AUROC), a parameter of overall scale accuracy, was also calculated for the SAAS, GAD-2 and GAD-7. The ROC curve provides a graphical display of all the possible combinations of sensitivity and specificity as a result of varying the cut-off score in a scale, and the AUROC provides a single index of overall diagnostic performance (Brace et al., 2013). The value of the AUROC can vary between 0.5 and 1, with a value of 0.5 indicating that a scale does not differentiate at all between subjects with and without the target condition and a value of 1 indicating a perfect scale which screens positive all those with the target condition and negative all those without the target condition. A value of 0.90 or above is considered excellent, with an AUROC between 0.80 and 0.90 indicating good discriminative accuracy, and values lower than 0.80 suggesting only a moderate accuracy (Pallant, 2013). Analyses of the AUROC are also particularly useful when comparing the screening performance of two or more scales. In this study, the AUROC of the SAAS, GAD-2 and GAD-7 were calculated and examined, as discussed in the Results.

Internal consistency of the SAAS and GAD-7

An initial inspection of item parameters for all SAAS items was conducted using descriptive statistics (response distributions, means and standard deviations) 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. All inter-item and item-total correlations were also inspected. The criteria presented in the pilot study to evaluate these parameters were applied (floor/ceiling effect $\geq$ 90%, item-total correlations $\geq$ 0.30, inter-item correlations $\geq$ 0.20 and $\leq$ 0.80).

Similarly to the pilot study, scale reliability was assessed by examining the internal consistency of the SAAS and the GAD-7 as measured by Cronbach's Alpha. It was noted earlier that various authors (Abell, 2009; DeVellis, 2012) indicate that the internal consistency of a scale developed for use in healthcare settings to inform clinical decisions about individuals should ideally approximate a value of $\alpha = 0.90$ (Kline, 2005; DeVellis, 2012). Values of Cronbach's Alpha much higher than this ($\alpha = \sim 0.95$) are likely to indicate item redundancy (Terwee et al., 2007). Nunnally's criterion (1978) of $\alpha > 0.70$ indicating

good internal consistency, and α > 0.80 indicating very good internal consistency are also often mentioned in the literature (Streiner & Norman, 2008).

Convergent and discriminant validity

As discussed in the Method chapter, convergent and discriminant validity are two psychometric properties that are commonly used to assess the construct validity of a scale (Abell et al., 2009). In the COSMIN checklist, these two psychometric properties are included under the *'hypotheses testing'* box. It is expected that a scale will correlate in the moderate to high range with scales purported to measure similar constructs and will have lower correlations with scales measuring constructs that are partially or fully unrelated to the construct of interest (Furr, 2011). Correlation coefficients were thus used to describe the magnitude of the relationship between two or more variables (e.g. respondents' scores on two scales). 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 (≥ 0.50).

Convergent validity: The correlation between SAAS and GAD-7 scores was calculated to test convergent validity. Although the GAD-7 was developed as a measure of a specific anxiety disorder (Generalised Anxiety Disorder), it is a popular measure of anxiety which has been used to assess a range of anxiety disorders in multiple populations (Spitzer et al., 2007; Plummer et al., 2016). As both the SAAS and the GAD-7 aim to assess problematic anxiety symptoms, a large correlation between the two scales was hypothesised, in the range 0.60-0.80.

Discriminant validity: The correlation between SAAS and EPDS scores was examined to test discriminant validity. As discussed earlier, the majority of EPDS items assess symptoms of depression (although a 3-item anxiety subscale was also identified). The EPDS was chosen as a well-established measure of perinatal depression, with the hypothesis that the two scales measured partially overlapping but distinct constructs. However, considering the evidence regarding the 3-item anxiety subscale within the EPDS, and the well-documented comorbidity between anxiety and depressive symptoms, a medium to large correlation was hypothesised between the two scales, in the range 0.40-0.60. The correlation between the EPDS and the GAD-7 was also calculated.

Pearson's *r* (Pearson product-moment correlation coefficient) is often reported in the psychometric literature as measure of the correlation between two scales (Pallant, 2013) and it was thus the primary choice for reporting scale correlations. However, this correlation test assumes that the data approximate the normal distribution and should not be used if this assumption is not met (Brace et al. 2013). Considering that scales measuring psychological symptoms are often characterised by a positively skewed distribution of scores (i.e. most people report few or no symptoms), the assumption of normality was assessed by calculating the Kolmogorov-Smirnov statistic. Normality is determined by a value of more than .05, indicating a non-significant result (Pallant, 2013). If total scores were found to be not normally distributed, Spearman's correlation (rho or r_s), the non-parametric equivalent of Pearson's *r*, was planned to be used. Spearman's correlation is based on the conversion of scores into ranked variables and is thus affected to a much lesser degree by outliers and non-normal distributions (Brace et al., 2013).

Factor analysis (Structural validity)

The rationale behind a range of data reduction techniques commonly known as factor analytic techniques or factor analysis was discussed in Chapter 4. Factor analysis is generally used to reduce variables that share common variance, such as items in a scale, into sets of clusters, known as components or factors (Bartholomew et al., 2011). 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, and factor analysis was carried out to provide evidence in support or against this hypothesis (i.e. to test its structural validity). The family of techniques known as factor analysis includes a number of exploratory and confirmatory techniques. Factors can also be described in terms of percentage of the total variance explained by each factor (eigenvalues) and by individual item loadings on one or more factors (coefficients of item loadings can vary between -1 and +1). Principal Factors Analysis, a type of exploratory factor analysis recommended when the aim is to evaluate theoretical predictions such as latent constructs (Tabachnick & Fidell, 2007; Wilson & McLean, 2011), was used to explore the factor structure of the SAAS.

Two tests were initially conducted to assess the suitability of the data for factor analysis, the Kaiser-Meyer-Olkin (KMO) measure of sample adequacy and Bartlett's test of sphericity (Pallant, 2013). A sample size above 100-150 is generally considered to be sufficient for factor analysis (DeVellis, 2012; Tabachnick & Fidell, 2007). However, the KMO statistic was calculated to ensure that the sample was sufficiently large. For this purpose, a KMO value of > 0.60 is generally considered suitable for factor analysis (Williams et al., 2012). Pallant (2013) indicates that factor analysis is also appropriate if the Bartlett's test of Sphericity is significant ($p < 0.05$).

In order to assess the factor structure of the SAAS, factors were extracted based on a combination of two rules, widely used in the research literature to determine the smallest number of factors providing a satisfactory account of the interrelationships among a set of items (Kline, 2005; Pallant, 2013). Kaiser's criterion, also known as the eigenvalue rule, is frequently used to determine how many factors should be retained (Brace et al., 2013). An eigenvalue is essentially a measure of the amount of variance captured by a single factor, and the rule indicates that factors with an eigenvalue less than 1 should not be retained in the factor solution (DeVellis, 2012). Another approach is the scree test (Catell, 1966), which involves the visual inspection of a scree plot (generated in SPSS) providing a graph of the eigenvalues of all the factors considered in the data set. Only factors above the point of inflexion (corresponding to the elbow of the plot) should be retained in a factor solution (Brace et al., 2013). Guidelines for exploratory factor analysis recommend that these two methods are considered in combination (Costello & Osbourne, 2005) and were thus both considered in this study. When more than one factor is identified, factors are rotated, using statistical procedures that do not modify the underlying solution, but increase the interpretability of clusters of item loadings (Pallant, 2013). In this study, a direct oblimin approach to rotation was planned if more than one factor was identified, as it is the recommended approach to factor rotation when possible factors are expected to be correlated (Williams et al., 2010). Item loadings on factors were also inspected, as they provide an indication of the strength of the association between an item and a given factor (Child, 2006). The conventions proposed by Tabachnick and Fidell (2007) were used to evaluate the strength of item loading coefficients:

- 0 – 0.44 = poor

- 0.45 – 0.54 = fair

- 0.55 – 0.62 = good

- 0.63 – 0.70 = very good

- \> 0.70 = excellent

8.4 Results

A total of 178 pregnant women returned by post completed questionnaires. Four respondents, however, did not complete the consent form ($n = 3$) or the reply slip ($n = 1$) and were thus not included in the analysis. All the results reported here refer to the remaining 174 women who completed questionnaires, consent forms and reply slips. The response rate was difficult to assess. As with the pilot study, it was not feasible for midwives to record all study packs that were handed out. A total of 750 study packs were given to midwives between February and July 2018. If all packs were handed to women the response rate would be 23.2%, slightly higher than for the pilot study. However, this may not be completely accurate. Recruitment was stopped after six months because of time constraints of the PhD, and once it was determined that the sample size was sufficient to conduct factor analysis. The initial recruitment target ($n = 200$) was only marginally not achieved (84%). It was not possible to assess whether there were differences between women who returned questionnaires and those who did not (i.e. response bias). However, the representativeness of the sample was examined based on information on respondents' age, parity, ethnicity and education, as discussed below.

With regard to missing data, 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. Four participants did not answer the two questions on ease of completion and acceptability of the SAAS. Furthermore, one participant did not complete any EPDS items. Thus, only 173 cases were included in the calculation of discriminant validity. All demographic and obstetric information was completed, with the exception of one woman who missed the '*Education*' question. The very small proportion of missing data, far below the commonly accepted rate of $\leq$ 5% (Furr, 2011), would appear to indicate that the questionnaire was adequately designed and formatted. Outliers were inspected by histograms, boxplots, and range checks. As mentioned earlier, while a number of outliers were observed, none were removed from

the analysis as no clear mistakes were identified and it was considered important to retain cases with extreme responses (Tabachnick & Fidell, 2007).

Subsample for the diagnostic interview

Thirty-seven women from the original sample were assessed with a structured diagnostic interview. The initial recruitment target of $n = 60$ was thus only partially reached (62%). A total of 71 study participants were invited to participate in the diagnostic interview. All women who were initially selected for the clinical interview and that eventually did not take part were contacted by phone ($n = 27$) or email ($n = 7$). However, a substantial proportion of those women did not answer the phone ($n = 21$), on four attempts on different days) or did not reply to two emails in two consecutive weeks ($n = 5$). A time and date for the telephone interview was arranged with eight women, but they did not answer the phone. Of the 37 women who were interviewed, eleven met the criteria for M.I.N.I diagnoses of anxiety disorder. Of those, four met the criteria for GAD, two for Posttraumatic stress disorder, two for OCD, one for Panic Disorder and two for multiple anxiety disorders. To ensure meaningful comparisons between the results of the scale and the structured diagnostic interview, the SAAS and GAD-7 were re-administered to women attending the diagnostic interviews, if a period of more than four weeks had passed since they first completed the scale ($n = 4$).

8.4.1 Demographic and obstetric characteristics of the sample

The mean age of the 174 women who participated in the study was 31.1 years (range 19 – 43). This was consistent with recent statistics indicating that 30.1 is the average age at which women in Scotland give birth (NHS Information Services Division [ISD], 2017). 54% of participants were in their first pregnancy ($n = 94$) while the remaining 46 % ($n = 80$) were in their second or subsequent pregnancy. Compared with maternity statistics both at the Scottish and UK level, with respectively 43% and 42% of first births in 2017 (Office for National Statistics [ONS], 2017; NHS Information Services Division [ISD], 2017), women at their first pregnancy were somewhat over-represented in this sample. Among women who had previously given birth, almost one third reported at least one experience of pregnancy

or birth complication (*n* = 24), consistently with the literature presented in Chapter 2. The mean gestational week of respondents at the time of completion of the questionnaire was 28.4 weeks, with a fairly similar representation of women in the second and third trimester of pregnancy, as illustrated in Table 18.

Table 18 – 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
Weeks pregnant	28.4 (6.93)	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

Socio-demographic data collected by study participants also included information in relation to ethnicity, education and marital status, as reported in Table 19.

Table 19 – 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 *(% among all people 16 years or over)*
Level 1 - 'O' Grade, Standard grade or equivalent (SVQ level 1 o 2)	13	7.5%	23%
Level 2 - Higher, A level or equivalent (SVQ Level 3)	16	9.2%	14.3%
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 (PhD)	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 *(% among 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 *(all engaged)*	1.7%	*Included in 'Single'*

With regard to ethnic origin, the sample in this study included a slightly lower proportion of *'White Scottish'* and *'White – Other British'* and a slightly higher proportion of *'White – Any other white ethnic group'* and *'Asian/Asian British'* compared to the 2011 census. This is arguably likely to result from migration to Scotland in recent years, and the fact that the sample was recruited in an urban area. The sample appeared, however, to represent adequately well the Scottish female population. Figures limited to women of childbearing age, which would have allowed more appropriate comparisons, were not available.

Notably, the sample was considerably more educated than the average of people living in Scotland, with a total of 67.1% of women reporting *'Level 4 or above'* as their highest level of qualification. While this is not uncommon among research participants (Bowling, 2014), it is a potential limitation of the study that needs to be considered. In relation to marital status, the questionnaire included the response category *'Cohabiting'* and *'Other'*. These were, however, not included in the 2011 Scotland census. If *'Single'*, *'Cohabiting'* and *'Other'* are considered as a single category as in the census, their proportion in the sample is similar to national-level statistics. As it could arguably be expected, married women were over-represented in this sample.

Women were also asked a single question to assess their general level of pregnancy-related anxiety symptoms. Specifically, women were asked: *"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"*. The mean score was 4.6 (SD 1.91). Scores to this single, pregnancy-related anxiety question were also 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. Specifically, scores of 7 or above were considered indicative of probable pregnancy-related anxiety. This score was chosen based on the observation that 18.4% of women scored at or above this cut-off for this question, consistently with prevalence estimates for pregnancy-related anxiety reported in previous studies (Heimstad et al., 2006; Lukasse et al., 2014).

8.4.2. Ease of completion and acceptability of the SAAS

The ease of completion and acceptability of the SAAS were evaluated with two questions, as described in the Data analysis section. Four women did not complete the two questions. Out of the remaining 170 study participants, 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 entirely acceptable for use in routine antenatal care by the vast majority of women who participated in the study. The proportion of women assigning a score < 7 was 10.6% ($n = 18$) for ease of completion and 4.7% ($n = 8$) for acceptability. The very high mean score for acceptability is of note, as developing a scale that was acceptable to the target population was considered critical at all stages of the scale development process.

8.4.3 Screening accuracy of the SAAS, GAD-2 and GAD-7 versus the M.I.N.I

As discussed earlier, it was clearly important to evaluate the screening accuracy of the SAAS in identifying women experiencing an anxiety disorder, as this was part of the primary aim of this programme of work. It was also of considerable importance to assess the screening accuracy of the GAD-7 and the GAD-2 (considered also as a separate scale), both because they are the NICE-recommended measures to screen for perinatal anxiety in antenatal care (NICE, 2014), and in consideration of the scarce evidence available to support this recommendation in perinatal populations (Nath et al., 2018). As also previously noted, NICE recommends that the two questions of the GAD-2 should be initially asked, and that women scoring a total of 3 or above should be either administered the GAD-7 (but a cut-off score is not recommended) or referred for further assessment. The sensitivity, specificity, positive and negative predictive values of the three scales (based on M.I.N.I diagnosis) at different cut-off points were calculated, and are presented in Table 20, 21 and 22 in the next pages. Values in bold indicate candidates as optimal cut-off scores for the three measures.

Table 20 – Sensitivity, Specificity, Positive and Negative Predictive Values for the SAAS at a range of cut-off points

SAAS cut-off score	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%)
≥13	54% (39% - 71%)	100% (Not computable)	100% (Not computable)	84% (72% - 96%)

95% confidence intervals in parentheses

At a cut-off score of 8 or above, the SAAS 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 of the true positive rate (i.e. sensitivity, 73%). Both cut-off scores thus resulted in good to excellent values of sensitivity and specificity. While sensitivity and specificity should both be at least above an acceptable level (> 70 %), a number of authors (Zimmerman & Mattia, 2001;: Streiner & Norman, 2008; Furr, 2011). have pointed out that there are many instances in clinical settings in which it might be preferable to prioritise one of the two indexes over the other (i.e. either maximise the

proportion of true positive cases identified or the proportion of true negative cases). This important point is further elaborated in the final Discussion chapter (Chapter 9).

Table 21 – Sensitivity, Specificity, Positive and Negative Predictive Values for the GAD-7 at a range of cut-offs points

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%)
≥11	36% (21% - 52%)	96% (89% - 100%)	80% (67% - 93%)	78% (65% - 91%)

Table 22 – Sensitivity, Specificity, Positive and Negative Predictive Values for the GAD-2 at a range of cut-offs points

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%)

The GAD-7 showed very good sensitivity (82%) and excellent specificity (96%) at a cut-off of 7 or above. As noted earlier, however, NICE does not specify a cut-off score for the GAD-7 in its guidance on mental health in the perinatal period (NICE, 2014), which may lead to assume 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. A similar, but arguably more significant problem was found with regard to the optimal cut-off score for the GAD-2 of 2 or above. This cut-off yielded 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).

Analysis of the AUROCs

ROC curves were generated for the three subscales. As previously discussed, the Area Under the ROC Curve (AUROC) provides a single summary measure of the ability of a scale or a test to discriminate between subjects with and without a specific target condition, and is thus particularly useful in comparing the screening accuracy of different scales (Bland, 2000;

Pallant, 2013). The three ROC curves for the SAAS, GAD-7 and GAD-2 are presented in Figures 6, 7, and 8.

Figure 6 – Receiver Operating Characteristic curve for the SAAS
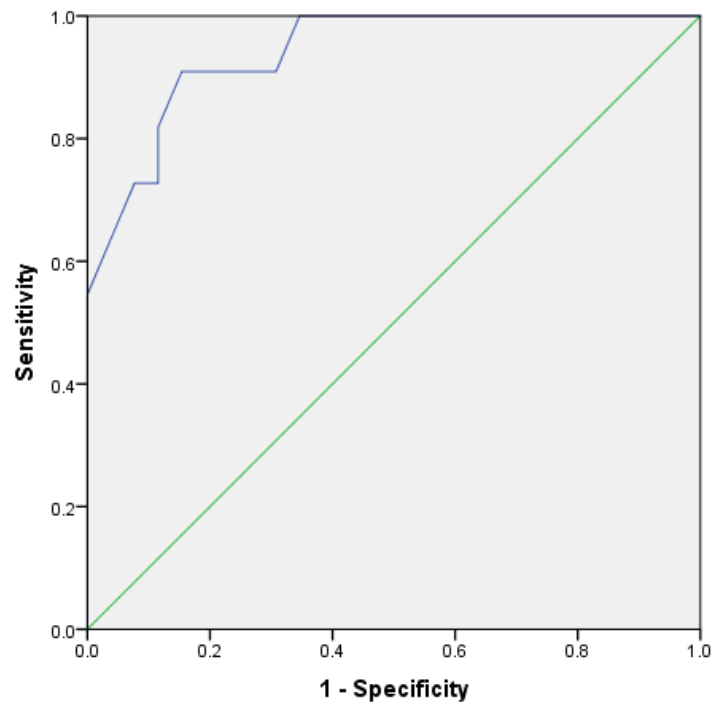
Figure 7 – Receiver Operating Characteristic curve for the GAD-7

Figure 8 – Receiver Operating Characteristic curve for the GAD-2

The test for the ROC curves conducted in SPSS were all statistically significant ($p < 0.05$). The AUROC for the SAAS was 0.94 (95% CI: 0.86-1), well above the threshold of 0.90 indicating excellent discriminative ability (Bland, 2000). The AUROC thus confirmed the particularly good screening accuracy of the SAAS, as indicated by values of sensitivity and specificity. The AUROC for the GAD-7 was 0.93 (95% CI: 0.85-1), only marginally lower tha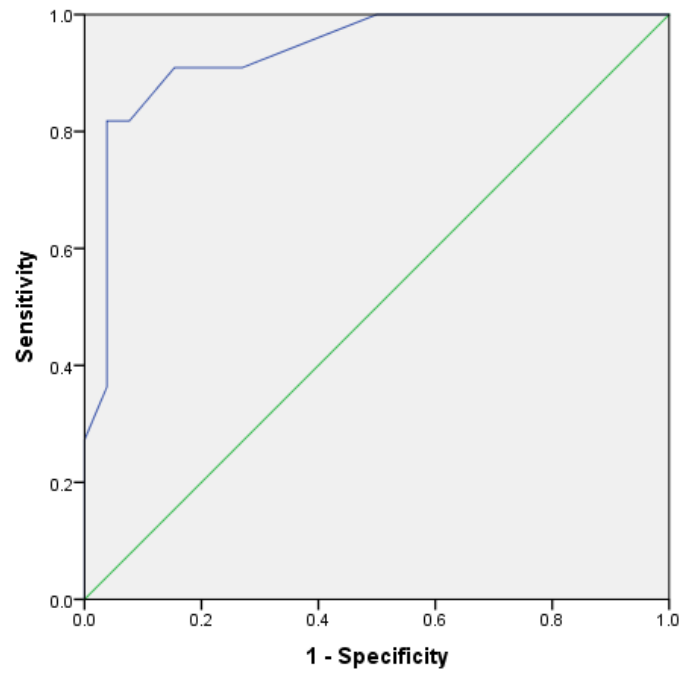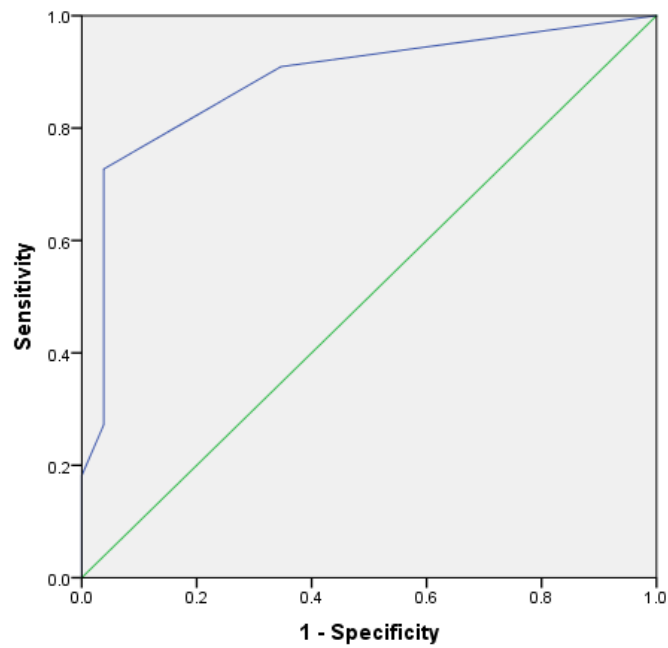n the SAAS. The two scales both showed in fact very good to excellent sensitivity and specificity to the optimal cut-off points. Finally, the AUROC for the GAD-2 was 0.88 (95% CI: 0.75-1), a value appearing to confirm the poorer discriminative accuracy of this ultra-brief scale in an antenatal sample compared to the GAD-7 and the SAAS. The relatively large confidence intervals of the values of sensitivity, specificity, PPV, NPV and AUROCs are arguably the result of the relatively small sample ($n = 37$, with 11 M.I.N.I diagnoses of anxiety disorders) used for the calculation of the screening accuracy of the three scales, a potentially important limitation of this study which is further considered in the Discussion chapter.

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

The SAAS, consistently with the construct definition of antenatal anxiety proposed in Chapter 6, was designed to screen for a range of problematic anxiety symptoms during pregnancy, including symptoms of anxiety disorders and symptoms of pregnancy-related anxiety. However, the 'gold standard' used in this study (i.e. M.I.N.I) only provided evidence of the screening accuracy of the SAAS in identifying women with an anxiety disorder. Consequently, it was deemed important to assess its screening performance, and compare it to the GAD-2/7, when both anxiety disorders and pregnancy-related anxiety symptoms were considered. As noted above, for this purpose a single question was included in the questionnaire to assess the general level of pregnancy-related anxiety symptoms in study participants: "*From 1 to 10, how do you feel about your pregnancy and about giving birth?*" with possible scores ranging from "*1= completely calm*" to "*10= extremely anxious*".

In secondary statistical analyses, the sensitivity, specificity, PPV and NPV for the SAAS and the GAD-2/7 were thus re-calculated, including among the 'positive' cases both women with a M.I.N.I-diagnosed anxiety disorder and women scoring 7 or higher to the single

pregnancy-related anxiety question, assumed to be an indication of significant PrA symptoms. For reasons of brevity, only the most relevant cut-off scores for these parameters of screening accuracy are reported for the three scales (Tables 23, 24, 25).

Table 23 – Sensitivity, Specificity, PPV and NPV for the SAAS in identifying women with an anxiety disorder and/or scoring ≥ 7 to the PrA question

SAAS cut-off score	Sensitivity	Specificity	PPV	NPV
≥ 7	**76%** (62% - 90%)	**90%** (80% - 100%)	**87%** (76% - 198%)	**82%** (70% - 95%)
≥ 8	71% (56% - 86%)	90% (80% - 100%)	86% (75% - 97%)	78% (65% - 91%)
≥9	59% (43% - 75%)	90% (80% - 100%)	83% (71% - 95%)	72% (58% - 87%)

Table 24 – Sensitivity, Specificity, PPV and NPV for the GAD-7 in identifying women with an anxiety disorder and/or scoring ≥ 7 to the PrA question

GAD-7 cut-off score	Sensitivity	Specificity	PPV	NPV
≥ 7	**47%** (31% - 63%)	**95%** (88% - 100%)	**89%** (79% - 99%)	**68%** (53% - 83%)
≥ 8	**47%** (31% - 63%)	**95%** (88% - 100%)	**89%** (79% - 99%)	**68%** (53% - 83%)
≥9	41% (25% - 57%)	95% (88% - 100%)	87% (76% - 98%)	66% (51% - 81%)

Table 25 – Sensitivity, Specificity, PPV and NPV for the GAD-2 in identifying women with an anxiety disorder and/or scoring ≥ 7 to the PrA question

GAD-2 cut-off score	Sensitivity	Specificity	PPV	NPV
≥ 2	**47%** (31% - 63%)	**95%** (88% -100%)	**89%** (79% - 99%)	**68%** (53% - 83%)
≥ 3	18% (6% - 30%)	95% (88% -100%)	75% (61% -89%)	76% (62% -90%)
≥4	12% (2% -23%)	100% (Not computable)	100% (Not computable)	57% (41% -73%)

As evident from these tables, the SAAS showed a superior screening accuracy when both anxiety disorders and PrA symptoms were considered (76% sensitivity; 90% specificity). Once more, the screening performance of the GAD-2 at the NICE recommended cut-off score of 3 or above was particularly 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. Similarly the GAD-7, at its optimal cut-off scores of ≥ 7, exhibited only moderate sensitivity and excellent specificity. The superior accuracy of the SAAS, compared to the GAD-2/7, in screening both for anxiety disorders and for pregnancy-related anxiety symptoms may not be particularly surprising when considering the constructs measured by these scales. Nonetheless, it provides further preliminary evidence to support the use of the SAAS as a brief screening scale for a range of problematic anxiety symptoms during pregnancy. At the same time, it suggests that the GAD-2 and GAD-7 might not be sufficiently accurate to screen for the whole spectrum of problematic anxiety symptoms that women can experience in the antenatal period. These findings, further commented in the Discussion chapter, also indicate that the SAAS in this sample showed very good criterion validity.

8.4.4 Internal consistency

Cronbach's Alpha coefficient as a measure of internal consistency was examined for the SAAS and the GAD-7. The internal consistency of the GAD-2 was not calculated as it has been noted that calculating the internal consistency of a two-item measure is inappropriate (O'Brien et al., 2008). For the SAAS, Cronbach's Alpha was found to be α = 0.88. This can be considered an excellent value, which closely approximates the value for clinical applications of a scale (α ~ 0.90) often suggested in the literature (Kline, 2005; DeVellis, 2012). Internal consistency for the GAD-7 was α = 0.87, comparable to the SAAS. This considerably high value of internal consistency for the GAD-7 was expected when considering that the scale was constructed to asses a single anxiety disorder (GAD) and is consistent with previous studies in the general and pregnant populations (Spitzer et al., 2006; Zhong et al., 2015). At the same time, such a high Cronbach's Alpha value for the SAAS seemed to support the unidimensionality of the construct measured by the scale, as hypothesised in the definition of the construct of antenatal anxiety. This hypothesis could be further tested through factor analysis, as presented in a subsequent section.

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 & Norman, 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. As previously discussed, item parameters within these ranges contribute to provide evidence of the psychometric robustness of the scale (DeVellis, 2012). Inspections of response distributions did not reveal any floor or ceiling effect among the 10 item composing the SAAS.

8.4.5 Convergent and divergent validity

Correlation coefficients were calculated between total scores of the SAAS, GAD-7 and EPDS to evaluate convergent and discriminant validity. Prior to this, the assumption of normal distribution of scores was assessed by calculating the Kolmogorov-Smirnov statistic, which was required to determine whether a parametric or non-parametric test was appropriate for the calculation of correlation coefficients. A non-significant result ($p > 0.05$) indicates normality (Pallant, 2013). The Kolmogorov-Smirnov test showed a value of $p <$

0.01, indicating a violation of the assumption of normality. Consequently, as detailed in the Data analysis section, Spearman's correlation (rho or r_s) was used to calculate and report correlation coefficients. A significant, positive correlation was found between the SAAS and GAD-7 ($r_s = 0.70$, $n = 174$; $p < 0.01$). 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 ($r_s = 0.68$, $n = 174$; $p < 0.01$). Subsequently, the magnitude of the correlation between SAAS and EPDS scores was examined. The findings related to this correlation were somewhat unexpected. There was a significant, positive correlation between the two measures ($r_s = 0.73$, $n = 173$; $p < 0.01$), which was in strength slightly larger 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 correlation between the GAD-7 and the EPDS was also calculated at this stage to examine whether the large correlation between the SAAS and EPDS was indicative of potential issues with the construct validity 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 ($r_s = 0.73$). These findings would appear to confirm that the EPDS is not a unidimensional measure of depression, as indicated in various studies reported earlier. Furthermore, while the different correlations between the three scales (r_s range: 0.70-0.73) suggested a considerable degree of overlap between the constructs measured, correlation coefficients were not sufficiently large to indicate that the scales measured highly similar constructs (a Spearman's correlation coefficient of 0.73 indicates that 53% of the variance is shared between two scales). Hypotheses which may contribute to explain these findings are presented in the final chapter of the thesis.

8.4.6 Factor structure

Principal Factors Analysis was carried out to examine the factor structure of the SAAS. The KMO measure of sample adequacy was excellent (0.902), far exceeding the limit for acceptable sample size of > 0.60 (Kline, 2005). Bartlett's test of sphericity also reached statistical significance (X^2 (45) = 706.90, $p < 0.01$), thus supporting further the suitability of the data for factor analysis. Principal Factors Analysis was thus conducted, and initially the magnitude of the eigenvalues were inspected. These are reported in Table 26.

Table 26 – Eigenvalues from Principal Factors Analysis on SAAS scores

| Factors | Eigenvalues | | |
	Total	% 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

As discussed earlier, the eigenvalue rule or Kaiser's criterion recommends that only factors with eigenvalues above 1 are retained. This provided an initial indication of a unidimensional structure of the SAAS, with a single eigenvalue above one explaining approximately half of the total variance. A second eigenvalue, however, 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 inflexion, as they are also used to inform the decision on which factors should be retained in the factor structure of a scale (DeVellis, 2012; Brace et al., 2013). The scree plot for the SAAS is presented in Figure 9.

Figure 9 – Scree plot generated with Principal Factors Analysis on SAAS scores
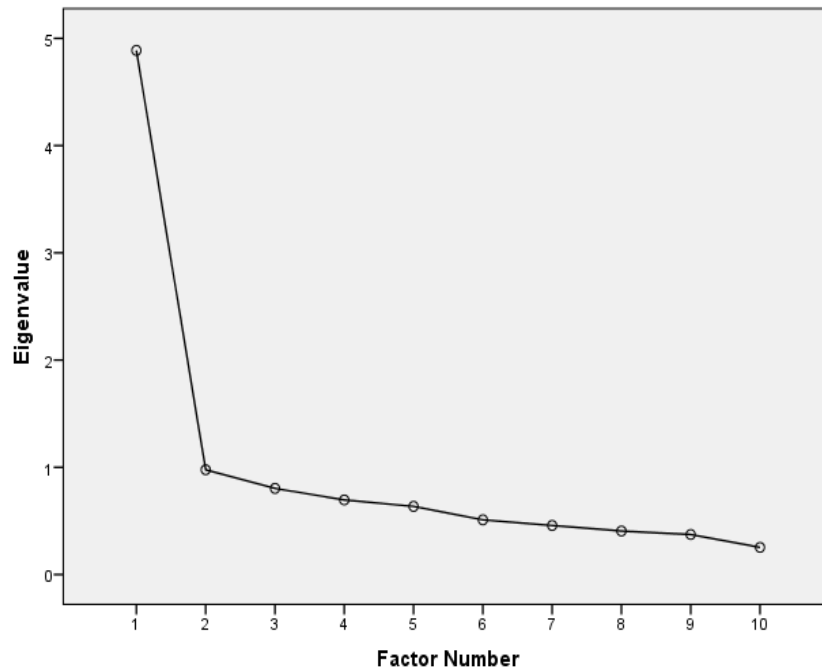
Inspection of the scree plot clearly confirmed that a one-factor solution for the SAAS was the most appropriate, as only one factor was found well above the point of inflexion. Consequently, based on the recommended combination of the eigenvalue rule and examination of the scree plot, a single factor was retained and a one-factor solution for the SAAS was proposed.

Individual item loadings on this single factor were also inspected. 7 of the 10 items included in the SAAS showed very good or excellent item loadings (≥ 0.63) according to the criteria proposed by Tabachnick and Fidell (2007) as detailed earlier. The remaining three items exhibited respectively good (item 4: 0.58) and fair item loadings (item 8: 0.51; item 9: 0.48). These items were the three pregnancy-related anxiety items included in the scale. While their not particularly high item loading coefficients might suggest a sub-component related to PrA within the 10-item SAAS, the magnitude of the loadings would not support the hypothesis of a distinct factor, as also confirmed by the findings of the factor analysis.

8.5 Final considerations

In conclusion, the findings presented in this chapter are an early indication that the SAAS has very good psychometric properties, and is potentially appropriate as a screening tool for use in clinical settings. Its measurement properties will clearly need to be further tested in future research studies, as discussed in the final chapter. In this sample, its internal consistency close to excellent ($\alpha = 0.88$) confirmed the reliability of the scale and suggested a unidimensional construct, 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 unexpected finding related to the correlation between the SAAS and the EPDS do not allow to reach conclusions regarding the discriminant validity of the SAAS. Possible explanations for these findings are presented in the final Discussion chapter.

As documented in the Results, the screening performance of the SAAS in this study was close to excellent, both in the identification of women experiencing an anxiety disorder (sensitivity: 91%, specificity: 85%) and significant pregnancy-related anxiety symptoms (sensitivity: 76%; specificity: 90%), providing evidence in support of the criterion validity of the scale. 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. This is an important distinction, which is also elaborated further in the final chapter. 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 pregnancy-related anxiety symptoms 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. The particularly poor screening performance of the GAD-2 at the NICE-recommended cut-off score of ≥ 3 was somewhat alarming. Finally, a potentially important limitation of this study was that the target sample size was not entirely achieved. A note of caution is thus appropriate when commenting and interpreting these findings. This is also

discussed further in the final chapter among the limitations of the research presented in this thesis.

Chapter 9   Discussion

9.1   Revisiting the study aim and research questions

In this final chapter, the evidence and findings presented in the thesis are discussed in relation to their rigour, limitations and future potential. In light of the findings presented in the five experimental chapters, it is initially useful to re-examine the study aim and research questions as presented in the introductory chapter and in 3.3.2. This programme of work primarily aimed to develop a brief scale with robust psychometric properties. Internal consistency, as well as several forms of validity were quantitatively tested. Other types of validity (i.e. face, content, and partially construct validity) are not easily summarised using statistical parameters, but require careful consideration at all stages of the scale development process (Simms, 2008). The methods and procedures used to develop the SAAS provide evidence of methodological rigour in the attempts to maximise the psychometric properties of the final version of the scale. The following paragraphs summarise the evidence that was provided to support the claim that the SAAS has overall desirable psychometric properties.

The primary aim of the research was the development and preliminary psychometric validation of a brief self-report scale specifically devised to screen for problematic anxiety symptoms in pregnant women. As noted in Chapter 1 and in other parts of the thesis, the scale was also developed with the objective of making it potentially feasible to implement in routine antenatal care. The SAAS was thus constructed with the target of producing a final version consisting of less than 12 items, a prerequisite for the scale to be considered for use as a screening tool in maternity care in the UK (NICE, 2014). For the same reason, the acceptability of the scale to the target population was also considered to be of primary importance. The scale construction process, and the subsequent phase of preliminary psychometric validation, were informed by five distinct studies, specifically aimed to answer one or more research questions, as discussed below:

- *Research question 1*: What should a construct definition of antenatal anxiety include in order to cover the core domains of problematic anxiety symptoms in pregnancy?
- *Research question 2:* Which items are the most appropriate to operationalise the proposed construct of antenatal anxiety into a self-report rating scale?

Two studies were conducted to answer Research Questions 1 and 2. A systematic review of the psychometric properties and item content of anxiety scales used in studies with pregnant women was initially carried out (Chapter 4). The aim was to identify anxiety symptoms and domains showing good or excellent evidence of their psychometric value when used to assess general or pregnancy-related anxiety in antenatal populations. The findings from the systematic review were complemented by qualitative interviews with women with experience of problematic anxiety symptoms during pregnancy (Chapter 5). The affective, cognitive, behavioural and somatic content areas of the construct of antenatal anxiety identified through the two studies were subsequently combined, using predefined criteria to evaluate the relative strength of evidence for each content area to be considered an important domain of the target 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 face and content validity in the initial item pool (Netemeyer et al., 2003; Simms, 2008). Furthermore, it can be argued that these two studies also supported the construct validity of the final version of the SAAS. As noted in the Method chapter (3.1.3), while it is common practice to assess this form of validity by examining convergent and discriminant validity with other scales through psychometric testing (DeVellis, 2012), a number of authors have proposed that construct validity encompasses all forms of validity, and should be evaluated by examining and considering the procedures used throughout the scale development process (Simms, 2008; Streiner & Norman, 2008; Abell et al., 2009). Based on this approach, evidence in support of the construct validity of a scale begins with an accurate, evidence-based definition of the construct of interest, and is further demonstrated through the robustness of the research methods used at all stages of scale development (Newton, 2012).

On the basis of the findings from the systematic review and the qualitative interviews, definitions of the construct of antenatal anxiety were formulated (6.2), and an initial item pool was generated to reflect the proposed construct. The conceptual definition of antenatal anxiety delineated the general boundaries of the construct (i.e. anxiety symptoms in pregnancy perceived as distressing and/or having a negative impact on individual functioning, and experienced for a sufficiently prolonged period of time), while the operational definition detailed specific content areas and facets of the construct (e.g. feelings

of panic or intense fear; excessive worry about the baby's health). A clear and well-delineated definition of the construct, providing the basis for the generation and selection of item relevant to its measurement, is also typically used as evidence of content validity of a scale (Clark & Watson, 1995; Streiner & Norman, 2008). Notably, in contrast with several authors (Huizink et al., 2004; Blair et al., 2011) who have proposed that pregnancy-related anxiety should be considered a specific and entirely distinct syndrome (see 2.3.2), the definition of antenatal anxiety proposed in this thesis considered 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.

- *Research question 3:* Which items are considered clear, relevant and acceptable by the target population and experts, and can thus be used to create a short and psychometrically robust self-report scale for the assessment of antenatal anxiety?

The initial item pool of 55 items generated in the previous phase was initially reviewed by three MMHS Change Agents, who provided feedback on the clarity and acceptability of all items (6.3.3). Their comments were used to modify a small number of items not deemed to be sufficiently clear or acceptable. This procedure further contributed to increase the chances of developing a scale acceptable to the target population, composed of items clearly worded and with good face validity. The mean scores in the psychometric validation study with regard to ease of completion and acceptability of the 10-item SAAS indicate that these objectives were achieved, as the scale was considered highly acceptable and very easy to complete. At this stage, three additional items were also suggested for inclusion in the initial item pool by two clinicians with specific expertise in the area, in an attempt to maximise further the content validity of the scale.

The Delphi study presented in Chapter 6 made use of the knowledge and clinical expertise of a group of health professionals working in the area of perinatal mental health to considerably reduce the number of items in the initial item pool, from 59 to 30. While this initial selection of items based on expert opinion served the purpose of producing a shorter, preliminary version of the scale for pilot psychometric testing, it also ensured that the face and content validity of the resulting scale (as defined in 3.1.3) were considered and enhanced (Netemeyer et al., 2003). A sufficiently large number of items were retained at this stage,

thus ensuring item coverage of different facets of the construct (i.e. content validity). It is also reasonable to assume that items that were considered less important by experts lacked sufficient face validity, or were not considered sufficiently relevant to the assessment of antenatal anxiety (i.e. lack of content validity). As noted above in relation to construct validity, this procedure of item reduction through expert opinion can also be considered to contribute to further support claims of this form of validity.

The pilot study presented in Chapter 7 addressed the second part of Research Question 3, related to items to be used to produce a brief and psychometrically robust final version of the scale. As noted in 3.2, this stage of pilot psychometric testing is concerned with maximising the internal consistency reliability of the scale, while retaining a sufficient number of items to reflect all the core domains of the construct of interest in order to maximise content validity. The study was thus predominantly based on quantitative analyses and items were selected for the final version of the scale by examining a number of item statistics, as detailed in Chapter 7. The particularly high internal consistency ($\alpha = 0.96$) of the 30-item version of the SAAS allowed the selection of 10 items which were eventually found to retain an excellent level of internal consistency. A brief and internally consistent measure for the assessment of the target construct was thus produced. As previously noted, the selection of items composing the final version of the SAAS was also partially based on considerations related to item content and coverage of a range of facets of the target construct, in an attempt to preserve as far as possible content and construct validity in the 10-item SAAS. The psychometric soundness of this final version of the scale was tested by addressing Research Questions 4 and 5:

- Research question 4: What is the evidence in relation to the convergent and discriminant validity, internal consistency and factor structure of the final version of the scale?

- Research question 5: How does the new scale perform when compared to the measure currently recommended by NICE (GAD-2/7), and to expert assessment using a structured diagnostic interview; and what are the optimised cut-off points for maximising sensitivity and specificity of the scales?

The 10-item SAAS, along with the GAD-7 and the EPDS, were completed in a second cross-sectional survey by a sample of pregnant women in their second and third trimester of pregnancy. As noted above, the internal consistency of the SAAS was found to be close to excellent (α = 0.88), approaching the value that has been recommended for clinical applications (Kline, 2005). Despite the inclusion of three pregnancy-related anxiety items, the SAAS was also found to have a single-factor structure, consistently with the definition of the construct proposed in this thesis (6.2). Although factor analysis often requires replication in a number of studies, this preliminary indication would support the hypothesis that pregnancy-related anxiety symptoms can be considered a dimension of the more general construct of antenatal anxiety.

Construct validity of the SAAS in this psychometric validation of the scale was evaluated through its convergent and discriminant validity, by hypothesising and testing the correlations of the scale with the GAD-7 and the EPDS. As noted in Chapter 8, while evidence of the convergent validity of the SAAS was supported by a large correlation with the GAD-7, the somewhat surprising finding related to the large correlation coefficients between the EPDS and the two anxiety measures (r_s range: 0.70 - 0.73) which appeared to question the discriminant validity of the scales. A possible explanation for these large correlations relates to previous findings documenting the existence of a 3-item anxiety subscale within the EPDS, as documented in Chapter 4. An alternative hypothesis, not incompatible with the previous one, is that the well-documented comorbidity between anxious and depressive symptoms, both in the general and in perinatal populations (Staneva et al., 2015a), might have resulted in a partial overlap of symptoms and thus of constructs as measured by the three scales. This would also be consistent with the tripartite model of depression and anxiety by Watson & Clark (1991), discussed earlier in the thesis (2.2). The similarly large correlation found between the GAD-7 and the EPDS appears to confirm problems 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. Additionally, as discussed in 3.3.1 and briefly noted above, others have argued that a global assessment of construct validity should be conducted by examining evidence provided throughout the scale development process (Simms, 2008; Abell et al., 2009).

In relation to the screening performance of the SAAS in the psychometric validation against the M.I.N.I., the scale showed good to excellent sensitivity and specificity at two distinct cut-off scores, thus providing evidence of very good criterion validity. A cut-off point of $\geq$

8 maximised the sensitivity of the scale (91%), and thus the proportion of true positives. Using a score of 12 or above as cut-off allowed to maximise the specificity of the SAAS (92%), but reduced substantially its sensitivity (73%). Here the implications of choosing one cut-off over the other are briefly discussed. It is, however, important to note that the potentially important limitation of the relatively small number of women assessed using a diagnostic interview (i.e. the 'criterion' or gold standard) in this study should be considered when interpreting these findings. This limitation is discussed further in 9.2.

In the context of screening for problematic anxiety symptoms in routine antenatal care, a case can arguably be made for both cut-offs. 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 unmotivated 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 & 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 simply 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. This approach, however, is likely to increase the proportion of false positives at the initial stage of screening. The decision to prioritise sensitivity over specificity or vice-versa in the context of screening for common perinatal mental health problems is complex, and a number of important factors need to be taken into account (Milgrom & Gemmill, 2014). In this section, some brief comments were provided to highlight the clinical implications of favouring one parameter over the other. To conclude the considerations on the SAAS, notably the scale also performed well in identifying women with an anxiety disorder AND/OR pregnancy-related anxiety symptoms, thus providing further evidence in support of its criterion validity.

Finally, the screening performance of the GAD-2 in this study also deserves a further comment. As previously observed in the thesis (2.6), the NICE recommendation to use the

GAD-2 (NICE, 2014) was exclusively based on evidence of its good screening accuracy for anxiety disorders in the general population, as no studies in perinatal populations were available at the time of the publication of the guidance. The findings presented in this thesis question the decision to rely on psychometric evidence from populations other than perinatal women. They also imply that, if the NICE-recommended cut-off score of ≥ 3 is currently used in clinical practice, a substantial number of pregnant women experiencing clinically significant anxiety symptoms are likely to go undetected (sensitivity 27%). Although the same cut-off score yielded an excellent specificity (96%), the particularly poor sensitivity of the GAD-2 at this cut-off cannot be overlooked. The psychometric validation study also indicated that the GAD-2 and the GAD-7 are not sufficiently accurate to screen for the whole spectrum of problematic anxiety symptoms (i.e. including PrA symptoms) that women can experience in the antenatal period.

9.2 Strengths and limitations of the research

The five studies conducted as part of this programme of work had several strengths and limitations. Initially, some general strengths of the research, as well as of specific studies presented in the thesis, are discussed. A number of other strengths of the research have arguably been discussed as part of the previous section. The remaining part of the section is focused on several limitations of the research, which are particularly important to consider when interpreting and commenting study findings.

The research presented in this thesis has a number of strengths. Research methods used to address sequentially the different research questions were carefully chosen following best practice in scale development, as detailed in Chapter 3. The methods and procedures that were eventually used to conduct the five studies were thus based on a solid theoretical and evidence-based background. A significant strength of the research was the combination of different sources of evidence to inform construct definition and the subsequent generation and selection of items to measure the target construct with a brief, psychometrically robust scale. Sources of evidence included the psychometric literature on anxiety scales used in pregnancy, women with experience of the target condition, health professionals with expertise in the area of perinatal mental health, and the intended population of respondents through pilot psychometric testing of candidate items.

Both for the systematic review presented in Chapter 4 and for the qualitative interviews, predefined criteria were used to determine the strength of evidence and relevance of specific symptoms to the assessment of antenatal anxiety, thus enhancing the credibility of the findings. In relation to the qualitative interviews with women with experience of problematic anxiety symptoms (Chapter 5), Evans and colleagues (2015) in their review of anxiety measures used with pregnant populations found no studies that used interviews with the target population to inform tool development. Although several measures have been developed to assess anxiety symptomatology during pregnancy, to my knowledge this is the first study to incorporate the experience and symptoms of pregnant women with problematic anxiety in the tool development phase. A number of authors have indicated that interviews with the target population can considerably contribute to strengthen the content validity of a measure (Abell et al., 2009; Streiner & Norman, 2008). Similarly, the involvement of MMHS Change Agents in relation to the initial review of the item pool for clarity and acceptability (Chapter 6), as well as the input provided with regard to the design and outline of the scale, have arguably contributed to the potential applicability of the scale to intended respondents, both in research and clinical settings.

In relation to the Delphi study, one clear advantage of involving a panel of experts in the item reduction phase of scale development is that, when carefully chosen, they provide a wealth of accumulated knowledge in the area of interest, which can be used to select the most relevant and appropriate indicators of the target construct (Clark & Watson, 2003). The use of a structured diagnostic interview as part of the psychometric validation of the SAAS can also be considered a significant strength of this research. Recent systematic reviews of psychometric studies in the area of perinatal anxiety (Meades et al., 2011; Evans et al., 2015) and my own review (Sinesi, Maxwell, O'Carroll & Cheyne, 2019) concluded that validation against a reference standard is still rarely included in the psychometric testing of scales used with perinatal populations.

In relation to the limitations of this programme of work, these can also be discussed by considering the different studies that have informed tool development. In relation to the interviews with women with experience of problematic anxiety symptoms, while they added an important qualitative component to the process of scale development (Hunsley & Mash, 2008; Streiner & Norman, 2008), the most significant limitation was perhaps that a proportion of women who took part in the interview did not have a diagnosis of an anxiety disorder. They were rather recruited based on a subjective self-assessment of problematic

anxiety during pregnancy, according to the criterion of "self-identified antenatal anxiety" detailed in 5.2.2. This might be considered a limitation of this study, although it allowed the inclusion of women who had experienced elevated levels of anxiety symptoms during pregnancy but were not formally diagnosed. Moreover, according to the definition of antenatal anxiety proposed in Chapter 6, anxiety disorders do not represent the whole spectrum of problematic anxiety symptoms in pregnancy and thus including women who may have not met formal diagnostic criteria was considered important.

The Delphi study detailed in Chapter 6 made use of expert opinion to aid the process of item selection and reduction. While the strengths of this approach in relation to item reduction were noted above, it can be argued that the purposive sampling technique adopted in this study, which relied on health professionals included in the mailing list of Maternal Mental Health Scotland, may have led to the over-representation of some professional backgrounds (e.g. clinical psychology; mental health nursing) and the under-representation of others (e.g. psychiatry), as illustrated in 6.5. Another potential drawback of this study is that four items were suggested for inclusion in the item pool by experts between round one and round two of the eDelphi. These four items were consequently only rated in a single round, and thus not subject to the iterative process of consensus building typical of Delphi studies. The potential limitation related to the small number of items rated only in one round was, however, arguably offset by the advantage of retaining a higher number of panellists.

One of the limitations of both the pilot and psychometric validation study was that the response rate was not accurately determined, for the reasons discussed in the previous chapter. Although the response rates for the two studies were relatively low (20.6% and 23.2% respectively), the actual response rate may have thus been higher. Despite the relatively low response rate it has to be noted, however, that in relation to the representativeness of the sample the only socio-demographic characteristic that considerably diverged from the most recent Scottish Census was the level of education. Women who took part in the psychometric validation of the SAAS reported significantly higher levels of educational attainment that the general Scottish population. Other obstetric and socio-demographics characteristics were all fairly representative of national-level statistics.

Perhaps the most significant limitation of the psychometric validation of the SAAS and GAD-2/7 against the M.I.N.I was that the target sample size for the diagnostic interviews ($n = 60$) was not achieved ($n = 37$). This limitation is potentially important, as it arguably had

a direct impact on the confidence intervals for the various parameters of screening accuracy. The relatively large confidence intervals reported for sensitivity, specificity, PPV, NPV and the Area Under the Curve (AUROC) for the three scales (8.4.3) would seem to indicate that some caution is needed when interpreting the parameters of screening accuracy reported in the study. The target sample size of 60 women to be assessed with a diagnostic interview would have provided width of the confidence interval of 10% for the estimates of sensitivity and specificity, according to the formula for the calculation of sample size requirements (Buderer, 1996) in studies of screening accuracy detailed in 8.3.1. As noted above this target was, however, not reached. Furthermore, as only 11 out of the 37 women assessed via a structured diagnostic interview were found to have an anxiety disorder, the values of the parameters of screening accuracy presented in Chapter 8 were based on a relatively small number of cases. As an example, the sensitivity of 91% showed by the SAAS at a cut-off score of 8 or above indicated that 10 out of 11 women with a M.I.N.I diagnosis of anxiety disorder had a score ≥ 8. It is evident, however, that a difference in the total score of a single participant could have caused a significant change in the estimate of sensitivity. A similar observation can be made for other parameters of screening accuracy reported in the study. The estimates of sensitivity, specificity, PPV and NPV for both the SAAS and the GAD-2/7, thus, need to be interpreted with caution. Further testing of the scale in larger samples, with a sufficient number of women assessed with a 'gold standard' structured diagnostic interview, would be highly desirable, as further elaborated in 9.4.

To conclude this section on the limitations of the psychometric study presented in Chapter 8, two further limitations that may be considered lessons to be learned for future studies are briefly noted. It was evident through the issue of under-recruitment for the diagnostic interviews that in a future study, it would be useful to consider inviting at least twice as many study participants to interview as needed. This is primarily based on the observation that, among the 71 women selected to be invited for interview, only 37 eventually took part. Additionally, on reflection, the choice of the EPDS as the scale used to evaluate discriminant validity was probably misjudged, when considering the existing evidence for an anxiety subscale within the measure.

9.3 Implications for policy and clinical practice

Perinatal mental health problems are of major importance as a public health issue (Bauer et al., 2014; Austin & Kingston, 2016; RCGP, 2017). Over the last few years, the importance of perinatal mental health has been increasingly recognised by a number of national and international clinical guidelines (SIGN 2012; NICE, 2014, COPE 2017), key policy and strategic documents (NHS England, 2016; Scottish Government, 2017) and health professional bodies (RCGP, 2017; Royal College of Midwives [RCM], 2017; Royal College of Psychiatrists [RCPSYCH], 2018). In the UK, third sector organisations such as Maternal Mental Health Alliance (MMHA) with its *Everyone's business* campaign (Ayers & Shakespeare, 2015) and Maternal Mental Health Scotland have also contributed to pressure national governments to improve perinatal mental health care.

Recent reports, however, show that the provision of care is patchy (Thompson & Rodell, 2014; Khan, 2015, MMHA, 2017). An NHS report indicated that in England specialist perinatal mental health services are present in less than 15% of all Trusts. Furthermore, 40% of Trusts provide no specialist service at all (NHS Improving Quality, 2015). In Scotland, only two Health Boards have a Mother and Baby Unit and specialist perinatal mental health community teams are only present in seven out of fourteen Health Boards (MMHA, 2017). As a result, although women with severe conditions are more likely to be identified, they are often referred to general adult psychiatric services which may not be well-equipped in the assessment and treatment of perinatal disorders (NICE, 2014). For the most common mental health difficulties such as perinatal anxiety and depression, detection rates are estimated to be lower than 50% (Bauer et al., 2014; Biaggi et al., 2016).

The evidence and findings presented in this thesis have potentially important implications for service commissioners and future clinical guidelines. Despite the variability in estimates of prevalence of antenatal anxiety, the body of research discussed in Chapter 2 highlighted that anxiety is a common mental health problem throughout pregnancy, with the most conservative prevalence estimates based on formal diagnostic criteria for anxiety disorders indicating that approximately 15% of pregnant women experience at least one anxiety disorder (Dennis et al, 2017). The specific, adverse impact of antenatal anxiety on maternal postnatal mental health (i.e. as a strong predictor of postnatal depression) and on a range of birth and child developmental outcomes is also now well-documented in the research literature (see 2.4). The preventative opportunities provided by early identification and support for women experiencing antenatal anxiety are thus evident, particularly when considering the range of adverse health outcomes for mother and child associated with

antenatal anxiety. It was also observed that the negative consequences of clinically significant anxiety during pregnancy have a considerable social and economic cost, well-documented in a recent report from the London School of Economics [LSE] and the Centre for Mental Health detailed in 2.5 (Bauer et al., 2014). The combined evidence discussed above in relation to the high health, social and economic costs associated with poor mental health in the perinatal period, including antenatal anxiety, should alert commissioners of maternity care services and other key stakeholders to take the appropriate steps to improve early identification and support for women experiencing mental health difficulties at this time. These may include, for example, the development of a fully integrated care pathway in perinatal mental health care, and actions aimed to reduce the stigma associated with poor mental health at this time which might, in turn, increase the likelihood of women disclosing problems and seeking support.

The findings presented in Chapter 8 also have potentially important implications for clinical practice. Brief screening tools have become increasingly popular in a range of clinical settings over the last few decades (First, 2008). There are arguably a number advantages to short screening measures such as self-report scales for use in routine maternity care to identify women experiencing poor mental health. This type of measures are relatively inexpensive, generally acceptable to intended respondents and intuitive to complete (Zimmerman et al., 2018). Moreover, standardise information can be gathered and compared at different visits, and validated cut-off scores can provide midwives and other health professionals with an easy-to-interpret summary related to specific psychological symptoms. This, in turn, can trigger a more detailed conversation on the subject, or prompt referral to the appropriate pathway of care if required. The most recent NICE guidelines on antenatal and postnatal mental health (2014) seem to support this approach, with the inclusion of brief scales such as the PHQ-2 and the GAD-2 (with the GAD-7 recommended for further assessment) for the identification of common perinatal mental health problems. However, the findings presented in Chapter 8 and commented in 9.1 with regard to the screening accuracy of the GAD-2 at the cut-off recommended by NICE pose serious questions regarding its appropriateness as a first-line screening measure for problematic anxiety in routine antenatal care. While the cautionary note regarding the limitations of the psychometric validation study discussed earlier need to be considered when interpreting these findings, a recent study published only last year on the screening accuracy of the GAD-2 (Nath et al., 2018) in a pregnant population reported findings consistent with those

presented in Chapter 8, but in a considerably larger sample. Nath and colleagues (2018) tested the GAD-2 in a British sample of over 500 pregnant women, who were all also assessed with a structured diagnostic interview. At the recommended cut-off of 3 or above, while the sensitivity and specificity of the GAD-2 in identifying women experiencing Generalised Anxiety Disorder were acceptable (sensitivity 69%; specificity 91%), the sensitivity of the scale in identifying women experiencing ANY anxiety disorder was poor (sensitivity 26%, specificity 91%). Notably, these parameters of screening accuracy for the GAD-2 are almost identical to those found in the psychometric validation study presented in Chapter 8 at the same cut-off score of ≥ 3 (sensitivity 27%, specificity 96%).

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. This, in turn, has important implications for clinical practice considering that, as noted earlier, using the GAD-2 at this problematic cut-off score is likely to result in a high rate of incorrect classifications, with a high proportion of false negatives. Conversely, the findings presented in Chapter 8 indicate that the SAAS may provide a suitable alternative, as it was found to be an effective screening measure for a range of problematic anxiety symptoms in pregnancy, highly acceptable to the target population and sufficiently brief to meet the NICE prerequisite (i.e. less than 12 items) for being considered as a screening tool to be used in routine maternity care.

9.4 Directions for future research

While the findings regarding the very good screening accuracy of the SAAS can be considered promising, further psychometric validation studies testing a range of psychometric properties of the SAAS need to be carried out before considering the scale for applications in research or clinical settings. Some possible direction for future research are briefly outlined in this section.

Further psychometric testing of the SAAS compared to a unidimensional measure of depression would be desirable to provide further evidence regarding its discriminant validity, and help clarify the issue of the large correlation between the scale and the EPDS. Factor analysis appeared to provide evidence against the distinction between symptoms of general anxiety experienced during pregnancy and pregnancy-related anxiety symptom. Future studies testing the factor structure of the SAAS in different and ideally larger samples would provide further evidence in support or against its structural validity. The test-retest reliability of the scale (i.e. the consistency of scale scores in an individual over subsequent administrations) was not assessed in this research. However, this is an important form of scale reliability, especially for a scale that may be considered for use in clinical settings. The evidence indicating that levels of anxiety symptomatology tend to fluctuate during pregnancy (Dennis et al., 2017) should, however, be taken into account when assessing the test-retest reliability of the SAAS.

With regard to the screening accuracy of the SAAS, its potential for use in the identification of a range of problematic anxiety symptoms in pregnancy was discussed above. Further studies testing its screening performance in other, larger antenatal samples, are obviously required before the scale can be considered for use in clinical practice. During the last year of the PhD documented in this thesis, the NMAHP Research Unit at the University of Stirling was awarded (in collaboration with City University, London) a grant from the National Institute for Health Research (NIHR) to identify the most accurate and acceptable screening scale for the identification of anxiety in routine maternity care. The significant contribution of this thesis to having already undertaken a substantive (published) literature review alongside the work to develop and test the SASS resulted in being invited to join the study team as a co-applicant. This study will test four self-report scales of anxiety and psychological distress in a sample of almost 2,000 women. The screening accuracy of the four scales will be evaluated through structured diagnostic interview with over 400 women, while a range of other psychometric properties (e.g. structural validity, internal consistency) will be assessed on the whole sample. In the final phase of this large, multi-site research, the

feasibility of implementation in routine maternity care of the scale exhibiting the best screening performance and acceptability to the target population will be tested through cases studies in different healthcare services (i.e. maternity services, psychological services and primary care) in England and Scotland. Based on the promising findings documented in this thesis, the SAAS was included among the four measures (GAD-2/7, SAAS, CORE-10, Whooley questions) which will undergo this robust psychometric testing. Consequently, over the course of the next three years important evidence will be gathered regarding a range of psychometric properties of the SAAS and other scales.

9.5 Conclusion

Despite the considerable gaps in the recognition and management of women experiencing perinatal mental health problems, significant steps have been taken in recent years to start addressing these issues. In England, as a result of the *Five Year Forward View for Mental Health* (NHS England, 2016), the Department of Health has invested £365 million in specialist perinatal mental health services for the period 2016-2021 to enable access to specialist perinatal mental health care for at least 30,000 additional women a year. Scotland has also seen an increased emphasis on improving perinatal mental health care in its recently published *Mental Health Strategy 2017-2027* (Scottish Government, 2017). As part of this strategy, a significant step forward has been the introduction in 2017 of a Scottish Perinatal Mental Health Network. The overall aim of this national managed clinical network is to improve the identification and treatment of perinatal mental health problems in Scotland, by identifying current gaps in current provision of care and advising the Scottish Government on service development and pathways of care. The lack of identification of women experiencing poor mental health in the perinatal period, however, currently remains a significant problem. Antenatal and postnatal care provide critical opportunities for early identification and support of women experiencing mental health difficulties at this time. While the early identification and management of health problems is recommended in many clinical areas, this appears to be particularly important in the area of perinatal mental health, in which timely treatment and support can potentially improve health outcomes for both mothers and children (Milgrom & Gemmill, 2015).

The research documented in this thesis 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 very 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 et al., 2014), an effective, acceptable and easy-to-complete screening tool that may be used by midwives and other health professionals to identify pregnant women experiencing a range of problematic anxiety symptoms could be a valuable addition, and allow for more efficient targeting of resources and care.

Reference list

Abell, N., Springer, D.W., & Kamata, A. (2009). *Developing and validating rapid assessment instruments*. New York: Oxford University Press.

Aday, L.A., & Cornelius, L.J. (2006). *Designing and conducting health surveys. A comprehensive guide.* San Franciso: Jossey-Bass.

Aizer, A., Stroud, L., & Buka, S. (2009). Maternal Stress and Child Well-Being: Evidence from Siblings. National Bureau of Economic Research Working Paper Series. *The Journal of Human Resources; 51*(3): 523–555.

Alexander, D. C. (2004). A Delphi study of the trends or events that will influence the future of California charter schools. *Digital Abstracts International, 65* (10), 3629.

American Psychiatric Association [APA]. (2000). Diagnostic and statistical manual of mental disorders: DSM-IV-TR. Washington, DC: American Psychiatric Association

American Psychiatric Association [APA]. (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. Washington, DC: American Psychiatric Association

American Psychiatric Association [APA]. (2013). *Diagnostic and statistical manual of mental disorders: DSM-V*. Arlington, VA: American Psychiatric Association.

Arch, J. J. (2013). Pregnancy-specific anxiety: Which women are highest and what are the alcohol-related risks? *Comprehensive Psychiatry*, *54*(3), 217–228.

Armstrong, D.S. (2004). Impact of prior perinatal loss on subsequent pregnancies. *Journal of Obstetric, Gynecologic, & Neonatal Nursing, 33*, 765–773.

Atkinson, M. J., Sinha, A., Hass, S. L., Colman, S. S., Kumar, R. N., Brod, M., & Rowland, C. R. (2004). Validation of a general measure of treatment satisfaction, the Treatment Satisfaction Questionnaire for Medication (TSQM), using a national panel study of chronic disease. *Health and Quality of Life Outcomes*, *2*(1), 12.

Ayers, S., & Shakespeare, J. (2015). Should perinatal mental health be everyone's business? *Primary Health Care Research & Development*, *16*(04), 323–325.

Austin, M.-P., Hadzi-Pavlovic, D., Leader, L., Saint, K., & Parker, G. (2005). Maternal trait anxiety, depression and life event stress in pregnancy: relationships with infant temperament.

Early Human Development, *81*(2), 183–190.

Austin, M. P. (2004). Antenatal screening and early intervention for "perinatal" distress, depression and anxiety: Where to from here? *Archives of Women's Mental Health*. Springer-Verlag.

Austin, M. P., Tully, L., & Parker, G. (2007). Examining the relationship between antenatal anxiety and postnatal depression. *Journal of Affective Disorders*, *101*(1–3), 169–174.

Austin, M.P., Highet, N. & the Expert Working Group (2017). *Mental Health Care in the Perinatal Period: Australian Clinical Practice Guideline*. Centre of Perinatal Excellence (COPE). Melbourne, VIC, Australia.

Austin, M. P., & Kingston, D. (2016). Psychosocial assessment and depression screening in the perinatal period: Benefits, challenges and implementation.

Ayers, S., Bond, R., Bertullies, S., & Wijma, K. (2016). The aetiology of post-traumatic stress following childbirth: a meta-analysis and theoretical framework. *Psychological Medicine*, *46*(06), 1121–1134.

Barlow, D.H. (2002). *Anxiety and its disorders: the nature and treatment of anxiety and panic (2nd ed.)*. Guilford Press, New York, NY

Barker, D.J. (1998). In utero programming of chronic disease. *Clinical Science, 95*:115–128.

Bar-Shai, M., Gott, D., Kreinin, I., & Marmor, S. (2015). Atypical Presentations of Pregnancy-Specific Generalized Anxiety Disorders in Women Without a Previous Psychiatric Background. *Psychosomatics*, *56*(3), 286–291.

Bartholomew, D., Knotts, M., & Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach.* (3rd ed.). John Wiley & Sons.

Bauer A., Parsonage M., Knapp M., Iemmi V., & Adelaya B. (2014). *Costs of perinatal mental health problems*. London School of Economics and Political Science.

Bayrampour, H., McDonald S., Fung T., Tough S. (2014). Reliability and validity of three shortened versions of the State Anxiety Inventory scale during the perinatal period. *Journal of Psychosomatic Obstetrics & Gynecology, 35*(3): 101–7.

Baxter, A. J., Scott, K. M., Vos, T., & Whiteford, H. A. (2013). Global prevalence of anxiety disorders: A systematic review and meta-regression. *Psychological Medicine*, *43*(5):897-910.

Berger, R. (2015). Now I see it, now I don't: researcher's position and reflexivity in qualitative research. *Qualitative Research*, *15*(2), 219–234.

Berle, J.O., Mykletun, A., Daltveit, A.K., Rasmussen, S., Holsten, F., & Dahl, A.A. (2005). Neonatal outcomes in offspring of women with anxiety and depression during pregnancy: a linkage study from The Nord-Trøndelag Health Study (HUNT) and Medical Birth Registry of Norway. *Archive of Womens Mental Health; 8*(3) 181-9.

Betrán, A. P., Temmerman, M., Kingdon, C., Mohiddin, A., Opiyo, N., Torloni, M. R., … & Downe, S. (2018). Interventions to reduce unnecessary caesarean sections in healthy women and babies. *The Lancet*, *392*(10155), 1358–1368.

Betts, K. S., Williams, G. M., Najman, J. M., & Alati, R. (2014). Maternal depressive, anxious, and stress symptoms during pregnancy predict internalizing problems in adolescence. *Depression and Anxiety*, *31*(1), 9–18.

Biaggi, A., Conroy, S., Pawlby, S., & Pariante, C. M. (2016). Identifying the women at risk of antenatal anxiety and depression: A systematic review. *Journal of Affective Disorders*, *191*, 62–77.

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. *Journal of Affective Disorders*, *197*, 251–258.

Blair, M. M., Glynn, L. M., Sandman, C., & Davis, E. P. (2011). Prenatal maternal anxiety and early childhood temperament. *Stress*, *14*(6), 644–651.

Bland, M. (2000). *An Introduction to Medical Statistics. 3rd edn*. Oxford; New York: Oxford University Press.

Bodecs, T., Szilagyi, E., Cholnoky, P., Sandor, J., Gonda, X., Rihmer, Z., & Horvath, B. (2013). Prevalence and psychosocial background of anxiety and depression emerging during the first trimester of pregnancy: data from a Hungarian po- pulation-based sample. *Psychiatria Danubina, 25*, 352–358.

Bosanquet, K., Bailey, D., Gilbody, S., Harden, M., Manea, L., Nutbrown, S., & McMillan, D. (2015). Diagnostic accuracy of the Whooley questions for the identification of depression: a diagnostic meta-analysis. *BMJ Open*, *5*(12), e008913.

Boulkedid, R., Abdoul, H., Loustau, M., Sibony, O., & Alberti, C. (2011). Using and

reporting the Delphi method for selecting healthcare quality indicators: a systematic review. *PLoS One 6* (6), e20476.

Bowling, A. (2014). *Research methods in health: Investigating health and health services.* Maidenhead, UK: McGraw Hill.

Brace, N., Kemp, R., & Snelgar, R. (2013). *SPSS for psychologists*. London: Routledge.

Braun, V., & Clarke, V. (2016). Using thematic analysis in psychology Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Brouwers, E. P. M., van Baar, A. L., & Pop, V. J. M. (2001a). Does the Edinburgh Postnatal Depression Scale measure anxiety? *Journal of Psychosomatic Research*, *51*, 659–663.

Brouwers, E. P. M., van Baar, A. L., & Pop, V. J. M. (2001b). Maternal anxiety during pregnancy and subsequent infant development. *Infant Behavior and Development*, *24*(1), 95–106.

Brunton, R. J., Dryer, R., Saliba, A., & Kohlhoff, J. (2015). Pregnancy anxiety: A systematic review of current scales. *Journal of Affective Disorders*, *176*, 24–34.

Buderer, N. M. (1996). Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Academic Emergency Medicine, 3,* 895-900.

Cantrill, J.A., Sibbald, B., & Buetow, S. (1996). The Delphi and nominal group techniques in health services research. *International Journal of Pharmacy Practice; 4*:67–74.

Cao-Lei, L., Laplante, D.P., & King S. (2016). Prenatal maternal stress and epigenetics: review of the human research. *Current Molecular Biology Reports, 2*(16)–25.

Capron, L. E., Glover, V., Pearson, R. M., Evans, J., O'Connor, T. G., Stein, A., … & Ramchandani, P. G. (2015). Associations of maternal and paternal antenatal mood with offspring anxiety disorder at age 18 years. *Journal of Affective Disorders*, (187), 20–26.

Carmona Monge, F. J., Peñacoba-Puente, C., Marín Morales, D., & Carretero Abellán, I. (2012). Factor structure, validity and reliability of the Spanish version of the Cambridge Worry Scale. *Midwifery*, *28*(1), 112–119.

Centre for Reviews and Dissemination (CRD) (2009). *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care*. Centre for Dissemination and Reviews.

Chaudron, L. H., & Nirodi, N. (2010). The obsessive–compulsive spectrum in the perinatal period: a prospective pilot study. *Archives of Women's Mental Health*, *13*(5), 403–410.

Child, D. (2006). *The essentials of factor analysis (3rd ed.).* Continuum International Publishing Group.

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric eveidence and taxonomic implications. *Journal of Abnormal Psychology*, *100*(3), 316–336.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues inobjective scale development. *Psychological Assessment,7*, 309-319.

Clark, L. A. & Watson, D. (2003). *Constructing validity: Basic issues in objective scale development*. Washington, D.C.: American Psychological Association.

Coates, R., Ayers, S., & Visser, R. De. (2016). Factor Structure of the Edinburgh Postnatal Depression Scale in a Population-Based Sample. *Psychological Assessment*.

Coolican, H. (2009). *Research methods and statistics in psychology*. London, United Kingdom: Hodder.

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: Theory and application. *American Journal of Medicine*, *119*(2).

Cosco, T.D., Doyle, F., Ward, M., & McGee, H. (2012). Latent structure of the Hospital Anxiety And Depression Scale: A 10-year systematic review. *Journal of Psychosomatic Research; 72*: 180–4.

Costello, A.B., & Osbourne, J.W. (2005). Best practices in explor-atory factor analysis: Four recommendations for getting themost from your analysis. *Practical Assessment, Research, and Evaluation, 10*, 1–9.

Cox, J. L., Holden, J. M., & Sagovsky, R. (1987). Detection of Postnatal Depression: Development of the 10-item Edinburgh Postnatal Depression Scale. *British Journal of Psychiatry*, *150*(06), 782–786.

Craske, M. G., Rauch, S. L., Ursano, R., Prenoveau, J., Pine, D. S., & Zinbarg, R. E. (2009). What is an anxiety disorder? *Depression and Anxiety*, *26*(12), 1066–1085.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*. 16, 297-334.

Data Protection Act. (1998). *Data Protection Act*. London: Information Commissioner.

Davis, E. P., & Sandman, C. A. (2012). Prenatal psychobiological predictors of anxiety risk in preadolescent children. *Psychoneuroendocrinology*, *37*(8), 1224–1233.

Dennis, C. L., Falah-Hassani, K., & Shiri, R. (2017). Prevalence of antenatal and postnatal anxiety: Systematic review and meta-analysis. *British Journal of Psychiatry*, *210*(5), 315–323.

Department of Health [DoH]. (1997). *The Caldicott Committee Report on the Review of Patient Identifiable Information*. London: Department of Health

Department of Health [DoH]. *Research Governance Framework for Health and Social Care, 2nd edn*. London: Department of Health, 2005.

Derogatis, L. (1977). *SCL-90-R. Administration, Scoring and Procedures Manual for the Revised Version of the SCL-90*. John Hopkins University Press.

De Vaus, D. (2014). *Surveys in Social Research*. 6th ed. London: Routledge.

DeVellis, R. F. (2012). *Scale development: Theory and applications*. Thousand Oaks, CA: Sage.

Ding, X. X., Wu, Y. Le, Xu, S. J., Zhu, R. P., Jia, X. M., Zhang, S. F., … & Tao, F. B. (2014, April 20). Maternal anxiety during pregnancy and adverse birth outcomes: A systematic review and meta-analysis of prospective cohort studies. *Journal of Affective Disorders*, *159*:103-10.

DiPietro, J.A., Hilton, S.C., Hawkins, M., Costigan, K.A., & Pressman, E.K. (2002). Maternal stress and affect influence fetal neurobehavioral development. *Developmental Psychology;38*(5):659-68.

Doyle, L., Brady, A.-M., & Byrne, G. (2009). An overview of mixed methods research. *Journal of Research in Nursing*, *14*(2), 175–185.

Dueker, G., Chen, J., Cowling, C., & Haskin, B. (2016). Early Human Development Early developmental outcomes predicted by gestational age from 35 to 41 weeks. *Early Human Development*, *103*, 85–90.

Dunkel-Schetter, C., Lobel, M. (2011). Pregnancy and birth: a multilevel analysis of stress and birth weight. In: Revenson, T., Baum, A., Singer, J., editors: 427-53. *Handbook of health*

psychology 2. Lawrence Erlbaum.

Dunkel Schetter, C., & Tanner, L. (2012). Anxiety, depression and stress in pregnancy: implications for mothers, children, research, and practice. *Current Opinion in Psychiatry*, *25*(2), 141–148.

Edwards, R., & Holland, J. (2013). *What is Qualitative Interviewing?* London: Bloomsbury.

Embretson, S.E., & Reise, S.P. (2000). *Item Response Theory for Psychologists*. London, UK: Lawrence Erlbaum Associates Inc.

Emmelkamp & Thomas Ehring, T. (2014). *The Wiley Handbook of Anxiety Disorders, First Edition*. Edited by Paul & John Wiley & Sons, Ltd. Published, by John Wiley & Sons, Ltd.

Eriksson, C., Jansson, L., & Hamberg, K. (2006). Women's experiences of intense fear related to childbirth investigated in a Swedish qualitative study. *Midwifery*, *22*(3), 240–248.

Evans, K., Spiby, H., & Morrell, C. J. (2015). A psychometric systematic review of self-report instruments to identify anxiety in pregnancy. *Journal of Advanced Nursing*, *71*(9), 1986–2001.

Fairbrother, N., Janssen, P., Antony, M. M., Tucker, E., & Young, A. H. (2016). Perinatal anxiety disorder prevalence and incidence. *Journal of Affective Disorders*, *200*, 148–155.

Faisal-Cury, A., & Rossi Menezes, P. (2007). Prevalence of anxiety and depression during pregnancy in a private setting sample. *Archives of Women's Mental Health, 10*, 25–32.

Fenaroli, V., & Saita, E. (2013). Fear of childbirth: A contribution to the validation of the Italian version of the Wijma Delivery Expectancy/Experience Questionnaire (WDEQ). *TPM - Testing, Psychometrics, Methodology in Applied Psychology*, *20*(2), 131–154.

Field, T.A., Diego, M., Hernandez-Reif, M., Figueiredo, B., Deeds, O., Ascencio, A., … Kuhn, C. (2010). Comorbid depression and anxiety effects on pregnancy and neonatal outcome. *Infant Behaviour and Development, 33*, 23-29.

Flick, U. (2002). *An introduction to qualitative research*. London: Sage.

First, M. B. (Ed.). (2008). Standardized Evaluation in Clinical Practice. Arlington, VA: American Psychiatric Publishing, Inc.

Fisher, J., Cabral de Mello, M., Patel, V., Rahman, A., Tran, T., Holton, S., & Holmes, W. (2011). Prevalence and determinants of common perinatal mental disorders in women in low-

and lower-middle-income countries: a systematic review. *Bulletin of the World Health Organization*, *90*, 139-149.

Francis, J. J., Johnston, M., Robertson, C., Glidewell, L., Entwistle, V., Eccles, M. P., & Grimshaw, J. M. (2010). What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychology & Health*, *25*(10), 1229–1245.

Furr, R. M. (2011). *Scale construction and psychometrics for social and personality psychology*. Thousand Oaks, CA: Sage.

Garthus-Niegel, S., Størksen, H. T., Torgersen, L., Von Soest, T., & Eberhard-Gran, M. (2011). The Wijma Delivery Expectancy/Experience Questionnaire - a factor analytic study. *Journal of Psychosomatic Obstetrics and Gynaecology*, *32*(3), 160–163.

Gavin, N. I., Gaynes, B. N., Lohr, K. N., Meltzer-Brody, S., Gartlehner, G., & Swinson, T. (2005, November). Perinatal depression: A systematic review of prevalence and incidence. *Obstetrics and Gynecology*, *106*(5 Pt 1):1071-83.

Gavin, N., Meltzer-Brody, S., Glover, V., & Gaynes, B.N. (2015). Is Population-Based Identification of Perinatal Depression and Anxiety Desirable?: A Public Health Perspective on the Perinatal Depression Care Continuum. In Milgrom, J. & Gemmill, A.W., editors: 11-31. *Identifying Perinatal Depression and Anxiety*. John Wiley & Sons.

Gellatly, R., & Beck, A. T. (2016). Catastrophic Thinking: A Transdiagnostic Process Across Psychiatric Disorders. *Cognitive Therapy and Research*, *40*(4), 441–452.

Gill, F.J., Leslie, G.D., Grech, C., Latour, J.M. (2013). Using a web-based survey tool to undertake a Delphi study: application for nurse education research. *Nurse Education Today;33*(11):1322-1328.

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 Psychiatrica Scandinavica*, *119*(5), 350–364.

Gitau, R., Menson, E., Pickles, V., Fisk, N. M., Glover, V., & MacLachlan, N. (2001). Umbilical cortisol levels as an indicator of the fetal stress response to assisted vaginal delivery. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, *98*(1), 14–17. 6.

Gladstone, G.L., Parker, G.B., Mitchell, P.B., Malhi, G.S., Wilhelm, K.A., & Austin, M.P.

(2005). A brief measure of worry severity (BMWS): personality and clinical correlates of severe worriers. *Journal of Anxiety Disorders; 19*(8): 877–92.

Glazier, R.H., Elgar, F.J., Goel, V., & Holzapfel, S., (2004). Stress, social support, and emotional distress in a community sample of pregnant women. *Journal of Psychosomatic Obstetrics & Gynecology, 25*, 247–255.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New York: Aldine de Gruyter.

Glover, V. (2014). Maternal depression, anxiety and stress during pregnancy and child outcome; What needs to be done. *Best Practice and Research: Clinical Obstetrics and Gynaecology*, *28*(1), 25–35.

Glover, V. (2016). Maternal stress during pregnancy and infant and child outcomes. In A. Wenzel (Ed.), *The Oxford Handbook of Perinatal Psychology*. New York: Oxford University Press;268-283.

Goodman, J. H., Chenausky, K. L., & Freeman, M. P. (2014). Anxiety Disorders During Pregnancy: A Systematic Review. *The Journal of Clinical Psychiatry*, *75*(10), 1153–1184.

Goodman, J. H., Watson, G. R., & Stubbs, B. (2016). Anxiety disorders in postpartum women: A systematic review and meta-analysis. *Journal of Affective Disorders*, *203*:292-331

Gourounti, K., Lykeridou, K., Taskou, C., Kafetsios, K., & Sandall, J. (2012). A survey of worries of pregnant women: Reliability and validity of the Greek version of the Cambridge Worry Scale. *Midwifery*, *28*(6), 746–753.

Grant, K.-A., McMahon, C., & Austin, M.-P. (2008). Maternal anxiety during the transition to parenthood: a prospective study. *Journal of Affective Disorders*, *108*(1–2), 101–111. https://doi.org/10.1016/j.jad.2007.10.002

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. *Journal of Health Psychology*, *8*(6), 753–764.

Green, S. M., Haber, E., Frey, B. N., & McCabe, R. E. (2015). Cognitive-behavioral group treatment for perinatal anxiety: a pilot study. *Archives of Women's Mental Health*, *18*(4).

Green J, & Murray D. (1994). The use of the Edinburgh Postnatal Depression Scale in

research to explore the relationship between antenatal and postnatal dysphoria. In Cox, J., & Holden, J., editors: 180–198. *Perinatal Psychiatry: Use and Misuse of the Edinburgh Postnatal Depression Scale*. Gaskell.

Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews are Enough? An Experiment with Data Saturation and Variability. *Field Methods, 18*, 59–82.

Guillemin, M., & Gillam, L. (2004). Ethics, Reflexivity, and "Ethically Important Moments" in Research. *Qualitative Inquiry*, *10*(2), 261–280.

Hagberg, H., & Wennerholm, U. B. (2000). Spontaneous premature birth: physiopathology, predictors and management. The frequency is constant--early detection can improve therapeutic possibilities. *Lakartidningen*, *97*(4), 301–306, 308–310.

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. *Sexual and Reproductive Healthcare*, *6*(4), 204–210.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Handelzalts, J. E., Fisher, S., Lurie, S., Shalev, A., Golan, A., & Sadan, O. (2012). Personality, fear of childbirth and cesarean delivery on demand. *Acta Obstetricia et Gynecologica Scandinavica*, *91*(1), 16–21.

Harper, R., & Reeves, B. (1999). Reporting of precision of estimates for diagnostic accuracy: a review. *BMJ, 318,* 1322.

Harvey, N., & Holmes, C. A. (2012). Nominal group technique: An effective method for obtaining group consensus. *International Journal of Nursing Practice*, *18*(2), 188–194.

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, *32*(4), 1008–1015.

Haug, T. T., Mykletun, A., & Dahl, A. A. (2004). The Association Between Anxiety, Depression, and Somatic Symptoms in a Large Population: The HUNT-II Study. *Psychosomatic Medicine*, *66*(6), 845–851.

Heimstad, R., Dahloe, R., Laache, I., Skogvoll, E., & Schei, B. (2006). Fear of childbirth and

history of abuse: Implications for pregnancy and delivery. *Acta Obstetricia et Gynecologica Scandinavica*, *85*(4), 435–440.

Heron, J., O'Connor, T. G., Evans, J., Golding, J., & Glover, V. (2004). The course of anxiety and depression through pregnancy and the postpartum in a community sample. *Journal of Affective Disorders*, *80*(1), 65–73.

Hocher, B. (2014). More than genes: the advanced fetal programming hypothesis. *Journal of Reproductive Immunology*, *104–105*, 8–11.

Hoertel, N., López, S., Peyre, H., Wall, M. M., González-Pinto, A., Limosin, F., & Blanco, C. (2015). Are symptom features of depression during pregnancy, the post partum period and outside the peripartum period didtint? Results from a nationally representative sample using item response theory (IRT). *Depression and Anxiety*, *32*(2), 129–140.

Howard, Louise M, Molyneaux, E., Dennis, C.-L., Rochat, T., Stein, A., & Milgrom, J. (2014). Non-psychotic mental disorders in the perinatal period. *The Lancet*, *384*(9956), 1775–1788.

Howard, Louise Michele, Ryan, E. G., Trevillion, K., Anderson, F., Bick, D., Bye, A., … & Pickles, A. (2018). Accuracy of the Whooley questions and the Edinburgh Postnatal Depression Scale in identifying depression and other mental disorders in early pregnancy. *British Journal of Psychiatry*, *212*(1), 50–56.

Hsu, C.C., Sandford, B.A. (2007). The Delphi technique: making sense of consensus. *Practical assessment, research & 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 Human Development*, *79*(2), 81–91.

Huizink, A. C., Delforterie, M. J., Scheinin, N. M., Tolvanen, M., Karlsson, L., & Karlsson, H. (2016). Adaption of pregnancy anxiety questionnaire–revised for all pregnant women regardless of parity: PRAQ-R2. *Archives of Women's Mental Health*, *19*(1), 125–132.

Hunsley, J. & Mash E. J. (2008). *A guide to assessment that work*. New York: Oxford University Press.

Hunter, A. L., Minnis, H., & Wilson, P. (2011). Altered stress responses in children exposed to early adversity: A systematic review of salivary cortisol studies. *Stress*, *14*(6), 614–626.

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).

Iqbal, S. & Pipon-Young, L. (2009). The Delphi method. *Psychologist; 22*(7): 598–600.

NHS Information Services Division [ISD]. (2017). Retrieved from:

https://www.isdscotland.org/Health-Topics/Maternity-and-Births/Publications/data-tables.asp

Johanson, G. A., & Brooks, G. P. (2010). Initial Scale Development: Sample Size for Pilot Studies. *Educational and Psychological Measurement, 70*(3), 394–400.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research*, *1*(2), 112–133.

Johnson, R., & Slade, P. (2002). Does fear of childbirth during pregnancy predict emergency caesarean section? *BJOG: An International Journal of Obstetrics and Gynaecology*,*109* (11), 1213-1221.

Jomeen, J., & Martin, C. (2004). Is the hospital anxiety and depression scale (HADS) a reliable screening tool in early pregnancy? *Psychology & Health*, *19*(6), 787–800.

Jomeen, J., & Martin, C. R. (2005a). Confirmation of an occluded anxiety component within the Edinburgh Postnatal Depression Scale (EPDS) during early pregnancy. *Journal of Reproductive and Infant Psychology*, *23*(2), 143–154.

Jomeen, J., & Martin, C. R. (2005b). The factor structure of the Cambridge Worry Scale in early pregnancy. *Journal of Prenatal & Perinatal Psychology & Health*, *20*(1), 25–48.

Karimova, G., & Martin, C. (2003). A psychometric evaluation of the Hospital Anxiety and Depression Scale during pregnancy. *Psychology, Health & Medicine*, *8*(1), 89–103.

Keedwell, P., & Snaith, R. P. (1996). What do anxiety scales measure? *Acta Psychiatrica Scandinavica*, *93*(3), 177–180.

Keim, S. A., Daniels, J. L., Dole, N., Herring, A. H., Siega-Riz, A. M., & Scheidt, P. C. (2011). A prospective study of maternal anxiety, perceived stress, and depressive symptoms in relation to infant cognitive development. *Early Human Development*, *87*(5), 373–380.

Kean, J., & Reilly, J. (2014). Item response theory. In: Hammond, F.M., Malec, J.M, Nick, T.G., Buschbacher, R.M., editors. *Handbook for Clinical Research: Design, Statistics and*

Implementation. New York, NY: Demos Medical Publishing; 195–198.

Kessler, R.C., & Üstün, T.B. (2004). The World Mental Health (WMH) Survey Initiative version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research; 13*(2): 93–121.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 593.

Kessler, R. C., Petukhova, M., Sampson, N. A., Zaslavsky, A. M., & Wittchen, H. U. (2012). Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. *International Journal of Methods in Psychiatric Research, 21*, 169–184.

Khan, L. (2015). *Falling through the gaps: Perinatal mental health and general practice*. London, Centre for Mental Health.

Kirschbaum, C., Prussner, J. C., Stone, A. A., Federenko, I., Gaab, J., Lintz, D., … Hellhammer, D. H. (1995). Persistent High Cortisol Responses to Repeated Psychological Stress in a Subpopulation of Healthy Men. *Psychosomatic Medicine*, *57*(5), 468–474.

Klabbers, G. A. (2016). Severe Fear of Childbirth : Its Features, Assesment, Prevalence, Determinants, Consequences and Possible Treatments. *Psychological Topics, 25*, 107–127.

Kline, T. (2005). *Psychological testing : a practical approach to design and evaluation*. Sage Publications.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2003). The Patient Health Questionnaire-2 : validity of a two-item depression screener. *Medical Care*, *41*(11), 1284–1292.

Kumar, A. (2015). Review of the steps for Developnebt of Quantitative Reserach tools. *Journal of Advanced Practices in Nursing*, *1*(1), 10–13.

Lamers, F., van Oppen, P., Comijs, H. C., Smit, J. H., Spinhoven, P., van Balkom, A. J. L. M., … Penninx, B. W. J. H. (2011). Comorbidity Patterns of Anxiety and Depressive Disorders in a Large Cohort Study. *The Journal of Clinical Psychiatry*, *72*(03), 341–348.

Landy, F.J. (1986). Stamp collecting versus science. *American Psychologist. 41*, 1183-1192

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. *Obstetrics and Gynecology*, *110*(5), 1102–1112.

Leeflang, M. M. G., Scholten, R. J. P. M., Rutjes, A. W. S., Reitsma, J. B., & Bossuyt, P. M. M. (2006). Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. *Journal of Clinical Epidemiology*, *59*(3), 234–240.

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? *Journal of Abnormal Child Psychology*, *42*(1), 161–171.

Levin, J. S. (1991). The factor structure of the pregnancy anxiety scale. *Journal of Health and Social Behavior*, *32*(4), 368–381.

Li, Y., & Baser, R. (2012). Using R and WinBUGS to fit a generalized partial credit model for developing and evaluating patient-reported outcomes assessments. *Statistics in Medicine; 31*(18): 2010–26.

Lin, Y., Xu, J., Huang, J., Jia, Y., Zhang, J., Yan, C., & Zhang, J. (2017). Effects of prenatal and postnatal maternal emotional stress on toddlers' cognitive and temperamental development. *Journal of Affective Disorders*, *207*.

Linstone, H. A., & Turoff, M. (Eds.). (2002). *The Delphi method: Techniques and applications* [Electronic version]. Newark, NJ: New Jersey Institute of Technology.

Lobel, M., Cannella, D. L., Graham, J. E., DeVincent, C., Schneider, J., & Meyer, B. A. (2008). Pregnancy-specific stress, prenatal health behaviors, and birth outcomes. *Health Psychology, 27*(5), 604-615.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3*, 635–694.

Lukasse, M., Schei, B., & Ryding, E. L. (2014). Prevalence and associated factors of fear of childbirth in six European countries. *Sexual & Reproductive Healthcare*, *5*(3), 99–106.

MacLellan, J. (2010). Antenatal anxiety in pregnant women from rural Cambodia. *British Journal of Midwifery*, *18*(11), 735–743.

Mancuso, R. A., Schetter, C. D., Rini, C. M., Roesch, S. C., & Hobel, C. J. (2004). Maternal

prenatal anxiety and corticotropin-releasing hormone associated with timing of delivery. *Psychosomatic Medicine*, *66*(5), 762–769.

Marchesi, C., Ampollini, P., Paraggio, C., Giaracuni, G., Ossola, P., De Panfilis, C., Tonna, M., Viviani, D. (2014). Risk factors for panic disorder in pregnancy: a cohort study. *Journal of Affective Disorders, 156*, 134–138.

Marchesi, C., Ossola, P., Amerio, A., Daniel, B. D., Tonna, M., & De Panfilis, C. (2016). Clinical management of perinatal anxiety disorders: A systematic review. *Journal of Affective Disorders*, *190*, 543–550.

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). *British Journal of Clinical Psychology; 31*: 301–306.

Martin, C. R. (2012). *Perinatal Mental Health: A clinical guide*. UK, British Library Catalogue in Publication Data. M&K Update. Kindle Edition.

Martini, J., Petzoldt, J., Einsle, F., Beesdo-Baum, K., Höfler, M., & Wittchen, H. U. (2015). Risk factors and course patterns of anxiety and depressive disorders during pregnancy and after delivery: A prospective-longitudinal study. *Journal of Affective Disorders*, *1*;175:385-95.

Mason, J. (2009). *Qualitative researching*. London: Sage.

Matthey, S., Barnett, B., Howie, P. and Kavanagh, D.J. (2003). Diagnosing postpartum depression in mothers and fathers: whatever happened to anxiety? *Journal of Affective Disorders, 74*, 139–147.

Matthey, S., Fisher, J., & Rowe, H. (2013). Using the Edinburgh postnatal depression scale to screen for anxiety disorders: Conceptual and methodological considerations. *Journal of Affective Disorders*, *146*(2), 224–230.

Matthey, S., Valenti, B., Souter, K., & Ross-Hamid, C. (2013). Comparison of four self-report measures and a generic mood question to screen for anxiety during pregnancy in English-speaking women. *Journal of Affective Disorders*, *148*(2–3), 347–351.

Meades, R., & Ayers, S. (2011). Anxiety measures validated in perinatal populations: A systematic review. *Journal of Affective Disorders*, *133*(1–2), 1–15.

Meltzer-Brody, S., Zerwas, S., Leserman, J., Holle, A. Von, Regis, T., & Bulik, C. (2011). Eating Disorders and Trauma History in Women with Perinatal Depression. *Journal of Women's Health*, *20*(6), 863–870.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741–749.

Milgrom, J., Gemmil, A., Bilszta, J.L., Hayes, B., Barnett, B., Brooks, J., et al. (2008). Antenatal risk factors for postnatal depression: a large prospective study. *Journal of Affective Disorders; 108*(1-2): 147-157.

Milgrom, J., & Gemmill, A. W. (2014). Screening for perinatal depression. *Best Practice and Research: Clinical Obstetrics and Gynaecology*, *28*(1), 13–23.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., & The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Annals of Internal Medicine; 151*(4): 264-70.

Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., & Knol, D.L., et al. (2010a). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*; *63*(7): 737-45.

Mokkink, L.B., Terwee, C.B., & Knol, D.L. (2010b). Stratford, P.W., Alonso, J., Patrick, D.L., et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology; 10*(22): 1-8.

Murray, D., & Cox, J.L. (1990). Screening for depression during pregnancy with the Edinburgh Depression Scale (EDDS). *Journal of Reproductive and Infant Psychology; 8*: 99–107.

Nasreen, H.E., Kabir, Z.N., Forsell, Y., Edhborg, M. (2011). Prevalence and associated factors of depressive and anxiety symptoms during pregnancy: a population based study in rural Bangladesh. *BMC Womens Health, 11*, 22.

Nath, S., Ryan, E. G., Trevillion, K., Bick, D., Demilew, J., Milgrom, J., … Howard, L. M.

(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.

NHS England Mental Health Taskforce. (2016). *Five year forward view for mental health*. London: NHS England.

National Institute for Health and Clinical Excellence (2007). *Antenatal and postnatal mental health: clinical management and service guidance (CG45)*. London: National Institute for Health and Clinical Excellence.

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

National Institute for Health and Care Excellence (2014). *Antenatal and postnatal mental health: clinical management and service guidance (CG192).* London: National Institute for Health and Clinical Excellence.

Netemeyer, R. G., Bearden, W.O. & Sughash, S. (2003). *Scaling procedures: Issues and applications*. Thousand Oaks, Calif.: Sage.

Norhayati, M.N., Hazlina, N.H., Asrenee, A.R., Emilin, W.M. (2015). Magnitude and risk factors for postpartum symptoms: a literature review. *Journal of Affective Disorders, 175*, 34–52.

Norton, P. J., & Paulus, D. J. (2017). Transdiagnostic models of anxiety disorder: Theoretical and empirical underpinnings. *Clinical Psychology Review*, *56*, 122–137.

Nunnally, J. C & Bernstein, L. H. (1994). *Psychometric theory* (2nd ed,). New York: McGraw-Hill.

O'Connor, T.G., Heron, J., Golding, J., & Glover, V. (2003). Maternal antenatal anxiety and behavioural/emotional problems in children: a test of a programming hypothesis. *Journal of Child Psychology and Psychiatry;44*(7):1025-36.

O'Connor, E., Rossom, R. C., Henninger, M., Groom, H. C., & Burda, B. U. (2016). Primary Care Screening for and Treatment of Depression in Pregnant and Postpartum Women.

Journal of American Medical Association JAMA, *315*(4), 388.

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. *Development and Psychopathology*, *26*(2), 393–403.

Office for National Statistics [ONS]. (2017). Retrieved from:

https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths /bulletins/birthcharacteristicsinenglandandwales/2017

Öhman, S. G., Grunewald, C., & Waldenström, U. (2003). Women's worries during pregnancy: Testing the Cambridge Worry Scale on 200 Swedish women. *Scandinavian Journal of Caring Sciences*, *17*(2), 148–152.

Olde, E., van der Hart, O., Kleber, R., & van Son, M. (2006). Posttraumatic stress following childbirth: A review. *Clinical Psychology Review*, *26*(1), 1–16.

Orr, S.T., Blazer, D.G., James, S.A., Reiter, J.P. (2007). Depressive symptoms and indicators of maternal health status during pregnancy. *Journal of Womens Health;16*(4):535–42.

Owen, D.J., Wood, L., Tomenson, B., Creed, F., & Neilson, J.P. (2017). Social stress predicts preterm birth in twin pregnancies. *Journal of Psychosomatic Obstetrics & Gynecology; 38*(1): 63–72.

Pallant, J. (2013). The SPSS Survival Guide 5th Ed. Open University Press. McGraw- Hill Education. Berkshire, England.

Patton, M. Q. (2002). *Qualitative research and evaluation methods 3rd ed*. Thousand Oaks, CA: Sage.

Paulden, M., Palmer, S., Hewitt, C., & Gilbody, S. (2009). Screening for postnatal depression in primary care: cost effectiveness analysis. *British Medical Journal 339*, b5203.

Petersen, J. J., Paulitsch, M. A., Guethlin, C., Gensichen, J., & Jahn, A. (2009). A survey on worries of pregnant women--testing the German version of the Cambridge worry scale. *BMC Public Health*, *9*(1), 490.

Petrie, K. J., & Weinman, J. (2012). Patients' Perceptions of Their Illness. *Current Directions in Psychological Science*, *21*(1), 60–65.

Petrillo, J., Cano, S. J., McLeod, L. D., & Coon, C. D. (2015). Using Classical Test Theory,

Item Response Theory, and Rasch Measurement Theory to Evaluate Patient-Reported Outcome Measures: A Comparison of Worked Examples. *Value in Health*, *18*(1), 25–34.

Phillips, D.J.W. (2002) Endocrine programming and fetal origins of adult disease. *Trends Endocrinology and Metabolism, 13*:363.

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 (PROMIS™): Depression, anxiety, and anger. *Assessment;18*:263–283.

Pinninti, N. R., Madison, H., Musser, E., & Rissmiller, D. (2003). MINI International Neuropsychiatric Schedule: clinical utility and patient acceptance. *European Psychiatry*, *18*(7), 361–364.

Plummer, F., Manea, L., Trepel, D., & McMillan, D (2016). Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis. *Gen Hosp Psychiatry 39:*24–31.

Pluye, P., & Nha Hong, Q. (2014). Combining the Power of Stories and the Power of Numbers: Mixed Methods Research and Mixed Studies Reviews. *Annual Review of Public Health*, *35*, 29–45.

Pop, V.J., Komproe, I.H., & van Son, MJ. (1992). Characteristics of the Edinburgh Post Natal Depression Scale in The Netherlands. *Journal of Affective Disorders; 26*: 105–10.

Prady, S.L., Pickett, K.E., Croudace, T., Fairley, L., Bloor, K., Gilbody, S., Kiernan, K.E., & Wright, J. (2013). Psychological distress during pregnancy in a multi-ethnic community: findings from the born in Bradford cohort study. *PLoS One, 8*, e60693.

Rattray, J., & Jones, M. C. (2007). Essential elements of questionnaire design and development. *Journal of Clinical Nursing*, *16*(2), 234–243.

Reed, V., Gander, F., Pfister, H., Steiger, A., Sonntag, H., Trenkwalder, C., … Wittchen, H.-U. (1998). To what degree does the Composite International Diagnostic Interview (CIDI) correctly identify DSM-IV disorders? Testing validity issues in a clinical sample. *International Journal of Methods in Psychiatric Research*, *7*(3), 142–155.

Remes, O., Brayne, C., van der Linde, R., & Lafortune, L. (2016). A systematic review of reviews on the prevalence of anxiety disorders in adult populations. *Brain and Behavior*, *6*(7), 1–33.

Reynolds, R. M. (2013). Glucocorticoid excess and the developmental origins of disease: Two decades of testing the hypothesis – 2012 Curt Richter Award Winner. *Psychoneuroendocrinology*, *38*(1), 1–11.

Reichardt, C. S., & Rallis, S. F. (1994). Qualitative and quantitative inquiries are not incompatible: A call for a new partnership. *New Directions for Program Evaluation, 61*, 85-91.

Ritchie, J., Lewis, J., & Elam, G. (2003). Designing and selecting samples. In: Ritchie, J., & Lewis, J. eds, *Qualitative research practice: a guide for social science students and researchers*. London: Sage, 77-108.

Riley, W.T., Pilkonis, P., & Cella, D. (2011). Application of the National Institutes of Health Patient-reported Outcome Measurement Information System (PROMIS) to mental health research. *The journal of mental health policy and economics;14*:201–208.

Roesch, S.C., Dunkel-Schetter, C., & Woo, G., et al. (2004) Modeling the types and timing of stress in pregnancy. *Anxiety, Stress & Coping, 17*: 87–102.

Royal College of General Practitioners [RCGP] (2017). *Perinatal mental health is a clinical priority for the College*. RCGP.

Royal College of Obstetricians and Gynaecologists [RCOG] & National Collaborating Centre for Women's and Children's Health, Caesarean Section. (2012). London; Royal College of Obstetricians and Gynaecologists.

Royal College of Obstetricians and Gynaecologists [RCGO]. (2017). *NMPA Project Team*. National Maternity and Perinatal Audit: Clinical Report.

Royal College of Midwives [RCM]. (2017*). Every mother must get the help they need*.

Royal College of Psychiatrists [RCPsych]. (2018). *Routine Clinical Outcome Measurement in Perinatal Psychiatry.* London: Royal College of Psychatrists.

Rose, M., & Devine, J. (2014). Assessment of patient-reported symptoms of anxiety. Dialogues *Clinical Neuroscience;16*:197–211.

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. *Archives of Women's Mental Health*, *6*(1), 51–57.

Ross, Lori E, & McLean, L. M. (2006). Anxiety disorders during pregnancy and the postpartum period: A systematic review. *The Journal of Clinical Psychiatry*, *67*(8), 1285–1298.

Rubertsson, C., Hellström, J., Cross, M., & Sydsjö, G. (2014). Anxiety in early pregnancy: Prevalence and contributing factors. *Archives of Women's Mental Health*, *17*(3), 221–228.

Russel, B. H. & Ryan, G. W. (2010). *Analyzing qualitative data: Systematic approaches*. London: Sage.

Sackett, D.L., & Haynes, R.B. (2002). The architecture of diagnostic research. In: Knottnerus JA, editor. *The evidence base of clinical diagnosis*. London: BMJ Books.

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 Medical Research Methodology, 11,* 62, 1-8.

Sanderson, W. C., Beck, A. T, & Beck, J. (1990). Syndrome comorbidity in patients with major depression or dysthymia: Prevalence and temporal relationships. *American Journal of Psychiatry, 147*, 1025-1028.

Schmied, V., Johnson, M., Naidoo, N., Austin, M.-P., Matthey, S., Kemp, L., … & Yeo, A. (2013). Maternal mental health in Australia and New Zealand: A review of longitudinal studies. *Women and Birth*, *26*(3), 167–178.

Scottish Government. Mental health strategy: 2017-2027. (2017). Scotland: Scottish Government.

Scottish Intercollegiate Guidelines Network (SIGN) (2012). *Management of perinatal mood disorders* (127). SIGN.

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. *Journal of Clinical Psychiatry; 59*: 22–33.

Silverman, D. (2001). *Interpreting qualitative data: Methods for Analysing Talk, Text and Interaction, 2ⁿᵈ*. London: Sage.

Simms, L. J. (2008). Classical and Modern Methods of Psychological Scale Construction.

Social and Personality Psychology Compass, *2*(1), 414–433.

Simpson, H.B., Neria, Y., Lewis-Fernández, R., et al. (2010). *Anxiety disorders: theory, research and clinical perspectives*. New York (NY): Cambridge University Press.

Simpson, W., Glazer, M., Michalski, N., Steiner, M., & Frey, B. N. (2014). Comparative efficacy of the generalized anxiety disorder 7-item scale and the Edinburgh Postnatal Depression Scale as screening tools for generalized anxiety disorder in pregnancy and the postpartum period. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie*, *59*(8), 434–440.

Sit, D., Rothschild, A. J., & Wisner, K. L. (2006). A Review of Postpartum Psychosis. *Journal of Women's Health*, *15*(4), 352–368.

Smith, A. W., & Jensen, R. E. (2019). Beyond methods to applied research: Realizing the vision of PROMIS®. *Health Psychology, 38*(5), 347-350.

Spector, P. E. (1992). *Summated Ratings Scales Construction*. Newbury Park, CA: Sage Publications.

Spielberger, C.D., Gorsuch, R.L., Lushene, R., Vagge, P.R., Jacobs, G.A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Spinelli, M. G. (2009). Postpartum Psychosis: Detection of Risk and Management. *American Journal of Psychiatry*, *166*(4), 405–408.

Spitzer, R.L., Kroenke, K., Williams, J.B., & Löwe B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives Internal Medicine; 166*(10): 1092-7.

Staneva, A., Bogossian, F., Pritchard, M., & Wittkowski, A. (2015a). The effects of maternal depression, anxiety, and perceived stress during pregnancy on preterm birth: A systematic review. *Women and Birth*, *28*(3), 179–193.

Staneva, A. A., Hons, B. A., Psychology, M., & Bogossian, F. (2015b). The experience of psychological distress , depression , and anxiety during pregnancy : A meta-synthesis of qualitative research. *Midwifery*, *31*, 563–573.

Stewart, M. (2001). Towards a global definition of patient centred care. *British Medical Journal (Clinical Research Edition)*, *322*(7284), 444–445.

Streiner, D. L., Norman, G. R., & Cairney, J. (2008). *Health measurement scales: A practical*

guide to their development and use, (4th edn.). Oxford: Oxford University Press.

Streiner, D. L., & Kottner, J. (2014). Recommendations for reporting the results of studies of instrument and scale development and testing. *Journal of Advanced Nursing*, *70*(9), 1970–1979.

Swalm, D., Brooks, J., Doherty, D., Nathan, E., & Jacques, A. (2010). Using the Edinburgh Postnatal Depression Scale to screen for perinatal anxiety. *Archives of Women's Mental Health*, *13*(6), 515–522.

Sutter-Dallay, A.L., Giaconne-Marcesche, V., Glatigry-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. *European Psychiatry, 19* (8): 459–63.

Tabachnick, B.G., & Fidell, L.S. (2007). *Using multivariate statistics (5th ed.).* Pearson Education Inc.

Talge, N. M., Neal, C., & Glover, V. (2007). Antenatal maternal stress and long-term effects on child neurodevelopment: How and why? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *48*(3–4), 245–261.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55.

Tebbe, B. B. M., Terluin, B., & Koelewijn, J. M. (2013). Assessing psychological health in midwifery practice: A validation study of the Four-Dimensional Symptom Questionnaire (4DSQ), a Dutch primary care instrument. *Midwifery*, *29*(6), 608–615.

Tendais, I., Costa, R., Conde, A., & Figueiredo, B. (2014). Screening for depression and anxiety disorders from pregnancy to postpartum with the EPDS and STAI. *The Spanish Journal of Psychology*, *17*, 7.

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. *Journal of Clinical Epidemiology; 60*(1): 34-42.

Terwee, C.B. (2011). *Protocol for systematic reviews of measurement properties*. COSMIN.

Terwee, C.B., Mokkink, L.B., Knol, D.L., Ostelo, R.W., Bouter, L.M., de Vet, HC. (2012).

Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*; **21**(4): 651-7

Tolin, D. F., Gilliam, C., Wootton, B. M., Bowe, W., Bragdon, L. B., Davis, E., … Hallion, L. S. (2018). Psychometric Properties of a Structured Diagnostic Interview for *DSM-5* Anxiety, Mood, and Obsessive-Compulsive and Related Disorders. *Assessment*, *25*(1), 3–13.

Tucker, J., & McGuire, W. (2004). Epidemiology of preterm birth. *BMJ (Clinical Research Ed.)*, *329*(7467), 675–678.

Turner, R. R., Quittner, A. L., Parasuraman, B. M., Kallich, J. D., Cleeland, C. S., & Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. (2007). Patient-reported outcomes: Instrument development and selection issues. Value Health, 10(Suppl 2), S86–S93.

Van den Bergh, B. R. H. (2016). Maternal Anxiety, Mindfulness, and Heart Rate Variability During Pregnancy Influence Fetal and Infant Development. In *Fetal Development* (pp. 267–292). Cham: Springer International Publishing.

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. *Children Development, 75*, 1085–1097.

Van Den Bergh, B. R. H., Mulder, E. J. H., Mennes, M., & Glover, V. (2005). Antenatal maternal anxiety and stress and the neurobehavioural development of the fetus and child: Links and possible mechanisms. A review. *Neuroscience and Biobehavioral Reviews*, *29*(2), 237–258.

Van den Bergh, B. (1991). The influence of maternal emotions during pregnancy on fetal and neonatal behavior. *Pre- Peri-Nat Psychol Journal; 5* :119– 30.

Vanin, J. (2008). *Overview of anxiety and the anxiety disorders. Anxiety Disorders: A Pocket Guide for Primary Care*. Humana Press. Virginia.

Verreault, N., Da Costa, D., Marchand, A., Ireland, K., Dritsa, M., & Khalifé, S. (2014). Rates and risk factors associated with depressive symptoms during pregnancy and with postpartum onset. *Journal of Psychosomatic Obstetrics & Gynecology*, *35*(3), 84–91.

Wadhwa, P.D. (2005). Psychoneuroendocrine processes in human pregnancy influence fetal development and health. *Psychoneuroendocrinology, 30*, 724–43.

Waggoner J, Carline JD, Durning SJ. Is there a consensus on consensus methodology? Descriptions and recommendations for future consensus research. Acad Med. 2016;91:663–668.

Waqas, A., Raza, N., Lodhi, H.W., Muhammad, Z., Jamal, M., & Rehman, A. (2015). Psychosocial factors of antenatal anxiety and depression in Pakistan: is social support a mediator? *PLoS One 10*, e0116510.

Westerneng, M., de Cock, P., Spelten, E. R., Honig, A., & Hutton, E. K. (2015). Factorial invariance of pregnancy-specific anxiety dimensions across nulliparous and parous pregnant women. *Journal of Health Psychology*, *20*(2), 164–172.

Whiting, P., Westwood, M., Beynon, R., Burke, M., Sterne, J. A., & Glanville, J. (2011). Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. *Journal of Clinical Epidemiology*, *64*(6), 602–607.

Widiger, T. A., & Clark, L. A. (2000). Toward DSM—V and the classification of psychopathology. *Psychological Bulletin*, *126*(6), 946–963.

Wijma, K., Wijma, B., Zar, M., Wjma ', K., & Za, M. (1998). Journal of Psychosomatic Obstetrics & Gynecology Psychometric aspects of the W-DEQ; a new questionnaire for the measurement of fear of childbirth Psychometric aspects of the W-DEQ; a new questionnaire for the measurement of fear of childbirth. *Journal of Psychosomatic Obstetrics & Gynecologyl*, *1919*(2), 84–9784.

Wijma, K., Wijma, B. (2017). A woman afraid to deliver: how to manage childbirth anxiety. In *Bio-Psycho-Social Obstetrics and Gynecology*. New York, USA: Springer International Publishing; 3-31.

Williams, B., Brown, T. & Onsman, A. (2010). Exploratory Factor Analysis: A five step guide for novices. *Australasian Journal of Paramedicine. 8*(3), 1-13.

Wilson, S., Draper, H., & Ives, J. (2008). Ethical issues regarding recruitment to research studies within the primary care consultation. *Family Practice*, *25*(6), 456–461.

Wilson, S. & MacLean, R. (2011). *Research methods and data analysis for psychology*. Maidenhead, UK: McGraw Hill.

Witteveen, A. B., De Cock, P., Huizink, A. C., De Jonge, A., Klomp, T., Westerneng, M., & Geerts, C. C. (2016). Pregnancy related anxiety and general anxious or depressed mood and

the choice for birth setting: a secondary data-analysis of the DELIVER study. *BMC Pregnancy and Childbirth*, *16*(1), 363.

Woody, C. A., Ferrari, A. J., Siskind, D. J., Whiteford, H. A., & Harris, M. G. (2017). A systematic review and meta-regression of the prevalence and incidence of perinatal depression. *Journal of Affective Disorders*, *219*:86-92.

World Health Organization. [WHO]. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Geneva: World Health Organization.

World Health Organization [WHO]. (2017). *Depression and other common mental disorders: global health estimates*. Geneva: World Health Organization.

World Medical Association (2013) World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *Journal of American Medical Association (JAMA), 310*:2191–2194.

Worthington, R.L., & Whittaker, T.A (2006). Scale development research: a content analysis and recommendations for best practice. *The Counselling Psychologist, 34*, 6,806-838.

Zhong, Q. Y., Gelaye, B., Zaslavsky, A. M., Fann, J. R., Rondon, M. B., S??nchez, S. E., & Williams, M. A. (2015). Diagnostic validity of the generalized anxiety disorder - 7 (GAD-7) among pregnant women. *PLoS ONE*, *10*(4), 1–18.

Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand 1983; 67: 361–70.

Zimmerman, M, & Mattia, J.I. (2001). The Psychiatric Diagnostic Screening Questionnaire: development, reliability and validity. *Comprehensive Psychiatry; 42*:175–189

Zimmerman, M., Walsh, E., Friedman, M., Boerescu, D.A., & Attiullah, N. (2018). Are self-report scales as effective as clinician rating scales in measuring treatment response in routine clinical practice? *Journal of Affective Disorders, 225*:449–452.

Appendix 1   South East Scotland Research Ethics Committee 02 approval

Lothian NHS Board

**South East Scotland Research
Ethics Committee 02**

**NHS
Lothian**

Waverley Gate
2-4 Waterloo Place
Edinburgh
EH1 3EG
Telephone 0131 536 9000

www.nhslothian.scot.nhs.uk

Date   15 August 2016
Your Ref
Our Ref

Enquiries to: Joyce Clearie
Extension: 35674
Direct Line: 0131 465 5674
Email: Joyce.Clearie@nhslothian.scot.nhs.uk

15 August 2016

Mr Andrea Sinesi
NMAHP Research Unit
Unit 13 Scion House
Stirling University Innovation Park
FK9 4NK

Dear Mr Sinesi

Study title:	**The development and initial validation of a screening scale for antenatal anxiety**
REC reference:	16/SS/0131
Protocol number:	05246
IRAS project ID:	200130

Thank you for your letter of dated 5th August 2016. I can confirm the REC has received the documents listed below and that these comply with the approval conditions detailed in our letter dated 01 August 2016

Documents received

The documents received were as follows:

Document	Version	Date
GP/consultant information sheets or letters [GP letter]	1	04 August 2016
Other [Reply slip - Qualitative interviews - Appendix 11]	2	04 August 2016
Other [Reply slip - Validation study - Appendix 12]	3	04 August 2016
Participant consent form [Consent form - Validation study - Appendix 7]	2	04 August 2016
Participant information sheet (PIS) [PIS Qualitative interviews - Appendix 2]	2	03 August 2016
Participant information sheet (PIS) [PIS Pilot study - Appendix 3]	2	03 August 2016
Participant information sheet (PIS) [PIS Validation study -	2	03 August 2016

Appendix 4]		
Research protocol or project proposal [Research protocol]	2	04 August 2016
Response to Additional Conditions Met [Cover letter responding to FWAC]	1	05 August 2016

Approved documents

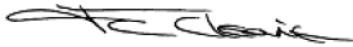The final list of approved documentation for the study is therefore as follows:

Document	Version	Date
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only)	1	29 February 2016
GP/consultant information sheets or letters [Information sheet for midwives - Validation study - Appendix 13]	1	25 May 2016
GP/consultant information sheets or letters [GP letter]	1	04 August 2016
Interview schedules or topic guides for participants [Interview guide - Qualitative interviews - Appendix 14]	1	26 May 2016
Letter from funder [Letter from funder and Annex A]	1	08 July 2015
Letters of invitation to participant [Qualitative interviews - Appendix 8]	1	24 May 2016
Letters of invitation to participant [Pilot study - Appendix 9]	1	23 May 2016
Letters of invitation to participant [Validation study - Appendix 10]	1	23 May 2016
Other [Reply slip - Qualitative interviews - Appendix 11]	2	04 August 2016
Other [Reply slip - Validation study - Appendix 12]	3	04 August 2016
Participant consent form [Consent form - Qualitative interviews - Appendix 5]	1	24 May 2016
Participant consent form [Consent form - Pilot study - Appendix 6]	1	25 May 2016
Participant consent form [Consent form - Validation study - Appendix 7]	2	04 August 2016
Participant information sheet (PIS) [PIS Qualitative interviews - Appendix 2]	2	03 August 2016
Participant information sheet (PIS) [PIS Pilot study - Appendix 3]	2	03 August 2016
Participant information sheet (PIS) [PIS Validation study - Appendix 4]	2	03 August 2016
REC Application Form [REC_Form_01072016]		01 July 2016
Research protocol or project proposal [Research protocol]	2	04 August 2016
Response to Additional Conditions Met [Cover letter responding to FWAC]	1	05 August 2016
Summary CV for Chief Investigator (CI) [CV Chief Investigator]	1	16 June 2016
Summary CV for student [CV student - same as CI]	1	16 June 2016
Summary CV for supervisor (student research) [CVs Academic supervisors]		20 May 2016
Summary, synopsis or diagram (flowchart) of protocol in non technical language [Summary of protocol]	1	22 May 2016

| Validated questionnaire [GAD-7 - Appendix 15] | 1 | 26 May 2016 |

You should ensure that the sponsor has a copy of the final documentation for the study. It is the sponsor's responsibility to ensure that the documentation is made available to R&D offices at all participating sites.

| 16/SS/0131 | **Please quote this number on all correspondence** |

Yours sincerely

Joyce Clearie
SESREC 2 Manager

E-mail: joyce.clearie@nhslothian.scot.nhs.uk

Copy to: *Ms. Joy Taylor*
            Ms Kayleigh Pender, NHS Greater Glasgow and Clyde - Clinical Research and
            Development

Appendix 2   NHS Greater Glasgow & Clyde R&D service approval

NHS
Greater Glasgow and Clyde

Senior Research Administrator: Kayleigh Pender
Telephone Number: 0141 232 1826
E-Mail: Kayleigh.pender@ggc.scot.nhs.uk
website www.nhsggc.org.uk/r&d

Clinical Research & Development
West Glasgow ACH
Dalnair Street
Glasgow G3 8SJ
Scotland, UK

03/10/2016

Mr Andrea Sinesi
University of Stirling
NMAHP Research Unit
Unit 13 Scion House
Stirling University Innovation Park
FK9 4NK

NHS GG&C Board Approval

Dear Mr Sinesi,

Study Title:	The development and initial validation of a screening scale for antenatal anxiety
Principal Investigator:	Mr Andrea Sinesi
GG&C HB site	Princess Royal Maternity Hospital
Sponsor	University of Stirling
R&D reference:	GN16OG073
REC reference:	16/SS/0131
Protocol no: (including version and date)	V2.0 04/08/16

I am pleased to confirm that Greater Glasgow & Clyde Health Board is now able to grant **Approval** for the above study.

Conditions of Approval

1. **For Clinical Trials** as defined by the Medicines for Human Use Clinical Trial Regulations, 2004

    a. During the life span of the study GGHB requires the following information relating to this site

        i. Notification of any potential serious breaches.

        ii. Notification of any regulatory inspections.

It is your responsibility to ensure that all staff involved in the study at this site have the appropriate GCP training according to the GGHB GCP policy (www.nhsggc.org.uk/content/default.asp?page=s1411), evidence of such training to be filed in the site file.

2. **For all studies** the following information is required during their lifespan.

    a. Recruitment Numbers on a quarterly basis

    b. Any change of staff named on the original SSI form

    c. Any amendments – Substantial or Non Substantial

Page 1 of 2	R&D Management Approval Letter	GN16OG073

d.   Notification of Trial/study end including final recruitment figures

e.   Final Report & Copies of Publications/Abstracts

Please add this approval to your study file as this letter may be subject to audit and monitoring.

Your personal information will be held on a secure national web-based NHS database.

I wish you every success with this research study

Yours sincerely,

Kayleigh Pender
Senior Research Administrator

CC: H.Cheyne (Supervisor)

Appendix 3   Participant information sheet for psychometric validation study

TESTING A NEW QUESTIONNAIRE TO

SCREEN FOR ANXIETY IN PREGNANCY

PARTICIPANT INFORMATION SHEET

Introduction

You are being invited to take part in a research study. Before you decide, it is important for you to understand why the research is being done and what taking part would involve for you. Please take time to read the following information carefully and discuss it with your midwife if you wish. Feel free to contact us if you would like more information or if anything is unclear (you can find our contact details at the end of this leaflet).

What is the purpose of the study?

Research studies show that many women experience anxiety in pregnancy. For most women this anxiety is mild but around one in seven women experience high levels of anxiety. We have developed a new questionnaire to assist midwives and other health professionals to recognise pregnant women who are experiencing high levels of anxiety, so that they can be offered the appropriate support. This study aims to find out whether this new questionnaire is useful in helping health professionals to identify which pregnant women are experiencing high levels of anxiety.

Why have I been asked to take part?

We aim to recruit 200 healthy pregnant women attending antenatal clinics at the Princess Royal Maternity and the Queen Elizabeth University Hospital to help us to test the questionnaire. We are aiming to include women from all different stages of pregnancy and we hope that you may be interested in taking part.

Do I have to take part?

No. It is up to you whether you decide to take part. If you do decide to take part you will be given this information sheet, a questionnaire booklet and we will ask you to sign a consent form. If you decide to take part, you can leave the study at any time without giving a reason. This will have no effect on any other care or treatment you are receiving. We are happy to answer any questions you may have before deciding whether you wish to take part in this study.

What will participation in the study involve?

If you decide to take part, we will ask you to complete three short questionnaires. The questionnaires ask about how you have been feeling recently and a few general questions about you and your pregnancy. We expect that filling in the questionnaires should take no longer than 10-15 minutes in total. Once you have completed the questionnaires, reply slip and consent form, we would ask you to post them to the Study Office in the pre-paid envelope provided with this study booklet.

Some women taking part in the study (60 out of 200) will also be contacted to attend a telephone interview (or face-to-face if preferred) with a mental health professional about their experience of anxiety and worries during pregnancy. We are doing these interviews so that we can compare the results of the questionnaires with the clinical judgement of an expert. The invitation to this interview will not be related to any concerns about your mental or physical health. You do not have to take part in this interview if you prefer not to. The interview will take place no longer than 4 weeks after you have posted back the study booklet and we anticipate that it will last between 15 and 20 minutes.

All information in the questionnaires and interviews will be kept completely confidential. It will not be possible to identify you or any information that you, or others, give us on any study reports or other documents. If you wish, when the study is complete we will send you information about the findings. Once the study ends we will not contact you again unless you have indicated to us that you would like information about the study results.

What are the possible disadvantages and risks of taking part?

We do not anticipate any specific risks to you from being involved in the study. Some of the questions on the questionnaires are routinely used in maternity care, and the questionnaires have been reviewed by mental health experts and service users. For this reason, we do not anticipate any question to be particularly distressing or upsetting. However, it is possible that you may feel distressed by the nature of the questions. If this is the case, we would advise you to speak to your GP or midwife in the first instance. If you have any concerns about your emotional wellbeing, you can also contact Breathing Space, a free, confidential phone service for anyone in Scotland experiencing low mood or anxiety. Their phone number is 0800838587 (calls are free from landlines and mobiles) and their website is http://breathingspace.scot/. If

you feel you require out of hours support, please phone NHS24 on 111 or the Samaritans on 0141 248 4488.

As described above, some women completing the questionnaires (60 out of 200) will also be invited to attend a telephone interview (or face-to-face if preferred) with a mental health professional about their experience of anxiety symptoms in pregnancy. If at the end of the interview you or the mental health professional have any concerns about your emotional wellbeing, they will be able to advise you on what to do if you want to access services that can provide specific support. You may also be advised to speak to your midwife, so that she can discuss this with you and, if appropriate, agree a support plan with you.

What are the possible benefits of taking part?

We do not expect the study to have direct benefits for participants but the information we get from this study may help improve the recognition of pregnant women who are experiencing high levels of anxiety. Identifying those women is a first, important step that can lead to an increase of women accessing emotional support when they need it.

What if there is a problem?

If you have a concern about any aspect of this study, please contact the researchers who will do their best to address your concerns. In the first instance contact: Andrea Sinesi, Study Manager, telephone 07522450407, email andrea.sinesi@stir.ac.uk. If you would like to contact an independent party you can do this by contacting Professor Jayne Donaldson (contact details provided at the end of this information sheet). If you remain unhappy and wish to complain formally, you can do this by contacting NHS GG&C at 0141 201 4500 (only complaints).

Will the information I provide be kept confidential?

Yes. All information collected about you will be kept strictly confidential. Paper questionnaires and demographic information will be kept in a locked cupboard, computerised data will be kept on a password protected computer and only those involved in the research will be permitted access to any of the files or data. The reply slip with your name and contact details and the consent form will always be kept separated from the rest of the study booklet. Your name and contact details will only be seen by the Study Manager, and if you are also invited to attend the interview by two members of the NHS GG&C Perinatal Mental Health Service (the administrator that will contact you to arrange the interview and the mental health practitioner conducting the interview). When the results of the study are written up, individuals who have taken part will not be identified in any way.

What will happen to the information I provide?

The results of the study will be written up as part of a doctoral thesis. They may also be published in an academic journal and presented to conferences and other academic events so that others can read and learn from the results of the study. You will not be identified in any report or publication with all data remaining strictly confidential.

Who is funding and organising the study?

The study has been funded by Chief Scientist Office Scotland. The study manager, Mr. Andrea Sinesi, is based in the Nursing, Midwifery and Allied Health Professions Research Unit at the University of Stirling.

Who has reviewed the study?

All research in the NHS is looked at by independent group of people, called a Research Ethics Committee, to protect your interests. This study has been reviewed and given favourable opinion by the South East Scotland Research Ethics Committee 02.

Further information and contact details

If you require further information or have any questions or concerns you can contact:

Researcher	Independent contact
Andrea Sinesi, PhD Candidate	Professor Jayne Donaldson
NMAHP Research Unit	School of Health Sciences
Unit 13 Scion House	University of Stirling
Stirling University Innovation Park	Stirling FK9 4LA
FK9 4NK	Tel: 01786 466354
Email: andrea.sinesi@stir.ac.uk	Fax: 01786 466344
Tel: 07522450407	Email: jayne.donaldson@stir.ac.uk

Appendix 4 Systematic review: search terms and example of search strategy

Overview of the search terms

The following main themes were combined, using the specified free text and MeSH terms, to locate studies potentially eligible for inclusion in the review. A combination of #1 AND #2 AND #3 AND #4 was used.

The four main themes and corresponding terms are presented below:

#1)   Anxiety: anxi* (anxiety, anxious), worr* (worry, worries, worrying), "anxiety disorder", "generali$ed anxiety disorder", GAD, panic, agoraphobia, "social anxiety", phobia, "pregnancy anxiety", "pregnancy-related anxiety", "pregnancy-specific anxiety", "fear of childbirth"

#2)   Pregnancy: Pregnan*(pregnancy, pregnant), antepartum, prepartum, prenatal, antenatal, trimester, perinatal, maternal

#3)   Measurement: Measur* (measure, measurement, measuring), scale, "self-report", rating, rated, instrument, tool, questionnaire, screen* (screen, screening)

#4)   Psychometrics: psychometric, reliability, validity, "content validity", "criterion validity", "concurrent validity", "convergent validity", "discriminant validity", "construct validity", factor* (factor, factors, factorial), "factor structure", "factor analysis", sensitivity, specificity, subscale.

Search strategy on PsycInfo

#	Query
S9	(S1 OR S2) AND (S3 OR S4) AND (S5 OR S6) AND (S7 OR S8)
S8	MM "Psychometrics"
S7	TI (psychometric or reliability or validity or "content validity" or "criterion validity" or "concurrent validity" or "convergent validity"

	or "discriminant validity" or "construct validity" or factor* or "factor structure" or "factor analysis" or sensitivity or specificity or subscale) OR AB (psychometric or reliability or validity or "content validity" or "criterion validity" or "concurrent validity" or "convergent validity" or "discriminant validity" or "construct validity" or factor* or "factor structure" or "factor analysis" or sensitivity or specificity or subscale)
S6	MM "Measurement"
S5	TI (measur* or scale or "self-report" or rating or rated or instrument or tool or questionnaire or screen*) OR AB (measur* or scale or "self-report" or rating or rated or instrument or tool or questionnaire or screen*)
S4	MM "Pregnancy"
S3	TI (pregnan* or antepartum or prepartum or prenatal or antenatal or trimester or perinatal or maternal) OR AB (pregnan* or antepartum or prepartum or prenatal or antenatal or trimester or maternal)
S2	MM "Anxiety" OR MM "Anxiety Disorders"
S1	TI (anxi* or worr* or "anxiety disorder" or "generalized anxiety disorder" or "generalised anxiety disorder" or GAD or panic or agoraphobia or "social anxiety" or phobia or "pregnancy anxiety" or "pregnancy-related anxiety" or "pregnancy-specific anxiety" or "fear of childbirth") OR AB (anxi* or worr* or "anxiety disorder" or "generalized anxiety disorder" or "generalised anxiety disorder" or GAD or panic or agoraphobia or "social anxiety" or phobia or "pregnancy anxiety" or "pregnancy-related anxiety" or "pregnancy-specific anxiety" or "fear of childbirth")

Appendix 5   Systematic review: PRISMA flowchart of the study selection process

Figure 1 – PRISMA flow diagram of the selection process (based on Moher et al.[39])
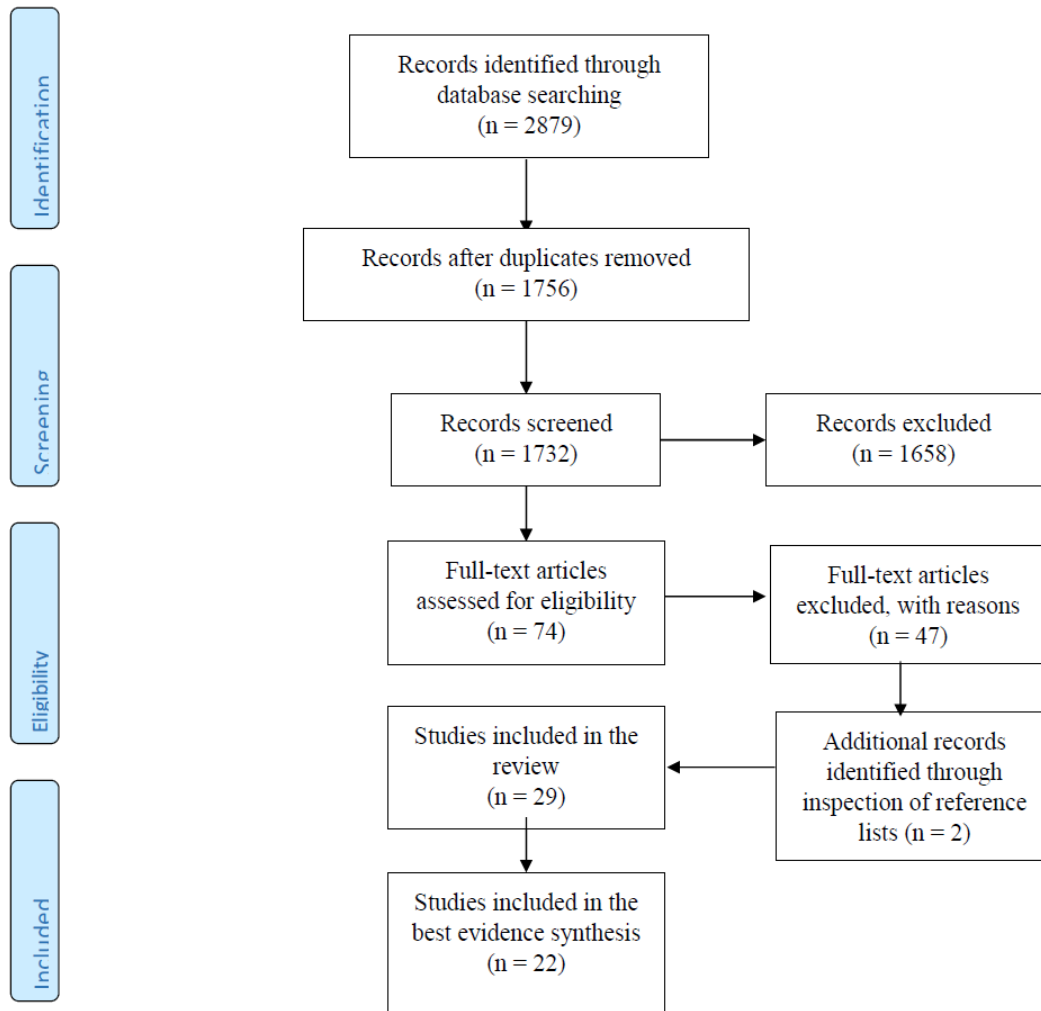
Appendix 6   Systematic review: Methodological quality of all included studies according to COSMIN checklist

Ratings for a study are only reported when a specific psychometric property was evaluated.

An overall score for the methodological quality of a study is determined by using a "worse score counts" system.

INDEX TEST / First author	Internal consistency	Content validity	Structural validity	Hypothesis testing	Criterion validity	Overall quality score
EPDS-A						
Brouwers (2001)	Excellent		Excellent	Good		**Good**
Jomeen (2005a)	Excellent		Excellent			**Excellent**
Swalm (2010)	Good		Good			**Good**
Matthey (2013)	Good			Good	Good	**Good**
Simpson (2014)					Fair	**Fair**
Coates (2016)	Excellent		Excellent			**Excellent**

HADS –A						
Karimova (2003)	Good		Good			**Good**
Jomeen (2004)	Good		Good			**Good**
Matthey (2013)	Good			Good	Good	**Good**
CWS						
Green (2003)	Good	Good	Good	Good		**Good**
Ohman (2003)	Poor					**Poor**
Jomeen (2005b)	Good		Good	Fair		**Fair**
Petersen (2009)	Good	Excellent	Good	Good		**Good**
Carmona Monge (2012)	Excellent	Excellent	Excellent	Excellent		**Excellent**
Gourounti (2012)	Excellent	Excellent	Excellent	Good		**Good**
STAI						
Martheau (1992)	Good	Good				**Good**
Grant (2008)	Good				Excellent	**Good**
Bayrampour (2014)	Excellent		Excellent	Good		**Good**
Tendais (2014)	Poor				Good	**Poor**

W-DEQ						
Wijma (1998)	Good	Excellent		Good		**Good**
Johnson (2002)	Good			Good		**Good**
Garthus-Niegel (2011)	Good		Good	Fair		**Fair**
Fenaroli (2013)	Good		Excellent	Good		**Good**
Haines (2015)	Poor					**Poor**
GAD-7						
Zhong (2015)	Excellent		Excellent		Excellent	**Excellent**
Simpson (2014)					Fair	**Fair**
PAS						
Levin (1991)	Poor		Fair			**Fair**
PRAQ-R/R2						
Huizink (2004)	Good		Good	Good		**Good**
Matthey (2013)	Good			Good	Good	**Good**
Westerneng (2015)	Excellent		Excellent			**Excellent**

Huizink (2016)	Good		Good			**Good**
BMWS Austin (2007)	Good			Good		**Good**

BMWS = Brief Measure of Worry Severity; CWS = Cambridge Worry Scale; EPDS-A = Edinburgh Postnatal Depression Scale – Anxiety subscale; GAD-7 = Generalised Anxiety Disorder – 7; HADS-A = Hospital Anxiety and Depression Scale – Anxiety subscale; PAS = Pregnancy Anxiety Scale; PRAQ-R and PRAQ-R2 = Pregnancy-Related Anxiety Questionnaire- Revised; STAI = State-Trait Anxiety Inventory; W-DEQ = Wijma Delivery Expectancy/Experience Questionnaire

Appendix 7   Anxiety items and domains identified by the systematic review

Measure	Item number / Item	Anxiety domain / Anxiety facet	Studies (First author, year)	Strength of evidence
EPDS	*(3) I have blamed myself unnecessarily when things went wrong*	*General distress* Self-critical	Brouwers, 2001 Jomeen 2005a Swalm 2010	Moderate *(0.56 in Coates 2016)*
EPDS	*(4) I have been worried or anxious for no good reason*	*Anxious apprehension* Worry	Brouwers, 2001 Jomeen 2005a Swalm 2010 Coates 2016	Strong
EPDS	*(5) I have felt scared or panicky for no very good reason*	*Fear/Panic* Scared	Brouwers, 2001 Jomeen 2005a Swalm 2010 Coates 2016	Strong
HADS-A	*(3) I get a sort of frightened feeling as if something awful is going to happen*	*Anxious apprehension* Worried that something bad may happen	Karimova 2003 Jomeen 2004	Strong
HADS-A	*(5) Worrying thoughts go through my mind*	*Anxious apprehension* Worry	Karimova 2003 Jomeen 2004	Strong
HADS-A	*(13) I get sudden feelings of panic*	*Fear/Panic* Fearful/terrified	Karimova 2003 Jomeen 2004	Strong
STAI (State form)	*(3) I am tense*	*Anxious apprehension* Nervous/motor tension	Marteau 1992 Bayrampour 2014	Strong
STAI (State form)	*(6) I am upset*	*General distress* Feeling upset	Marteau 1992 Bayrampour 2014	Strong
STAI (State form)	*(17) I am worried*	*Anxious apprehension* Worry	Marteau 1992 Bayrampour 2014	Strong
GAD-7	*(3) Worrying too much about different things*	*Anxious apprehension* Worry (generalised)	Zhong 2015	Strong

GAD-7	(2) *Not being able to stop or control worrying*	*Anxious apprehension* Worry (Inability to control worrying)	Zhong 2015	Strong
BMWS	*All items*	*Anxious apprehension* Worry	Austin 2007	Strong
CWS	(1)*The possibility of something being wrong with the baby*	*Pregnancy-related anxiety* Worried about baby's health	Green 2003 Petersen 2009 Carmona Monge 2012 Gourounti 2012	Strong
CWS	(2) *Giving birth*	*Pregnancy-related anxiety* Fear of childbirth	Green 2003 Petersen 2009 Carmona Monge 2012 Gourounti 2012	Strong
CWS	(3) *The possibility of miscarriage*	*Pregnancy-related anxiety* Worried about miscarriage	Green 2003 Petersen 2009 Carmona Monge 2012 Gourounti 2012	Strong
CWS	(7) *Coping with the new baby*	*Pregnancy-related anxiety* Worried about future parenting	Green 2003 Carmona Monge 2012 Gourounti 2012	Moderate *(0.58 in Petersen 2009)*
CWS	(8) *Going to hospital*	*Pregnancy-related anxiety* Worried about being in hospital	Green 2003 Petersen 2009 Carmona Monge 2012	Moderate *(0.47 in Gourounti 2012)*
CWS	(9) *Internal examinations*	*Pregnancy-related anxiety* Worried about being in hospital	Green 2003 Petersen 2009 Carmona Monge 2012	Moderate *(0.61 in Gourounti 2012)*
W-DEQ	*Several items*	*Pregnancy-related anxiety* Fear of childbirth	Wijma 1998 Jonhson 2002 Fenaroli 2013	Strong
W-DEQ	*Several items*	*Pregnancy-related anxiety*	Wijma 1998 Jonhson 2002	Strong

			Fenaroli 2013	
		Negative feelings towards childbirth		
W-DEQ	*Several items*	*Pregnancy-related anxiety* Worried about baby's safety	Jonhson 2002 Fenaroli 2013	Moderate *(Contrasting results in Wijma, 1998)*
PRAQ-R	*Two items*	*Pregnancy-related anxiety* Fear of childbirth	Huizink 2004 Westerneng 2015 Huizink 2016	Strong
PRAQ-R	*Three items*	*Pregnancy-related anxiety* Fear about baby's health	Huizink 2004 Westerneng 2015 Huizink 2016	Strong
PRAQ-R	*Three items*	*Pregnancy-related anxiety* Concern about one's appearance	Huizink 2004 Westerneng 2015 Huizink 2016	Strong

BMWS: Brief Measure of Worry Severity; CWS: Cambridge Worry Scale; EPDS-A: Edinburgh Postnatal Depression Scale – Anxiety subscale; GAD-7 = Generalised Anxiety Disorder – 7; HADS-A: Hospital Anxiety and Depression Scale – Anxiety subscale; PRAQ-R: Pregnancy-Related Anxiety Questionnaire- Revised; STAI: State-Trait Anxiety Inventory; W-DEQ: Wijma Delivery Expectancy/Experience Questionnaire

Appendix 8   Participant information sheet for qualitative interviews

TESTING A NEW QUESTIONNAIRE TO

SCREEN FOR ANXIETY IN PREGNANCY

PARTICIPANT INFORMATION SHEET

Introduction

You are being invited to take part in a research study. Before you decide, it is important for you to understand why the research is being done and what taking part would involve for you. Please take time to read the following information carefully. Feel free to contact us if you would like more information or if anything is unclear (you can find our contact details at the end of this leaflet).

What is the purpose of the study?

Research studies show that many women experience anxiety in pregnancy. For most women this anxiety is mild but around one in seven women experience high levels of anxiety. My name is Andrea Sinesi, I am based at the University of Stirling and I am conducting this research as part of a Doctoral degree with a group of experienced researchers. We are developing a new questionnaire to assist midwives and other health professionals to recognise pregnant women who have high levels of anxiety, so that they can be offered the appropriate support. As part of this study, we would like to obtain the views and hear about the experiences of women who have (or have had) anxiety during pregnancy.

Why have I been asked to take part?

You have been asked to take part because you have either past or current experience of anxiety during pregnancy. We are aiming to recruit 15 women with similar symptoms to be interviewed. Your experience of anxiety symptoms during pregnancy will help with the development of this new questionnaire.

Do I have to take part?

No. It is up to you whether you decide to take part. If you are interested in taking part you will be given this information sheet, a consent form and a reply slip. Participation in this research is completely voluntary and you can choose to withdraw from the research at any time. Any care or treatment you may be receiving will not be affected in any way if you decide not to take part now or if you withdraw from the study later. We are happy to answer any questions you may have before deciding whether you wish to take part in this study.

What will participation in the study involve?

If you are interested in taking part, please post the reply slip with your contact details to the Study Office using the enclosed pre-paid envelope. I will then get in touch with you to discuss what the interview would involve and answer any questions you have. If you decide to participate, an interview will be arranged at a time and place of your convenience. The interview will last between 30 and 60 minutes. During this interview, I will ask you questions about your experience of anxiety and worries during pregnancy. I would like to know more about the feelings, thoughts and physical symptoms of women who feel anxious during pregnancy. I will record the interview using an audio-recording device. We will not contact you again unless you have indicated to us that you would like information about the study results. If you decide to participate in the study, you will receive a small gift voucher for taking part.

What are the possible disadvantages and risks of taking part?

We appreciate that the content of this interview could be potentially upsetting or bring back unpleasant memories. You can choose not to answer questions that you do not feel comfortable answering. You can also pause or stop taking part in the interview at any time without giving a reason. I will spend some time at the end of the interview to check how you are feeling and reflect on how you found the interview. If you are feeling upset, I can give you information on what to do if you want to access services that can provide specific support. If you wish I can also contact a mental health professional working for NHS GG&C to discuss a support plan. If you are feeling distressed or upset in the days following the interview, we would advise you to contact and speak to your GP in the first instance. If you have any concerns about your emotional wellbeing, you can also contact Breathing Space, a free, confidential phone service for anyone in Scotland experiencing low mood or anxiety. Their phone number is 0800838587 (calls are free from landlines and mobiles) and their website is http://breathingspace.scot/. If you feel you require out of hours support, please phone NHS24 on 111 or the Samaritans on 0141 248 4488.

What are the possible benefits of taking part?

We do not expect the study to have direct benefits for you but the information we get from this study may help improve the recognition of pregnant women who are experiencing high levels

of anxiety. Identifying those women is a first, important step that can lead to an increase of women accessing emotional support during pregnancy if they need it.

What if there is a problem?

If you have a concern about any aspect of this study, please contact the researchers who will do their best to address your concerns. In the first instance contact: Andrea Sinesi, Study Manager, telephone 07522450407, email andrea.sinesi@stir.ac.uk. If you would like to contact an independent party you can do this by contacting Professor Jayne Donaldson (contact details provided at the end of this information sheet). If you remain unhappy and wish to complain formally, you can do this by contacting NHS GG&C at 0141 201 4500 (only complaints).

Will the information I provide be kept confidential?

Yes. All information collected about you will be kept strictly confidential. The consent form and the reply slip with your contact details will also be kept in a locked cupboard. The audio-recordings of the interview will be typed up into Word documents for analysis by an approved transcriber. No names or identifiable information will be attached to these 'transcripts". The audio-recordings will be erased after being transcribed. Paper transcripts will be kept in a locked cupboard at the University of Stirling. Computerised data including digital recordings will be kept on a password protected computer and only those involved in the research will be permitted access to any of the files or data. When the results of the study are written up, individuals who have taken part will not be identified in any way.

What will happen to the information I provide?

The information you provide will be used to help us develop this new questionnaire to screen for anxiety in pregnancy. The results of the study will be written up as part of a doctoral thesis. They may also be published in an academic journal and presented to conferences and other academic events so that others can read and learn from the results of the study. You will not be identified in any report or publication with all data remaining strictly confidential.

Who is funding and organising the study?

The study has been funded by Chief Scientist Office Scotland. The study manager, Mr. Andrea Sinesi, is based in the Nursing, Midwifery and Allied Health Professions Research Unit at the University of Stirling.

Who has reviewed the study?

All research in the NHS is looked at by independent group of people, called a Research Ethics Committee, to protect your interests. This study has been reviewed and given favourable opinion by the South East Scotland Research Ethics Committee 02.

Further information and contact details

If you require further information or have any questions or concerns you can contact:

Researcher	Independent contact
Andrea Sinesi, PhD Candidate	Professor Jayne Donaldson
NMAHP Research Unit	School of Health Sciences
Unit 13 Scion House	University of Stirling
Stirling University Innovation Park	Stirling FK9 4LA
FK9 4NK	Tel: 01786 466354
Email: andrea.sinesi@stir.ac.uk	Fax: 01786 466344
Tel: 07522450407	Email: jayne.donaldson@stir.ac.uk

Thank you for reading this and considering taking part in this study.

Appendix 9   Reply slip for the qualitative interviews

Participant Identification Number:

STUDY REPLY SLIP

If you are interested in taking part, please post this reply slip to the study office in the envelope provided.

Name: _____

Telephone(s): _____

Best time to call: _____

Email: _____ @ _____

How you would prefer us to contact you   Phone [＿＿]   Email [＿＿]

Please tick the box if you would like to receive information about study findings once the study ends. If you agree, we will email or text you a link to a webpage with a summary of study results. [＿＿]

Appendix 10 Interview guide for qualitative interviews

QUALITATIVE INTERVIEWS - THE EXPERIENCE OF ANXIETY DURING PREGNANCY

SEMI-STRUCTURED INTERVIEW GUIDE

Name of researcher: Andrea Sinesi

Please note that women interviewed will all have a current of past experience of anxiety during pregnancy. This interview guide uses the past tense as we expect that the majority of the women interviewed will be mothers with a past experience of anxiety in pregnancy. Verb tenses will be adapted depending on women's individual circumstances.

The interview guide may be slightly modified with the contributions of the research steering group.

Introduction

Thank you very much for agreeing to meet me today. My name is Andrea Sinesi and I work at the University of Stirling. As you know, we are carrying out some research to develop an anxiety questionnaire that can be used by midwives to recognise pregnant women who have high levels of anxiety. As part of this study I am interviewing women who have had anxiety in pregnancy and this will help us to develop this questionnaire.

I am interested in the experience of women who felt anxious during pregnancy, so I will be asking you some questions about how you felt, what kind of thoughts you had and how anxiety affected you day by day. The interview is very informal and completely confidential. I realise that sometimes people find it difficult speaking with a stranger about personal matters so you don't have to answer questions that you do not feel comfortable answering. You can also pause or stop taking part in the interview at any time and that is absolutely fine. Are you happy to go ahead knowing this?

With your permission I would like to audio-record the interview. This is so that I can concentrate on what you are telling me rather than spending the whole time taking notes. Is that OK? As soon as we have written up the study, all the recording will be destroyed

Do you have any questions before we start?

Initial questions

How many times have you been pregnant?

Have you had similar or different experiences during these pregnancies?

Feelings and physical symptoms

Can you tell me how you felt in pregnancy? *(in case, focus on pregnancy with higher levels of anxiety)*

Explore, as and when appropriate:

> *Differences among trimesters*
>
> *Physical symptoms that participant ascribes to anxiety*
>
> *Sleeping patterns, energy levels*
>
> *Presence of panic attacks/ panic-like symptoms*

What were the feelings that made you feel more distressed? *(give examples: worried, on edge, tense, irritable, excessive fear)*

Did you think that the anxiety you were having was different from the worries pregnant women usually experience? *If yes, explore how/why*

Thoughts

Did you notice any particular thoughts that usually went through your mind when you were feeling anxious?

Explore, as and when appropriate:

> *Obsessive thinking*
>
> *Worry (generalised or specific)*
>
> *Catastrophising*

Behaviours

How did anxiety affect you day to day? (*home, work, social*)

What sort of situations made you feel more anxious?

Some people tend to avoid certain things or situations when they feel anxious. Did anything like this happen to you? *(if yes, explore)*

Pregnancy-related anxiety

Some women during pregnancy have anxieties that are very much specific to their pregnancy (*give examples: about the health of the baby, possible pain during labour and delivery, worries about how they will cope once the baby is born*). Did you find yourself worrying about any of these or similar things? *(if yes, explore)*

Conclusion

We are about to finish with the interview. What is the most important thing that you wanted me to hear from talking with you today?

Is there anything else that we haven't discussed and that you would like to tell me?

How do you feel after talking to me today?

Reiterate that information on how to access support is available in the information sheet.

Assurances about confidentiality

Thank you very much for talking to me today.

Possible probes

Can you give me an example of that?

Can you tell me a bit more about that?

How did that make you feel?

What stands out in your mind about that?

What do you think was the main reason/trigger for that?

Possible empathy statements (to use if and when appropriate)

I can imagine this was really hard for you

I'm sorry to hear that

I realise this must have been distressing for you

Appendix 11 Electronic leaflet introducing the eDelphi

A new screening tool for antenatal anxiety

We need your expert opinion on questions that women should be asked

We are developing a questionnaire to screen for anxiety symptoms in pregnant women:
Reaching consensus among experts

Anxiety in pregnancy is both common and under-recognised. A Scottish Government funded research project at the University of Stirling is looking for your expert view on symptoms that should be included in this new scale for antenatal anxiety.

Next week an online survey will be launched as part of this study. If you are a health professional with at least two years of experience working in perinatal mental health we will ask you to take part. Your opinion will help us develop a screening tool that is reliable and grounded in clinical experience, so please help us by taking part.

An email with a link to register to the survey will be circulated to all members of Maternal Mental Health Scotland.

If you would like further information, please email andrea.sinesi@stir.ac.uk

Appendix 12   Home page of the eDelphi

INTRODUCTION

Thank you for agreeing to take part in this eDelphi to help us develop a screening tool for antenatal anxiety. The Delphi process is designed to combine expert opinion into group consensus and will take place over two rounds of eDelphi.

1st Round – You are now participating in round one. In the following pages you will be presented with a list of statements (e.g. "I felt fearful" or "Many situations made me worry") which represent possible indicators of problematic anxiety symptoms. These statements were formulated based on A) a systematic review of anxiety scales used in studies with pregnant women and B) interviews with women with experience of antenatal anxiety. **You will be asked to rate how important it is for each statement to be considered for inclusion in a screening tool for problematic anxiety symptoms in pregnancy**. At this stage you will also have the opportunity to suggest further questions to be put forward to the panel in the second round. Your responses will be strictly confidential. Further information on how to rate the statements is provided in the following page.

2nd Round - In round two, which will take place approximately two weeks after the beginning of this initial round, you will be shown the distribution of the group's scores from round one as well as your own score. **You will be asked to re-score the statements from round one, taking into consideration the group consensus and your own score**. You will receive an email when round two is launched. It is important that you complete both rounds. Through this process, we aim to arrive at a consensus regarding the most reliable and valid indicators of problematic anxiety symptoms in pregnant women.

A third round may be conducted only if a sufficient level of consensus is not achieved in the first two rounds.

Please register using the link below to access the eDelphi. If possible, we would recommend using a personal email address rather than an NHS or University email address when registering. This is because emails with a reminder of your login details are automatically sent at various stages of the survey process and some web mail servers may block these emails.

1st QUESTIONS PAGE

Below is a list of statements that were identified as possible indicators of problematic anxiety symptoms in pregnancy. The final version of the screening tool will be formatted as a self-report questionnaire, and women will be asked to read each statement and indicate how often they have experienced each symptom **in the past 7 days**. Each statement will have the same **five response options**: Never, Rarely, Sometimes, Often, and Always.

Please rate each item according to how much you consider it to be an important indicator of problematic anxiety in pregnant women, based on your own clinical experience.

If you would also like to comment on a statement, please check the box "*Provide Feedback?*" to display a textbox allowing you to provide feedback. You can complete the survey in one session or save it and return to it at a later time. If you require further information or have any technical difficulties with completing the survey, please do not hesitate to contact me (andrea.sinesi@stir.ac.uk)

Some points to consider when completing the eDelphi

- The statements that you are asked to rate are divided into five anxiety domains, which include both statements for the various anxiety disorders as well as fears and anxieties specific to the antenatal period. However, you will find that some common symptoms of anxiety (in particular physical symptoms such as being easily fatigued or sleep disturbance) are not included as they are common experiences in pregnancy without necessarily being the reflection of poor mental health.

- The final aim is to develop a *screening tool* that is able to detect **problematic anxiety** during the antenatal period rather than a *diagnostic tool* for specific anxiety disorders in the antenatal period. For the purpose of this study, problematic anxiety refers to anxiety that is likely to require clinical attention because of its negative impact on a person's normal routine, occupational functioning or social activities and/or because it causes significant distress. A formal diagnosis would still need to be made by a health professional based on clinical judgement.

- The domain of Pregnancy-related anxiety (i.e. fears and worries that are specific to pregnancy) was included in the list of statements as there is evidence that elevated levels of pregnancy-related anxiety can cause significant distress and impact on a woman's daily functioning.

- You will find that the domain of General distress/Negative affect includes statements assessing symptoms that are not specific only to anxiety disorders (e.g. "I felt upset", "I felt overwhelmed"). These symptoms emerged as consistent themes from the interviews with women who have experienced anxiety in pregnancy so it was considered important to include statements tapping into these symptoms.

- Please consider that the final version of the screening tool will include no more than 11 questions (i.e. statements), as longer scales are unlikely to be feasible and acceptable for routine use in maternity care. The initial list contains 55 statements as it is best practice in scale development to start with a large, over-inclusive pool of questions, which is then gradually reduced in number based on expert opinion (through this eDelphi) and pilot testing. This is why you will find that some statements ask about the same symptom (e.g. "worrying") in a number of different ways. When several statements ask about the same symptom, please consider the content, wording and clarity of statements when rating each statement.

Appendix 13   Information sheet for midwives

nmahp-ru
Nursing, Midwifery and Allied Health Professions | Research Unit
Improving health through research

NHS
Greater Glasgow and Clyde

UNIVERSITY of STIRLING

STUDY TITLE: TESTING A NEW QUESTIONNAIRE TO SCREEN FOR ANXIETY IN PREGNANCY

INFORMATION FOR MIDWIVES: HOW TO RECRUIT WOMEN INTO THE STUDY

1) Briefly describe the study to women who meet the inclusion criteria (you can use the script below to explain the study if you like).

2) Hand out the study pack (the A4 envelope) to women who are interested in taking part. They can complete the questionnaire at home and return it to us in a pre-paid envelope.

WHO TO REFER TO THE STUDY?

Inclusion criteria:	Exclusion criteria:
• Women between 6 and 38 gestational weeks and at least 18 years of age • Receiving routine prenatal care • Level of English sufficient to understand and complete a questionnaire • Able to provide written informed consent to take part in the study	• Major medical/obstetrical complication of pregnancy • Severe cognitive impairment • Current severe mental health condition (any psychotic illness or bipolar disorder as noted in medical history)

BRIEF SCRIPT TO EXPLAIN THE STUDY

We are working with the University of Stirling on a study to help women who have high levels of anxiety during pregnancy. These researchers have developed a short questionnaire to support midwives to identify pregnant women who are very anxious and they are looking for 200 women to fill in three questionnaires to help them to compare them. Some women taking part will also be invited to have a brief telephone interview about symptoms of anxiety. Would you be interested in reading some more information about the study? Participation is completely voluntary.

(If yes) I can give you this study pack, which has all the information you need to decide whether you want to take part. In case you want to participate, you can complete the questionnaires and a consent form - it should take around 5/10 minutes - and then return it to the researchers in a pre-paid envelope.

If you would like more information about the study please contact the Lead researcher
Email: andreasinesi@stir.ac.uk Telephone: 07522450407

Appendix 14   Letter of invitation for pilot study

LETTER OF INVITATION TO PARTICIPATE IN A RESEARCH PROJECT

STUDY TITLE: TESTING A NEW QUESTIONNAIRE TO SCREEN FOR ANXIETY IN PREGNANCY

Dear Madam,

You are being invited to take part in a research study. My name is Andrea Sinesi, I am based at the University of Stirling and I am conducting this research as part of a Doctoral degree, supported by a group of experienced researchers. We are developing a new questionnaire to assist midwives and other health professionals to recognise pregnant women who are experiencing high levels of anxiety, so that they can be offered the appropriate support. We need to test this questionnaire on a whole range of pregnant women.

We have enclosed an information sheet that describes the study in more detail. We hope that it will answer some of the questions you may have about taking part. However if you would like to ask any other questions about the research, please feel free to get in touch by using the contact details provided in the information sheet. While we would greatly appreciate your help in this project, it is important to know that participation is voluntary.

If you decide to take part, we would ask you to:

- Complete the enclosed questionnaire and consent form. We expect that this should take no longer than 5-10 minutes.
- Post them to us in the pre-paid envelope within two weeks of being given this study booklet (for your convenience we have stapled together all the documentation you need to return).

Please keep one copy of the consent form for your own records (2 copies are provided).

Thank you for your time and consideration. We look forward to hearing from you.

With kind regards,

Andrea Sinesi

Appendix 15   Questionnaire for pilot study

Testing a new questionnaire to screen for anxiety in pregnancy

QUESTIONNAIRE

Thank you for taking the time to take part in this study. We are asking you to **complete this questionnaire and the attached consent form, and return them to us in the enclosed pre-paid envelope.** This is still a temporary version of the questionnaire that we are developing. Your answers are completely confidential and will not be shared with any health professionals. However, if you have any concerns about how you are feeling emotionally, in the first instance please speak to your midwife or GP. Further information is provided in the enclosed Participant Information Sheet.

Instructions

The questions below ask about how you have felt in the past week. **Please complete each question by circling or marking ("✓") the number that best describes your experience in the past 7 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 7 days…

	Never	Rarely	Sometimes	Often	Always
1. I worried more than usual	1	2	3	4	5
2. My anxiety stopped me from doing things	1	2	3	4	5
3. I had sudden feelings of panic	1	2	3	4	5
4. I had a racing or pounding heart	1	2	3	4	5
5. I felt detached from pregnancy and the baby	1	2	3	4	5
6. My mind was racing	1	2	3	4	5

In the past 7 days...	Never	Rarely	Sometimes	Often	Always
7. I was much more irritable than usual	1	2	3	4	5
8. I felt panicky for no good reason	1	2	3	4	5
9. I did not feel like myself	1	2	3	4	5
10. I have felt scared about giving birth	1	2	3	4	5
11. I felt that my anxiety made me act impulsively	1	2	3	4	5
12. I found it hard to focus on anything other than my anxiety	1	2	3	4	5
13. I worried about losing my baby	1	2	3	4	5
14. I felt unable to cope	1	2	3	4	5
15. I felt something awful would happen	1	2	3	4	5
16. I worried that something may be wrong with my baby	1	2	3	4	5
17. I felt like I needed help for my anxiety	1	2	3	4	5
18. Thoughts got stuck in my head	1	2	3	4	5
19. I avoided people	1	2	3	4	5
20. I have felt so anxious that I had thoughts about terminating my pregnancy	1	2	3	4	5
21. I could not control my anxiety	1	2	3	4	5
22. I felt so anxious that I had thoughts of ending my life	1	2	3	4	5
23. I felt extremely anxious	1	2	3	4	5

In the past 7 days…	Never	Rarely	Sometimes	Often	Always
24. I have had negative thoughts about childbirth	1	2	3	4	5
25. My worries overwhelmed me	1	2	3	4	5
26. I did not feel worthy of being a mother	1	2	3	4	5
27. I felt like I was losing control	1	2	3	4	5
28. I had a feeling of impending doom	1	2	3	4	5
29. I felt overwhelmed	1	2	3	4	5
30. I needed someone to support me with my anxiety	1	2	3	4	5

Finally, we would like to ask some information about you and how you found filling in the questionnaire:

Your age _____ Weeks pregnant _____ Have you had a previous pregnancy? Yes ☐ No ☐

- From 1 to 10, how easy was the questionnaire to complete?   _____
  1 = not easy at all   10 = extremely easy
- How clear were the instructions?   _____
  1 = not at all clear   10 = extremely clear
- Would you find it acceptable to complete a questionnaire like this as part of routine antenatal care?
  1 = unacceptable   10 = completely acceptable   _____

If you found any of the questions unclear or have any other comments, please write the number of the question and your comment in the textbox below:

```



```

Please return this questionnaire and the attached consent form in the pre-paid envelope provided (no stamp is needed). Thank you for your participation.

Appendix 16   Questionnaire for psychometric validation study

Testing a new questionnaire to screen for anxiety in pregnancy

QUESTIONNAIRE

Thank you for taking the time to take part in this study. **We are asking you to complete this short questionnaire, the attached consent form and reply slip, and return them to us in the enclosed pre-paid envelope.** You may notice that some of the questions are similar. This is because we are comparing the questions to identify which are the most relevant to ask. Your answers are completely confidential and will not be shared with any health professionals. However, if you have any concerns about how you are feeling emotionally, in the first instance please speak to your midwife or GP. Further information is provided in the enclosed Participant Information Sheet.

First, we would like to ask some information about your pregnancy:

Your age _____

How many weeks pregnant are you currently? _____

From 1 to 10, how do you feel about your pregnancy and about giving birth? _____

1 = completely calm   10 = extremely anxious

Have you had a previous pregnancy?   Yes ☐   No ☐

If you had one or more previous pregnancies, have you ever experienced pregnancy or birth complications? If yes, please describe briefly:

_____

_____

Questionnaire for validation study – V1 04/12/2017
This project is funded by Chief Scientist Office Scotland (Project no. DTF/15/03)

311

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

Please tell us how you found filling in these questions:

- From 1 to 10, how easy were these questions to complete? _____
  1 = not easy at all   10 = extremely easy

- Would you find it acceptable to complete these questions as part of routine antenatal care? _____
  1 = unacceptable   10 = completely acceptable

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.**

Over the last 2 weeks, how often have you been bothered by the following problems?

	Not at all	Several days	More than half the days	Nearly every day
1. Feeling nervous, anxious or on edge	0	1	2	3
2. Not being able to stop or control worrying	0	1	2	3
3. Worrying too much about different things	0	1	2	3
4. Trouble relaxing	0	1	2	3
5. Being so restless that it is hard to sit still	0	1	2	3
6. Becoming easily annoyed or irritable	0	1	2	3
7. Feeling afraid as if something awful might happen	0	1	2	3

The questions below ask about how you have felt in the past 7 days. Please complete each question by marking ("✓") the answer that comes closest to how you have felt **IN THE PAST 7 DAYS**, not just how you feel today.

1. I have been able to laugh and see the funny side of things:

☐ As much as I always could
☐ Not quite so much now
☐ Definitely not so much now
☐ Not at all

2. I have looked forward with enjoyment to things:

☐ As much as I ever did
☐ Rather less than I used to
☐ Definitely less than I used to
☐ Hardly at all

3. I have blamed myself unnecessarily when things went wrong:

☐ Yes, most of the time
☐ Yes, some of the time
☐ Not very often
☐ No, never

4. I have been anxious or worried for no good reason:

☐ No, not at all
☐ Hardly ever
☐ Yes, sometimes
☐ Yes, very often

5. I have felt scared or panicky for no very good reason:

☐ Yes, quite a lot
☐ Yes, sometimes
☐ No, not much
☐ No, not at all

6. Things have been getting on top of me:

☐ Yes, most of the time I haven't been able to cope at all
☐ Yes, sometimes I haven't been coping as well as usual
☐ Definitely not so much now
☐ Not at all

7. I have been so unhappy that I have had difficulty sleeping:

☐ Yes, most of the time
☐ Yes, sometimes
☐ Not very often
☐ No, not at all

8. I have felt sad or miserable:

☐ Yes, most of the time
☐ Yes, quite often
☐ Not very often
☐ No, not at all

9. I have been so unhappy that I have been crying:

☐ Yes, most of the time
☐ Yes, quite often
☐ Only occasionally
☐ No, never

10. The thought of harming myself has occurred to me:

☐ Yes, quite often
☐ Sometimes
☐ Hardly ever
☐ Never

Finally, we would like to ask some general information about you:

What is your ethnic group?

1. White ☐

    If White, please specify: Scottish ☐   Other British ☐   Any other white ethnic group ☐

2. Asian/ Asian British ☐

3. Black/ African/ Caribbean/ Black British ☐

4. Mixed/ Multiple ethnic groups ☐

5. Other ethnic group ☐

6. Prefer not to say ☐

What is your highest educational qualification?

1. Level 1 - 'O' Grade, Standard grade or equivalent (SVQ level 1 o 2) ☐

2. Level 2 - Higher, A level or equivalent (SVQ Level 3) ☐

3. Level 3 - HNC/HND or equivalent (SVQ Level 4) ☐

4. Level 4 - Degree, Professional qualification (Above SVQ Level) ☐

5. No qualifications ☐

6. Other qualification (please specify) _____ ☐

What is your marital status?

1. Married ☐

2. Single ☐

3. Cohabiting ☐

4. Divorced ☐

5. Widowed ☐

6. Separated ☐

7. Other (please specify) _____ ☐

Finally, what is your postcode? _____ *(this will only be used to give us some general information about where women taking part in the study live)*

Please return this form and the attached consent form and reply slip in the pre-paid envelope provided (no stamp is needed). Thank you for your participation.

Appendix 17   Consent form for the validation study

Consent form – Study 3

Testing a new questionnaire to screen for anxiety in pregnancy

Name of Researcher: Andrea Sinesi   Email: andrea.sinesi@stir.ac.uk

Participant Identification Number:

Consent Statement	Please initial each box
I confirm that I have read and understand the information sheet dated 04/12/17 (version 4) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	
I understand that my participation is voluntary and may not benefit my own health, and that I am free to withdraw at any time, without giving reason, without my medical care or legal rights being affected.	
I agree that relevant data and my contact details can be stored, confidentially and securely, by the study office at University of Stirling.	
I understand that relevant data collected during the study may be looked at by responsible individuals from the research team or staff from NHS GG&C, where it is relevant to my taking part in research. I give permission for these individuals to have access to data I have provided.	
I understand that the researchers will let my GP know that I am taking part in the research. However, they will not share the questionnaire results or any other information with my GP.	
I understand that I <u>may</u> be asked to take part in a telephone interview with a member of the research team.	
I agree that confidentiality shall only be breached where the interviewer believes that something you have told them places you or others at serious risk. In this instance they are obliged to pass this information on to the relevant persons.	
I understand that the information collected about me will be used to support other research in the future, and may be shared anonymously with other researchers.	
I agree to take part in the above study.	

Name of Participant	Date	Signature

<u>Andrea Sinesi</u>

Name of Researcher	Date	Signature

Thank you for taking part in the study. Please keep a copy of this form for your records.

Appendix 18   Systematic review published on the BJPsych Open

Review

Anxiety scales used in pregnancy: systematic review

Andrea Sinesi, Margaret Maxwell, Ronan O'Carroll and Helen Cheyne

Background

Anxiety disorders and self-reported symptoms are highly prevalent in pregnancy. Despite their negative impact on maternal and child outcomes, uncertainty remains regarding which symptoms can be considered accurate indicators of antenatal anxiety.

Aims

To examine and synthesise the evidence in relation to the psychometric properties and content of self-report scales used to detect anxiety symptoms in pregnant women.

Method

A systematic search was carried out and the methodological quality of all included studies was assessed. Only those achieving a rating of good or excellent were considered in a synthesis of the best available evidence.

Results

Several anxiety symptoms and domains were identified as promising for screening for general antenatal anxiety and pregnancy-related anxiety, including elevated levels of worry, symptoms of panic, fear of childbirth and excessive worries about the baby's health.

Conclusions

This review contributes to the existing knowledge by identifying a number of anxiety symptoms that can be considered psychometrically robust indicators of antenatal anxiety.

Declaration of interest

None.

Keywords

Pregnancy; anxiety disorders; screening; psychometric properties; pregnancy-specific anxiety.

Anxiety during pregnancy is estimated to affect between 15 and 23% of women and is associated with increased risk for a range of negative maternal and child outcomes.[1–3] This has led to growing attention in research[4,5] and clinical guidelines[6] over recent years. Antenatal anxiety has been consistently found to be a strong predictor of postnatal anxiety and depression.[7–11] It has also been linked to adverse birth and child development outcomes, including low birth weight,[12,13] premature birth[3,14,15] and detrimental effects on neurodevelopmental, cognitive and behavioural child outcomes.[4,16,17] Adverse child developmental outcomes found to be associated with antenatal anxiety include, for example, increased risk of language delay,[16] attention-deficit hyperactivity disorder[16] and poorer emotional regulation.[17]

Assessing anxiety in pregnancy

The importance of promoting the detection of women experiencing antenatal anxiety has been reflected in recent clinical guidelines. In the UK, the National Institute for Health and Care Excellence (NICE) guidance on perinatal mental health[6] has for the first time recommended considering use of two screening questions (Generalised Anxiety Disorder scale, GAD-2)[18] for the case-identification of anxiety in pregnant and postnatal women, and the most recent Scottish guidelines have also called for further research in this area.[19] However, the evidence for recommending the GAD-2 is primarily based on its good screening accuracy in the general population,[20] with a very limited evidence base in perinatal populations. Although clinical diagnostic interviews are the optimal method of assessment for anxiety disorders, self-report rating scales such as the GAD-2 are often preferred in busy clinical practice and research because of their brevity.[21]

A recent systematic review found that self-reported anxiety symptoms during pregnancy had a pooled prevalence of 22.9% across trimesters.[1] For anxiety disorders based on DSM or ICD diagnostic criteria[22,23] the overall prevalence was 15.2%. Similar prevalence rates were reported in a number of studies showing that problematic anxiety symptoms affect approximately 15% of women, both in early pregnancy[5] and in later stages.[2,24] High levels of self-reported symptoms, as opposed to anxiety disorders, are of relevance as they have also been shown to be associated with negative maternal and child outcomes.[8,15] In research settings, antenatal anxiety has been measured with a heterogeneity of self-report scales, often in the absence of evidence of their psychometric accuracy in pregnant populations.[25]

Screening for antenatal anxiety using scales developed for the general population is problematic for various reasons, partly as a result of the unique nature of pregnancy. One of the main concerns relates to the emphasis of many self-report measures of general anxiety on somatic symptoms and their potential confounding role when questions around physical symptoms are used to screen for anxiety during pregnancy.[26,27] For instance, questions regarding sleep disturbances or palpitations, which are relatively common during pregnancy, may potentially lead to inflated scores. The assessment of antenatal anxiety is further complicated by the fact that anxiety symptoms that women can experience in pregnancy are not limited to the range of anxiety disorders determined by formal diagnostic criteria.[22,23]

Pregnancy-specific anxiety

The occurrence of pregnancy-specific anxiety has been proposed as a distinct syndrome[28] and a number of studies have investigated this unique anxiety type.[29–31] This emerging construct refers to a

particular anxiety response related to a current pregnancy, which can include fears and worries around labour and delivery, the health of the baby and expected changes in a woman's role.[32] There is now good evidence of the clinical distinctiveness of pregnancy-specific anxiety,[33,34] and some studies indicate that pregnancy-specific anxiety may be a stronger predictor of negative child outcomes than general antenatal anxiety.[34] However, women who may be significantly anxious because of pregnancy-related concerns might not meet the diagnostic criteria for a DSM/ICD anxiety disorder and consequently go unrecognised.

Aims

Recent reviews on the psychometric properties of scales to measure perinatal anxiety have highlighted this gap and the lack of anxiety scales with sound psychometric properties for use with pregnant women.[25,35,36] However, none of these reviews have examined the content of measures with published psychometric data in pregnant populations. Consequently, it remains crucial to establish which symptoms can be considered reliable and valid indicators of maternal antenatal anxiety.

The aim of the present paper was to systematically examine and synthesise both the psychometric properties and content of self-report scales used to assess anxiety in pregnancy in order to identify a core set of anxiety symptoms and anxiety domains with established psychometric properties in pregnant populations. This was achieved by conducting a systematic review of studies reporting at least one psychometric property (i.e. one aspect of reliability or validity) of a self-report measure used to assess antenatal anxiety and by appraising and summarising the best available evidence in the form of a narrative synthesis.

Method

The review was conducted based on guidance for undertaking reviews of clinical tests from the Centre of Reviews and Dissemination[37] and COSMIN (COnsensus-based Standards for the selection of health status Measurement INstruments) recommendations for systematic reviews of measurement properties,[38] and is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.[39] Ethical approval was not required as the study only involved secondary analysis of anonymised data.

Search strategy and selection criteria

Computerised searches were performed to query the following electronic bibliographic databases: MEDLINE, PsycINFO, and the Cumulative Index to Nursing and Allied Health Literature (CINAHL). The initial objective of the review was to locate primary research articles reporting psychometric properties of self-report rating scales used to assess anxiety symptoms in a pregnant population.

The databases were searched from 1991 up to and including February 2017 and searches were restricted to articles published in peer-reviewed journals and available in English. A combination of four main themes was used in the search. Specifically, the major concepts searched were 'anxiety', 'pregnancy', 'measurement' and 'psychometrics' and search terms included both free text and Medical Subject Headings (MeSH) terms. Major concepts and related synonyms for the four main themes were searched in the title and abstract fields, with several key terms also searched as a major concept within each database (see supplementary Appendix 1 available at https://doi.org/10.1192/bjo.2018.75).

Reference lists and citation records of papers included in the review were also inspected for potential inclusion of additional studies. Reports, commentaries, conference proceedings and other grey literature were not searched. Methodological search filters were not applied as there is evidence that, because of the variety of designs used in studies of diagnostic or screening test accuracy, applying methodological filters is likely to result in the omission of a significant number of relevant studies.[40,41] A predefined list of inclusion and exclusion criteria was applied in relation to type of study, population, construct of interest and type of measurement. A complete list of inclusion and exclusion criteria is provided in the Appendix.

Study selection and data extraction

All articles resulting from the electronic bibliographic database searches were imported into RefWorks and duplicates were removed. Titles and abstracts of articles resulting from the initial search were reviewed to identify potentially relevant studies. When there was an indication that an article may have met the inclusion criteria for the review, the full-text publication was obtained and reviewed. The lead reviewer (A.S.) screened titles and abstracts of all retrieved articles to determine their appropriateness for inclusion in the review. A second reviewer (H.C.) independently screened a sample (10%) of all retrieved articles to establish an index of interrater agreement determined as per cent agreement,[42] which was 98% for titles and abstracts screened by both reviewers. Discrepancies were discussed and resolved by applying the relevant study eligibility criteria to reach consensus.

The PRISMA flow diagram[39] was used to document the different stages of the study selection process (Fig. 1). In relation to data extraction, the full-text article of all studies included in the review was inspected and the full version of the rating scale used was obtained in order to extract information relevant to the review. Data extraction forms and summary tables were developed and piloted on a small number of studies ($n = 6$) identified as eligible for inclusion at an early stage of the review.

For each included study the following information was extracted: (a) author/s, (b) year of publication, (c) country, (d) name of index test, (e) sample size, (f) timing of assessment (expressed as trimester or mean gestational week), (g) construct of interest. For each of the rating scales, we extracted: (a) number of items, (b) type and number of response options (for example Likert scale, dichotomous), (c) time frame assessed (for example past week, past month), (d) score range, (e) total possible score, (f) cut-off score (if available). In order to determine which psychometric properties were evaluated in each study, the COSMIN taxonomy and definitions of measurement properties were used.[43] The following psychometric properties were extracted: internal consistency reliability, construct validity, convergent and discriminant validity, structural (i.e. factorial) validity and criterion validity. Definitions of all psychometric properties examined in this review and their corresponding indexes are presented in supplementary Appendix 2.

Quality assessment

An assessment of the methodological quality of each study included in the review was conducted using the COSMIN checklist, specifically developed to evaluate the study quality and risk of bias in systematic reviews of studies on the measurement properties of health measurement instruments.[44] In this review, five of the nine possible boxes in the checklist were employed as they were considered to be relevant to evaluate the methodological quality of studies assessing the construct of anxiety in pregnancy.
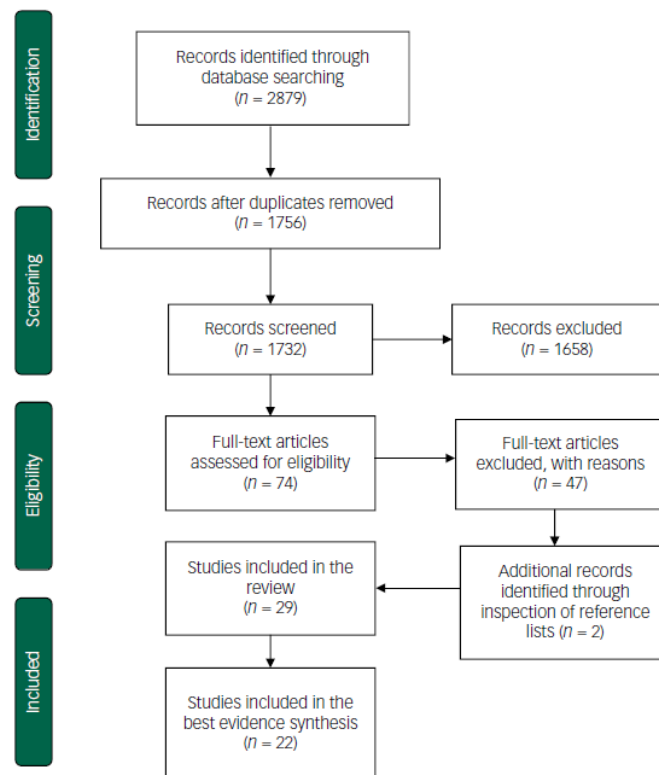
Fig. 1 PRISMA flow diagram of the selection process (based on Moher *et al*).

Specifically, these were box A (internal reliability), D (content validity), E (structural validity), F (hypotheses testing) and H (criterion validity). Each measurement property is scored on a four-point rating scale as 'poor', 'fair', 'good' or 'excellent'. An overall score for the methodological quality of a study is determined by using a 'worse score counts' system.[45] The lead reviewer (A.S.) performed the quality assessment for all studies included in the review, with the second reviewer (H.C.) assessing a random sample of studies (*n* = 5) to confirm the accuracy of the scoring system. It was decided that only studies that achieved an overall rating of good or excellent were considered in the best-evidence synthesis in order to guarantee the quality of the conclusions reached by the review.

Best-evidence synthesis

The main aim of this review was to examine the psychometric properties and content of anxiety measures used in pregnancy, both at the scale and at the item level, in order to identify specific items (i.e. questions) or anxiety domains with established psychometric properties in this population. A synthesis of the best available evidence is presented for each scale in a narrative form, as the considerable differences across studies in relation to measure used, sample size, time of administration and type of reliability or validity reported precluded a meta-analysis. At the scale level, the psychometric properties discussed above were examined and synthesised. The number of studies, their methodological quality and the consistency of findings were taken into account.

Specifically, the following criteria were used to classify the strength of evidence from one or more studies, based on COSMIN recommendations for quality criteria:[46] (a) strong evidence: consistent findings in multiple studies of good or excellent methodological quality or in one study of excellent quality, (b) moderate evidence: consistent findings in multiple studies of good or excellent quality, except for one study with contrasting findings, (c) limited evidence: one study of good methodological quality, and (d) unclear or conflicting evidence: contrasting results in multiple studies of good quality. Only items and anxiety domains with moderate or strong evidence of being accurate indicators of anxiety symptoms in pregnancy were considered psychometrically sound in assessing antenatal anxiety.

At the item level, the analysis was primarily based on factor analysis, and specifically on the examination and comparison of coefficients of item loadings on specific anxiety factors for each scale. In psychometrics, the examination of item loadings is recommended in order to determine which items within a scale possess the strongest psychometric properties in terms of their discriminative power,[47] and can be therefore considered to detect an important aspect of the construct assessed.[48] Factor analysis is used to reduce variables (i.e. single items) that share common variance into set of clusters (i.e. factors).[49]

In this review, the criteria proposed by Tabachnick & Fidell[50] and listed as follows were adopted to evaluate the strength of item loading coefficients: (a) 0–0.44, poor; (b) 0.45–0.54, fair; (c) 0.55–0.62, good; (d) 0.63–0.70, very good; (e) >0.70, excellent. Only items that

319

showed very good or excellent loadings (i.e. 0.63 or above), and for which the strength of evidence from one or multiple studies was moderate or strong according to the criteria discussed above, were considered to be psychometrically sound in measuring anxiety symptoms in pregnancy. When items forming a factor were found to be particularly homogeneous in relation to their content, the entire dimension or domain that the factor represented rather than individual items was selected as a domain identified as psychometrically sound. Secondary indexes that were examined at the item level when factor analysis was not conducted were the correlations between individual items and the remainder of items within a scale (corrected item-total correlations) and item discrimination parameters for analyses based on item-response theory models.

Results

The initial search yielded 2879 citations, which were reduced to 1756 following de-duplication. The titles and abstracts of remaining articles were screened for potentially eligible studies, resulting in 74 publications for which the full-text article was retrieved. At this stage 47 studies were excluded and 2 publications were added from hand searches of reference lists of included studies. This resulted in a final sample of 29 studies included in the review.[8,11,28,31,33,51–74] The main reasons for excluding studies after retrieving the full text were: (a) no psychometric data available, (b) construct of interest different from inclusion criteria (for example antenatal stress, general mental health), (c) study participants recruited exclusively from high-risk samples. The study selection process is summarised in the PRISMA flowchart (Fig. 1).

The 29 included studies used 9 different scales as index tests to measure antenatal anxiety. The most commonly reported psychometric properties were internal consistency reliability ($n = 27$; 93% of studies), convergent validity ($n = 21$; 72%) and structural validity ($n = 16$; 55%). The characteristics of included studies are presented in Table 1. Included studies showed a considerable degree of heterogeneity in relation to the construct assessed (i.e. general anxiety versus an anxiety disorder versus pregnancy-specific anxiety), gestational age of participants, sample size and type of psychometric properties reported.

As discussed in the Method, a quality assessment of all included studies was performed and only studies achieving a rating of good or excellent in relation to their methodological quality and risk of bias were included in the best-evidence synthesis. Seven studies were given a rating of poor[59,68,72] or fair[56,62,65,70] for their methodological quality and were thus not considered in the synthesis. The quality assessment of all 29 studies included in the review is presented in the supplementary Table 1. Further details about the criteria used to rate the methodological quality of all studies included are available from the corresponding author on request.

Best-evidence synthesis

Following an assessment of the methodological quality of all studies, 22 were included in the best-evidence synthesis phase of the review.[8,11,28,31,33,51–55,57,58,60,61,63,64,66,67,69,71,73,74] This section discusses the findings from these studies through an examination of the psychometric properties of each scale and a critical analysis of the content of their items and anxiety domains found to be psychometrically sound for the assessment of antenatal anxiety. This analysis was carried out accordingly to the criteria discussed in detail in the Method. For clarity of exposition, a synthesis is presented here separately for each scale, whereas the Discussion summarises the general findings of the review.

Edinburgh Postnatal Depression Scale (EPDS) – Anxiety subscale

The EPDS[75] is a ten-item self-report questionnaire originally developed to screen for postpartum depression, which asks respondents about symptoms of depression experienced in the previous week. Because of the lack of items specific to the postpartum period, the EPDS has also been validated for use with pregnant women.[76,77] Although the EPDS was developed as a unidimensional measure of depression, it was included in this review because of growing evidence that it contains a separate subscale measuring anxiety rather than depressive symptoms, in both antenatal and postnatal populations.[78–80]

Six studies included in this review examined the psychometric properties of the EPDS anxiety subscale in pregnant women. All studies except one[70] achieved an overall methodological quality rating of good[52,67,71] or excellent[54,63] and were thus included in the best-evidence synthesis. Four of the five studies examined the factor structure of the EPDS to investigate whether the existence of an anxiety subscale could be confirmed.

Brouwers and colleagues[52] performed exploratory factor analysis (EFA) of EPDS scores in women in their second trimester of pregnancy. The EFA revealed three components within the EPDS, namely two separate depressive (items 1, 2, 8) and anxiety (items 3, 4, 5) symptoms subscales and a third component consisting only of item 10 ('The thought of harming myself has occurred to me'). However, this third factor was not included in the final factor solution as the authors argued that a single-item loading could not plausibly identify a distinct latent factor.[52] A two-factor solution, comprising separate depression and anxiety subscales, was therefore proposed. The three items of the anxiety subscale (item 3 'I have blamed myself unnecessarily when things went wrong', item 4 'I have been anxious or worried for no good reason', item 5 'I have felt scared or panicky for no very good reason') were the only ones, among the ten EPDS items, with item loadings on a single factor above the predefined cut-off of 0.63, ranging from 0.68 (item 3) to 0.73 (item 4). An examination of their content appears to indicate that these questions, all loading highly on a single factor, tap important affective and cognitive components of anxiety (for example feeling panicky or worried).

Similar findings were reported by Jomeen & Martin[63] in women in their first trimester of pregnancy. EFA resulted in a three-factor solution that included depression and anxiety dimensions, and the same third factor identified by Brouwers and colleagues.[52] The items loading significantly (>0.63, range 0.73–0.85) onto the anxiety subscale were entirely consistent (items 3, 4, 5) with those identified in the previous study.[52] The authors then conducted confirmatory factory analysis (CFA), a more refined data reduction technique than EFA,[81] and tested various predefined factor models including the original unidimensional depression model,[75] as well as both a two- and a three-factor model identified by Brouwers and colleagues.[52] Results from the CFA revealed once again a clear superiority of the two-factor solution, thus confirming the previous finding that the EPDS both in early and in mid-pregnancy consistently measures two distinct dimensions of depression and anxiety.

A further study included in this review[67] used the three-item EPDS anxiety subscale (EDS-3A) identified in previous studies to examine its criterion and convergent validity in pregnancy when compared with other anxiety measures. The EDS-3A performed better than both the anxiety subscale of the Hospital Anxiety and Depression Scale (HADS-A[82]) and the Pregnancy Related Anxiety Questionnaire-Revised (PRAQ-R[28]) in detecting women with an anxiety disorder as determined by DSM diagnostic criteria. Furthermore, the EDS-3A showed a moderately high correlation

Table 1 General characteristics of studies included in the review

Authors	Year	Participants, n	Gestational age	Country	Index test	Time frame assessed	Target construct
Austin et al[11]	2007	748	3rd trimester	Australia	BMWS	'General experience'	Worry
Bayrampour et al[51]	2014	3021	2nd trimester	Canada	STAI (3 six-item short forms)	State (present time); trait (general feelings)	State/trait anxiety
Brouwers et al[52]	2001	197	24 weeks	Netherlands	EPDS-A	Previous week	General anxiety
Carmona Monge et al[53]	2012	285	Mean 14.1 weeks	Spain	CWS	Present time	Worry during pregnancy
Coates et al[54]	2016	5551	18 and 32 weeks	UK	EPDS-A	Previous week	General anxiety
Fenaroli & Saita[55]	2013	522	27–35 weeks	Italy	W-DEQ	Current expectations about childbirth	Fear of childbirth
Garthus-Niegel et al[56]	2011	1642	32 weeks	Norway	W-DEQ	Current expectations about childbirth	Fear of childbirth
Gourounti et al[57]	2012	132	11–14 weeks	Greece	CWS	Present time	Worry during pregnancy
Grant et al[8]	2008	100	35–39 weeks	Australia	STAI (state and trait forms)	State (present time); trait (general feelings)	State/trait anxiety
Green et al[58]	2003	1207	1st/2nd/3rd trimester	UK	CWS	Present time	Worry during pregnancy
Haines et al[59]	2015	1410	2nd trimester	Australia	W-DEQ	Current expectations about childbirth	Fear of childbirth
Huizink et al[28]	2004	172	1st/2nd/3rd trimester	Netherlands	PRAQ-R	Present time	Pregnancy-related anxiety
Huizink et al[31]	2016	1144	24 and 34 weeks	Finland	PRAQ-R2	Present time	Pregnancy-related anxiety
Johnson & Slade[60]	2002	424	3rd trimester	UK	W-DEQ	Current expectations about childbirth	Fear of childbirth
Jomeen & Martin[61]	2004	101	Mean 13.57 weeks	UK	HADS-A	Previous week	General anxiety
Jomeen & Martin[62]	2005	129	Mean 13.86 weeks	UK	CWS	Present time	Worry during pregnancy
Jomeen & Martin[63]	2005	101	Mean 13.57 weeks	UK	EPDS-A	Previous week	General anxiety
Karimova & Martin[64]	2003	100	12 and 34 weeks	UK and Uzbekistan	HADS-A	Previous week	General anxiety
Levin[65]	1991	266	2nd trimester	USA	PAS	Present time	Pregnancy-related anxiety
Marteau & Bekker[66]	1992	200	Gestational age not reported	UK	STAI (six-item short form)	Present time	State anxiety
Matthey et al[67]	2013	132	Mean 14.9 weeks	Australia	EPDS-A, HADS-A, PRAQ-R	Various time frames	General and pregnancy-related anxiety
Öhman et al[68]	2003	200	8–42 weeks	Sweden	CWS	Present time	Worry during pregnancy
Petersen et al[69]	2009	344	Mean 31.4 weeks	Germany	CWS	Present time	Worry during pregnancy
Simpson et al[70]	2014	240	1st trimester	Canada	EPDS-A, GAD-7	Previous week	General anxiety
Swalm et al[71]	2010	4706	10–42 weeks	Australia	EPDS-A	Previous week	General anxiety
Tendais et al[72]	2014	148	1st/2nd/3rd trimester	Portugal	STAI-S	Present time	State anxiety
Westerneng et al[33]	2015	6004	Mean 19.8 weeks	Netherlands	PRAQ-R	Present time	Pregnancy-related anxiety
Wijma et al[73]	1998	196	32 weeks	Sweden	W-DEQ	Current expectations about childbirth	Fear of childbirth
Zhong et al[74]	2015	946	Mean 9.6 weeks	Peru	GAD-7	Previous two weeks	Generalised anxiety disorder

BMWS, Brief Measure of Worry Severity; STAI, State-Trait Anxiety Inventory; EPDS-A, Edinburgh Postnatal Depression Scale – Anxiety subscale; CWS, Cambridge Worry Scale; W-DEQ, Wijma Delivery Expectancy/Experience Questionnaire; PRAQ-R and PRAQ-R2, Pregnancy-Related Anxiety Questionnaire-Revised; HADS-A; Hospital Anxiety and Depression Scale – Anxiety subscale; PAS, Pregnancy Anxiety Scale; GAD-7, Generalised Anxiety Disorder – 7.

with the HADS-A ($r = 0.68$) and a low to moderate correlation with the PRAQ-R ($r = 0.23$), which may be interpreted as an indication that the three measures tap into different aspects of antenatal anxiety.

Although a potential limitation of the three studies reported above is their relatively small number of participants ($n < 200$), the existence of an anxiety subscale within the EPDS was further confirmed in two subsequent studies with much larger numbers of participants ($n > 4000$). Swalm and colleagues[71] examined the EPDS factor structure in Australian women across the three trimesters of pregnancy. A two-factor solution consisting of anxiety and depression components was found once more to be optimal, accounting for 55% of the score variance (anxiety subscale 29.4%; depression subscale 25.4% of the total variance). Moreover, an analysis of individual item loadings confirmed that items 3, 4 and 5 were the only ones with loadings higher than 0.63 on the anxiety subscale (range 0.75–0.78).

A recent UK population-based study[54] conducted both EFA and CFA on a large number of participants at two time points (18 and 32 weeks' gestation). Although both EFA and CFA indicated a three-factor model as the best factor solution, this was primarily because of the 'depression' factor that was split into an anhedonia (items 1 and 2) and a depression (items 7–10) factor. Importantly, this was the only study in which item 3 'I have blamed myself unnecessarily when things went wrong' (0.56) did not reach the pre-defined item loading coefficient of 0.63.

In summary, according to the criteria previously discussed to evaluate the strength of evidence in relation to the psychometric properties of reviewed scales, item 3 of the EPDS showed moderate evidence of its psychometric value, and items 4 and 5 demonstrated strong evidence of being psychometrically sound in assessing antenatal anxiety, as their item loadings on the anxiety subscale consistently exceeded the 0.63 cut-off in all reviewed studies.

HADS – Anxiety subscale

The HADS[82] is a widely popular screening tool[83] originally developed to assess anxiety and depression in non-psychiatric patients. This 14-item measure consists of two subscales (anxiety: HADS-A; depression: HADS-D), both comprising seven items and enquiring about feelings over the past week with four response options.[82] It is particularly important to establish the psychometric properties of the HADS when used in the antenatal period, as a considerable number of studies have used this screening tool to assess anxiety and depression levels in pregnant women, including in recent years.[5,84]

Three studies included in this review examined psychometric aspects of the HADS in a pregnant population.[61,64,67] They all achieved a rating of good in relation to their methodological quality. Karimova & Martin[64] investigated the factor structure of the HADS in the third trimester of pregnancy by conducting EFA of HADS scores in nulliparous women, and a *post hoc* factor analysis revealed a two-factor solution. Specifically, six of the seven HADS-D items loaded higher on one factor and an equal number of HADS-A items loaded higher on a second factor. However, there was significant overlapping of item loadings on the two subscales, with only four HADS-A items (item 3 'I get a sort of frightened feeling as if something awful is going to happen'; item 5 'Worrying thoughts go through my mind'; item 9 'I get a sort of frightened feeling like "butterflies" in the stomach' and item 13 'I get sudden feelings of panic') loading above 0.63 on the anxiety factor. The authors therefore concluded that the seven-item HADS-A and HADS-D subscales do not reliably distinguish between anxiety and depressive symptoms in pregnancy.

Table 2 Item loading coefficients of the Hospital Anxiety and Depression Scale – Anxiety subscale (HADS-A) subscale in Karimova & Martin and Jomeen & Martin

HADS-A items	Karimova & Martin (2003) Anxiety factor (factor 2)	Jomeen & Martin (2004) Anxiety factor (factor 2)
1. I feel tense or wound up	0.18	0.31
3. I get a sort of frightened feeling as if something awful is going to happen	0.67[a]	0.74[a]
5. Worrying thoughts go through my mind	0.78[a]	0.69[a]
7. I can sit at ease and feel relaxed	0.33	0.07
9. I get a sort of frightened feeling like 'butterflies' in the stomach	0.65[a]	0.57
11. I feel restless as if I have to be on the move	0.57	0.36
13. I get sudden feelings of panic	0.67[a]	0.75[a]
a. Item loadings of 0.63 or above.		

A further study was conducted by Jomeen & Martin[61] on women in early pregnancy. Both EFA and CFA revealed a three-factor solution that confirmed that the HADS in pregnancy is not a bi-dimensional measure of anxiety and depression. However, a comparison of individual item loadings of the HADS anxiety subscale in the two studies was carried out in this review to examine psychometric information for each individual item within the HADS anxiety subscale. This is presented in Table 2.

The observation that three items of the HADS-A (items 3, 5, 13) are the only ones to reach an item loading above 0.63 on the anxiety factor in both studies is of particular importance. Although the two studies reached the conclusion that the seven-item HADS-A as a whole is not a psychometrically sound measure of anxiety in pregnancy, the three HADS-A items identified here showed a consistent pattern across the two studies, with significantly similar loadings on the anxiety factor. These items would therefore appear to have good psychometric value in assessing specific anxiety symptoms in pregnancy.

A subsequent study[67] compared the screening performance of the HADS-A with diagnosis of an anxiety disorder according to DSM criteria. The authors found that high anxiety scores on the HADS-A, defined as the top 15% of scores, had poor concordance (34%) with formal diagnosis of an anxiety disorder. The poor concordance with DSM diagnoses seems to confirm the previous findings indicating that the seven-item HADS anxiety subscale as a whole is not a reliable screening tool to assess anxiety in pregnancy. However, based on the evidence provided by the two studies discussed above on the factor structure of the HADS, we conclude that the three identified items represent a shortened version of the HADS-A which, unlike the entire HADS-A, has good evidence of its psychometric properties to measure antenatal anxiety.

State-Trait Anxiety Inventory (STAI)

The STAI[85] comprises two subscales, each composed of 20 items. It is based on a model of anxiety that distinguishes between state and trait anxiety.[86] State anxiety refers to the situation-specific, transient component of anxiety. Conversely, trait anxiety reflects a relatively stable personality trait, a dispositional anxiety proneness.[58] Response options range from one (not at all) to four (very much so) for both the state and trait form, and each scale includes ten anxiety-present (for example 'I am worried') and ten anxiety-absent (for example 'I feel secure') items. The state form asks participants about feelings at the present time, whereas the trait form

enquires about how a respondent generally feels. The STAI has been widely validated in the general population[11] and is one of the most common measures used in research to assess anxiety in perinatal women.[35]

This review located four studies reporting psychometric properties of the STAI in pregnant populations, one of which[72] was scored poor in relation to its methodological quality. Both the state and trait form of the STAI were used in an Australian study by Grant and colleagues[8] on women in the third trimester of pregnancy. Internal consistency was found to be high for the full version of the scale, with a Cronbach's alpha (α) of 0.95. A structured diagnostic interview was also used (Mini International Neuropsychiatric Interview)[87] to identify women meeting DSM-IV diagnostic criteria for an anxiety disorder. The authors found a cut-off score of 40 to yield the highest accuracy in identifying women with a diagnosed anxiety disorder, with a sensitivity of 80.9% and a specificity of 79.7%.[8] However, they also acknowledged the limited generalisability of the findings because of the relatively small number of participants. The study did not provide any psychometric data at the item level and it was thus not possible to reach conclusions on the psychometric qualities of individual items measuring specific symptoms.

A further study[66] tested various shortened versions of the STAI-S form to determine the smallest subset of items that preserved high correlations ($r > 0.90$) with the original, 20-item STAI-S. They found that a six-item version produced scores comparable with the full version ($r > 0.94$) while retaining a good level of internal consistency (α = 0.82). The six items selected were the ones with the highest correlations with the remaining 19 items of the STAI-S (i.e. corrected item-total correlations). Specifically, the authors identified three anxiety-present and three anxiety-absent items, corresponding to the following emotional states: calm, tense, upset, relaxed, content and worried. This is a significant finding, as it identifies a number of symptoms (i.e. feeling tense, upset or worried) that correlate highly with the 20-item STAI-S total score, providing an initial indication that these anxiety-present symptoms may be considered relatively accurate indicators of problematic anxiety in pregnancy.

This was confirmed in a further study by Bayrampour and colleagues[51] that examined the psychometric properties of three six-item shortened versions of the STAI-S when compared with the full state form. The three short versions are the ones discussed above[66] and two other versions developed in non-perinatal populations. The six-item version by Marteau & Bekker[66] had the highest correlation with the sum score of the full form ($r = 0.94$). Furthermore, confirmatory factory analysis was conducted and the version by Marteau & Bekker[66] was found once more to consistently have the best values for all fit indexes considered, with the three anxiety-present items (i.e. feeling tense/upset/worried) all found to have coefficient item loadings above 0.63, a further indication of their psychometric soundness.

In sum, the three items from the STAI-S short form discussed above were identified in two studies of good methodological quality[51,66] as potentially reliable indicators of anxiety symptoms during pregnancy.

GAD-7

The GAD-7[18] was developed in 2006 as a brief screening measure for generalised anxiety disorder. Its original psychometric validation study, in a large number of primary care patients indicated very good screening accuracy in identifying people with a diagnosis of generalised anxiety disorder.[18] The scale consists of seven items asking respondents about some of the core generalised anxiety disorder symptoms (for example excessive or persistent worry, trouble relaxing) experienced in the previous 2 weeks. As previously

discussed, the first two questions of the GAD-7 (GAD-2) have been recently recommended by NICE as a brief screening measure for anxiety in perinatal women.[6]

Only two studies examining the measurement properties of the GAD-7 in a pregnant population were identified by this review,[70,74] and only one[74] achieved a satisfactory rating for its methodological quality. Importantly, this was one of the few included studies that performed assessment of a scale against a gold-standard clinical interview, the Composite International Diagnostic Interview,[88] to determine the criterion validity of the scale. In this antenatal sample at a cut-off score of seven or above, notably different from the cut-off of ten identified in the general population, the measure yielded moderately good sensitivity (73%) and specificity (67%).[74] Internal consistency was close to excellent (α = 0.89).

Both EFA and CFA were conducted, and confirmed the unidimensional structure (i.e. a single factor) of the GAD-7 previously found in the general population.[18] The results of the factor analysis indicated that the seven items loaded on a single factor with item loadings all exceeding 0.63. In order to identify which items provided the most accurate screening performance we thus examined the item discrimination parameters, which are based on item-response theory and indicate how well individual items differentiate between different levels of the target condition among respondents.[89] Two items showed considerably higher discrimination parameter estimates than the remaining ones. These were item 3 'Worrying too much about different things' (2.05) and item 2 'Not being able to stop or control worrying' (2.04), which clearly tap into the experience of pervasive or persistent worry typical of generalised anxiety disorder. All other items exhibited substantially lower discrimination parameter estimates. Considering that this study was of excellent methodological quality, the two identified items have consequently strong evidence of their psychometric value in the antenatal period.

Brief Measure of Worry Severity (BMWS)

A single study[11] was located reporting psychometric data of the BMWS[90] in pregnant women. Self-report scales assessing the construct of 'worry' were included in this review (Appendix 1) as worry is a core clinical feature of generalised anxiety disorder.[22,91] A number of studies indicate that generalised anxiety disorder is the most common anxiety disorder in pregnancy[1,72] and for this reason worry can be hypothesised to be an important dimension of the construct of antenatal anxiety. The BMWS was developed as a unidimensional measure of the functional impact and severity of worry.[90] It includes eight items assessing different aspects of worry. Respondents are asked to rate their general or usual experience of worrying, with four verbally anchored response options (not true at all – definitely true).[90]

Austin et al aimed to determine whether the construct of worry as measured by the BMWS, defined as 'dysfunctional trait cognitive anxiety', was a significant predictor of postnatal depression.[11] Internal consistency was very good (α = 0.89) and the BMWS also showed good convergent validity with the STAI trait ($r = 0.71$). Although psychometric properties of the scale at the item level were not reported, there was evidence that the construct of worry as measured by the BMWS is a reliable indicator of antenatal anxiety. First, the BMWS was found to have good construct validity in these pregnant participants, as it showed significant correlations with a number of other variables linked to a current episode of anxiety and depression.[11] Moreover, it was a better predictor of postnatal depression than the STAI-S after controlling for possible confounding factors. As the literature indicates that antenatal anxiety is a predictor of postnatal depression,[7,8,10] it appears than

the BMWS taps into a core component of antenatal anxiety considering its good predictive validity.

Consequently, the construct of worry has strong evidence of being psychometrically robust according to the criteria used in this review (i.e. consistent findings in multiple studies of good or excellent methodological quality) as it was also identified as psychometrically sound in other studies previously discussed in this synthesis.

Cambridge Worry Scale (CWS)

The CWS is a 16-item measure assessing the extent and content of women's worries during pregnancy.[58] The 16 items in the CWS enquire both about worries specific to pregnancy, such as 'The possibility of miscarriage', 'The possibility of something being wrong with the baby' or 'Giving birth', and more general concerns including 'Money problems' and 'Your relationship with your family and friends'. Items are scored on a six-point Likert-type scale with verbally described anchors ranging from zero (not a worry) to five (major worry) and referring to the present time.[58]

Six studies examining psychometric aspects of the CWS in a pregnant population were included in this review, four of which are considered here. The other two studies were rated as poor[68] or fair[62] for their methodological quality.

Green and colleagues[58] were the first to investigate the structural validity (i.e. the factor structure) of the CWS. A longitudinal design was used in a large number ($n = 1207$) of British women completing the CWS at gestational weeks 16, 22 and 35. The authors analysed scores at these three time points by means of principal component analysis (PCA), a form of exploratory factor analysis. The PCA revealed a four-factor structure, consisting of the following factors: (a) socio-medical aspects of having a baby, (b) socio-economic issues, (c) health of mother and baby, and (d) relationships with partner, family and friends. This four-factor solution was subsequently replicated in all the other studies examined in this synthesis.[53,57,69] This can be considered robust evidence of factorial stability of the CWS in different populations and stages of pregnancy.

The convergent validity of the CWS was examined by comparing it with STAI state and trait scores[57,58,69] and with the anxiety subscale of the Symptom Checklist-90[92] by Carmona Monge and colleagues.[53] Two of the four CWS subscales were found to have the highest correlations with state anxiety (STAI-S) scores across studies. These were the 'socio-medical' and the 'health of mother and baby' factors. For the purpose of this review, we specifically focused on these two factors, both because of their higher correlations with state anxiety and because the content of items in these subscales appears to reflect worries more closely related to pregnancy. Thus, an examination of individual item loadings for these two factors was carried out.

In relation to the 'socio-medical' subscale, one item ('Giving birth') was found to load above the predefined criterion of 0.63 in all studies, thus demonstrating strong evidence of its psychometric properties in assessing a major worry in pregnancy. Another three items showed moderate strength of evidence as they loaded above 0.63 on the 'socio-medical' subscale in all studies apart from one. Specifically, 'Internal examinations' had an item-type loading coefficient of 0.61 in Gourounti and colleagues,[57] but item loadings above 0.63 in all the other studies; 'Going to hospital' (0.68–0.79), apart from Gourounti and colleagues[57] (0.47); and 'Coping with the new baby' (0.65–0.68), except for the study by Petersen and colleagues,[69] in which its loading was 0.58.

An inspection of the second factor examined, 'Health of mother and baby', indicated two further items with loadings >0.63 in all the studies, namely 'The possibility of miscarriage', which ranged between 0.75[58] and 0.85[53], and 'The possibility of something being wrong with the baby' (range 0.65–0.83[53,58]). The other two items included in this subscale, 'Own health' and 'Health of someone else close', consistently loaded below the pre-defined cut-off.

In summary, three items of the CWS ('Giving birth', 'The possibility of miscarriage', 'The possibility of something being wrong with the baby') demonstrated strong evidence of their psychometric properties. Three further items ('Internal examinations', 'Going to hospital' 'Coping with the new baby') showed a moderate strength of evidence of their psychometric value in pregnancy.

Wijma Delivery Expectancy/Experience Questionnaire (W-DEQ – Version A)

The W-DEQ[73] was developed in the late nineties to assess the construct of fear of childbirth. Within the research literature on pregnancy-specific anxiety, fear of childbirth or tokophobia has emerged as a central dimension of pregnancy-specific anxiety.[5,34,93] The W-DEQ Version A[73] includes 33 items enquiring about thoughts and feelings relating to the approaching childbirth, with six response options ranging from 'not at all' to 'extremely'.

Five studies included in the present review reported psychometric information on the W-DEQ in an antenatal population,[55,56,59,60,73] and three studies achieved a good or excellent methodological quality rating.[55,60,73] In the original development study of the W-DEQ,[73] internal consistency of the measure was excellent ($\alpha = 0.93$). The authors also provided good evidence of the face and construct validity of the W-DEQ, with all items formulated based on the clinical experience of the first two authors and incorporating women's input in the wording of items. The W-DEQ showed higher correlations with other anxiety measures than with extraversion or depression measures. However, these correlations were only moderate (STAI-T: $r = 0.54$; S-R Inventory of anxiousness: $r = 0.52$), thus showing a degree of conceptual overlap but also a sufficient level of variance left to indicate that the W-DEQ measures other than anxiety as a dispositional trait.[73]

At the item level, item-total correlations were ranked and the authors examined the ten items with the highest ranking. Two domains of fear of childbirth, 'Negative feelings towards childbirth' and 'Fear of labour and delivery', were identified among the items more strongly correlated with the sum score, thus suggesting a stronger relation with the overall construct of fear of childbirth. As single items composing the W-DEQ are very specific to a given feeling or cognitive appraisal, we considered it appropriate to focus on domains of fear of childbirth rather than individual items.

Two other studies[55,60] included in this synthesis conducted factor analysis of W-DEQ scores and found four distinct dimensions of the construct of fear of childbirth as measured by the scale. Johnson & Slade[60] named the four identified domains Fear, Lack of positive anticipation, Isolation and Riskiness. The latter two refer to feelings of isolation related to childbirth and to the extent to which women anticipate risks for the child during delivery. Fenaroli & Saita[55] also found a four-factor structure of the W-DEQ, and although the four domains were named with slightly different labels than those used by Johnson & Slade[60], the four factors were considerably similar and had a high degree of conceptual overlap. In this best-evidence synthesis two dimensions of pregnancy-specific anxiety, namely Fear of labour and delivery and Negative feelings towards childbirth (corresponding to Lack of positive anticipation in Fenaroli & Saita[55]), were thus found to exhibit strong evidence of being psychometrically sound in assessing this specific aspect of antenatal anxiety. A third dimension (Fear for baby's health) showed moderate strength of evidence as, although it was identified in two studies,[55,60] contrasting results were found in another study.[73]

PRAQ-R and PRAQ-R2

This pregnancy-specific anxiety measure is composed of ten items assessing various manifestations of anxiety related to a current pregnancy. Each item asks about feelings at the present time and has five response options ranging from 'never' to 'very often'. Its original version (PRAQ)[94] consisted of 58 items and was developed based on previous anxiety measures.

The first study testing the psychometric properties of the PRAQ was carried out by Huizink and colleagues[28] who initially tested a revised, 34-item version (PRAQ-R[28]) of the original PRAQ on 230 nulliparous women. The authors' aim was to examine the factorial structure of the PRAQ-R and test the hypothesis that pregnancy-specific anxiety could be differentiated from general anxiety by comparing STAI and PRAQ-R scores. They found that only between 8 and 27% of the PRAQ-R variance was accounted for by the index of general anxiety at different time points during pregnancy, with no linear association found between the two measures. This was interpreted as evidence of the distinctiveness of the pregnancy-specific anxiety construct[28] and highlighted once more that measures of general anxiety cannot be accurately used to identify women experiencing fears and worries specific to pregnancy.

The authors initially conducted EFA and removed a number of items because of high error variance, resulting in a final version comprising ten items (PRAQ-R). A subsequent CFA revealed that a solution with three factors provided the best fit to the data, with the three identified factors labelled by the researchers 'Fear of giving birth' (three items), 'Fear of bearing a physically or mentally handicapped child' (four items) and 'Concern about one's appearance' (three items). All individual items loaded on one of the factors above the cut-off of 0.63, except for one item (0.50), 'I am worried about not being able to control myself during labour and fear that I will scream'. Similarly to the approach used for the W-DEQ and discussed above, we considered the whole factors rather than individual items making up a given factor.

Two further studies[31,33] included here tested the measurement properties of the PRAQ-R, and both replicated the previous finding of a three-factor structure of the PRAQ-R by means of CFA. As the original participants of the ten-item PRAQ-R were exclusively composed of nulliparous women, Westerneng and colleagues[33] aimed to test the factorial stability of the three-factor solution of the PRAQ-R[28] on a large (n > 6000) data-set of both nulliparous and parous women. This involved the deletion of item 8 'I am anxious about the delivery because I have never experienced one before', as it was not suitable for use with women who had already experienced childbirth. CFA confirmed the same three-factor structure of the original ten-item PRAQ-R with good indexes of fit to the data for both nulliparous and parous women.

Three factors were also found in a recent study[31] that replaced item eight of the original PRAQ-R with the more generic 'I am anxious about the delivery' in order to preserve a ten-item scale while making it appropriate for all pregnant women irrespective of parity (PRAQ-R2)[31]. All item loadings were once more above 0.63 (range: 0.70–0.93) except for two items, 'I am worried about not being able to control myself during labour and fear that I will scream', similarly to Huizink and colleagues,[28] and 'I sometimes think that our child will be in poor health or will be prone to illnesses'.

In summary, across the three studies examined here[28,31,33] eight items from the PRAQ-R were found to consistently have high loadings on one of three factors (i.e. pregnancy-specific anxiety domains). These three pregnancy-specific anxiety domains, namely 'Fear of giving birth', 'Fear of bearing a physically or mentally handicapped child' and 'Concern about one's appearance', were all identified in studies of good or excellent methodological quality, thus providing strong evidence of being accurate indicators of pregnancy-specific anxiety.

There are several important findings to this study. First, this review has identified a number of anxiety items and domains from existing self-report scales with demonstrated psychometric value when used to assess symptoms of anxiety in pregnant women. To the best of our knowledge, this is the first study to analyse the content of self-report anxiety measures used in the antenatal period and provide recommendations for the accurate assessment of maternal antenatal anxiety based on a systematic synthesis of published psychometric data.

A second, significant finding of this paper is that it highlights the scarcity of studies reporting psychometric properties of scales employed to measure anxiety in pregnancy. A considerable number of studies using self-report scales to assess antenatal anxiety were not included in this review as no measurement properties of the scale used were reported. It would appear that in most cases researchers have selected a given anxiety measure only based on its widespread use and good psychometric properties in the general population.[36] However, assuming that the measurement properties of a psychological scale developed for the general population are preserved in pregnancy is incorrect for various reasons discussed earlier in this paper (i.e. undue emphasis on somatic symptoms, lack of validated cut-off scores and norms for pregnant populations, role of pregnancy-specific anxiety).

A further limitation of the literature is that only a dearth of studies located by this review (n = 5)[8,67,70,72,74] validated a measure against a reference 'gold' standard such as a structured diagnostic interview. Testing a scale against a reference standard provides evidence of the screening accuracy of a measure, also referred to as its criterion validity, arguably the single most important aspect of psychometric validation of a scale.[48]

Perhaps even more surprisingly, only two studies[70,74] were identified that examined the psychometric properties of the GAD-7 in a pregnant population, and only one[74] was found to have satisfactory methodological quality. As previously reported, the GAD-2 (i.e. the initial two questions of the GAD-7) is the measure currently recommended by NICE in the UK to screen for anxiety in pregnant women, followed by administration of the GAD-7 if a woman scores three or higher on the GAD-2.[6] The only methodologically robust study providing psychometric information on the GAD-7 in a pregnant population[74] was also somewhat limited by focusing exclusively on the screening accuracy of the GAD-7 for generalised anxiety disorder, without providing any evidence of its screening ability for other anxiety disorders in pregnancy. Furthermore, sub-analyses to assess the screening ability of the GAD-2 as opposed to the full GAD-7 were not conducted, thus leaving unanswered the question of whether the GAD-2 can be used as an ultra-brief screening scale for problematic anxiety symptoms in pregnancy, as per recent guidelines.[6]

Key best-evidence findings

Eight self-report measures were considered in the synthesis of the best available evidence presented above. One further scale located by this review (Pregnancy Anxiety Scale[65]) was not examined at the best-evidence stage as the single study reporting its psychometric properties was rated poor for its methodological quality.[65]

The key findings regarding anxiety items and domains identified as accurate indicators of antenatal anxiety, as discussed in the Results, are summarised here. A complete list of all the identified anxiety items and domains is also presented in the supplementary Table 2. Furthermore, a table summarising all the correlations between scales included in the review is available in supplementary Table 3.

Items assessing excessive, generalised worry were found to be psychometrically sound in the antenatal period in the EPDS, HADS-A, BMWS, GAD-7 and STAI-S. Overall, there was strong evidence of the psychometric robustness of items measuring the domain of worry, with consistent findings in multiple studies of good or excellent quality. Since excessive worry is essentially a cognitive symptom, it could be argued that it is less susceptible to the physical and physiological changes of pregnancy, and it remains thus a good indicator of problematic anxiety in pregnancy as it is in the general population.

A second anxiety domain that showed good evidence of its psychometric soundness in pregnant populations concerned items tapping into symptoms of fear or panic. Feelings of fear are another important component of different anxiety disorders, including panic disorder, agoraphobia, social anxiety disorder and specific phobia.[22,95] In this review, items assessing the fear/panic domain were identified as psychometrically sound for use in pregnancy in various scales, including the HADS-A, the EPDS and several pregnancy-specific anxiety scales.

Other specific symptoms identified by this review showed moderate evidence of their screening ability in the assessment of antenatal anxiety. These included being excessively self-critical (EPDS, item 3), feeling upset (STAI-S, item 6) and the experience of nervous or motor tension (STAI-S, item 3). Although these symptoms may not appear to be specific to anxiety disorders, these findings are in line with the well-established tripartite model of anxiety and depression. This model postulates that depressive and anxiety disorders share a common component of general emotional distress, and the symptoms above can be categorised as manifestations of general distress, which can be present in both depressive and anxiety symptomatology.[96]

In relation to anxiety symptoms specifically related to pregnancy, fear of childbirth was shown to be a good indicator of pregnancy-specific anxiety. Specifically, pregnancy-specific anxiety symptoms of fear related to giving birth exhibited strong evidence of their psychometric value in the W-DEQ (several items) and the PRAQ-R (two items related to fear of childbirth).

Items assessing persistent worries specifically related to pregnancy also showed good psychometric properties in the CWS, the W-DEQ and the PRAQ-R. The worries with the strongest evidence to support their screening accuracy related to concerns regarding the health or safety of the baby and the possibility of miscarriage. Other worries, including being in hospital and worrying about future parenting showed only moderate evidence of their screening value (see supplementary Table 2). It may be argued that most women are likely to experience some degree of concern regarding these aspects of pregnancy, but that in women experiencing clinical levels of anxiety these worries may be more intense or persistent (i.e. higher severity or frequency).

Strengths and limitations

The present review has a number of strengths. Only studies with good or excellent methodological quality as determined by the COSMIN checklist[45] were included in the best-evidence synthesis, thus guaranteeing that the conclusions reached were only based on the strongest evidence available. We also used a comprehensive search strategy that was devised to locate studies testing the psychometric properties of both general anxiety scales and pregnancy-specific anxiety measures, unlike previous reviews that were focused mostly or exclusively on general anxiety or pregnancy-specific anxiety scales.[35,36] A second reviewer independently checked a sample of studies, both in the initial phase of screening of titles and abstract and for the quality assessment of included studies, as per best practice recommendations for systematic reviews.[37] The review was reported according to the PRISMA reporting guidelines[39] (see supplementary Table 4).

Several limitations also have to be acknowledged. Searches were limited to research articles in English and restricted to publications from 1991 onwards, this being the year when the first pregnancy-specific anxiety scale was developed. The generalisability of the review findings may also be somewhat limited by the fact that we did not include studies from countries with substantial cultural differences compared with the UK (i.e. Asian and African countries) for which cultural equivalence of psychological symptoms cannot be assumed.[97,98]

Implications and future directions

The accurate identification of women experiencing high levels of anxiety symptoms in pregnancy is important and deserves clinical attention for several reasons. Whereas postnatal depression has been the focus of most research in perinatal mental health in the past decades,[21,24] there is now a substantial body of research indicating that anxiety in pregnant women is common and is associated with increased risk for negative maternal and child outcomes.[3,32,99] In the UK, the Royal College of General Practitioners has identified perinatal mental health as a clinical priority[100] and a recent report from the London School of Economics has estimated the costs of neglecting perinatal mental health problems in the UK to be a striking figure of £8.1 billion for every annual cohort of women, with approximately three-quarters of this cost related to the adverse long-term impact on children.[101]

Among the range of perinatal mental health problems that women can experience, anxiety disorders have the highest prevalence.[1] Consequently, a number of authors in recent years have advocated the use of a brief scale for the universal screening of antenatal anxiety.[36,102] To the best of our knowledge, no anxiety scales have been developed that are specific to the antenatal period and take into account both general and pregnancy-specific anxiety symptoms. Most studies have used measures of general anxiety, but the clinical importance of including screening for pregnancy-specific anxiety symptoms is supported by studies indicating that pregnancy-specific anxiety may be a better predictor of adverse birth and child development outcomes than general anxiety during pregnancy.[34,103]

Future research is needed to conduct robust psychometric studies of existing measures in sufficiently large samples and ideally including validation against a reference standard. The development of a new anxiety scale specifically constructed for use in pregnancy and that takes into account both general anxiety and symptoms of pregnancy-specific anxiety would also be highly desirable.

In sum, despite the research literature on prevalence, risk factors and treatment of antenatal anxiety having decisively grown in recent years,[1,8,104] this review clearly points out how evidence regarding the screening performance of anxiety scales for use in pregnancy, including the one currently recommended by NICE, remains insufficient. The lack of measures with a sufficient evidence base constitutes a considerable barrier to the identification of pregnant women experiencing problematic anxiety symptoms, the initial step if they are to be offered the appropriate support or treatment. This is, in turn, an important missed opportunity for early prevention of negative health outcomes for women and their children. This review improves the current understanding of anxiety symptomatology

in pregnant women and may contribute to provide the theoretical basis for the development of a psychometrically robust screening scale for antenatal anxiety.

Andrea Sinesi, MSc, PhD Student, Nursing, Midwifery and Allied Health Professions Research Unit, University of Stirling, UK; **Margaret Maxwell**, PhD, Professor of Health Services and Mental Health Research and Director, Nursing, Midwifery and Allied Health Professions Research Unit, University of Stirling, UK; **Ronan O'Carroll**, MPhil, PhD, Professor of Psychology, Department of Psychology, University of Stirling UK; **Helen Cheyne**, RGN, RM, PhD, Professor of Maternal and Child Health and Deputy Director, Nursing, Midwifery and Allied Health Professions Research Unit, University of Stirling, UK

Correspondence: Andrea Sinesi, Nursing, Midwifery and Allied Health Professions Research Unit (NMAHP RU), Unit 13 Scion House, Stirling University Innovation Park, University of Stirling, Stirling FK9 4NK, UK. Email: andrea.sinesi@stir.ac.uk

First received 2 Aug 2018, final revision 2 Nov 2018, accepted 9 Nov 2018

Appendix

Inclusion and exclusion criteria

	Inclusion criteria	Exclusion criteria
Type of study	Primary research articles reporting at least one psychometric property of a self-report rating scale used to assess anxiety symptoms in pregnancy. Published in a peer-reviewed journal in English in or after 1991.	Studies conducted in countries with substantial cultural differences with the UK (i.e. African and Asian countries) for which cultural equivalence cannot be assumed. Qualitative studies on the experience of anxiety symptoms during pregnancy.
Population	Pregnant or perinatal participants (for perinatal samples, subgroup analyses of psychometric properties of the measure available for the subsample of pregnant women).	Sample composed exclusively of women with high-risk pregnancies, because of obstetric complications (for example pre-eclampsia, ectopic pregnancy) or high psychosocial risk.
Construct of interest	A specific anxiety disorder, as determined by DSM-5 or ICD-10 diagnostic criteria. General 'anxiety' or 'worry' Pregnancy-related anxiety, as defined by Huizink and colleagues.[28] Fear of childbirth, as this is deemed to be a relevant component of pregnancy-specific anxiety.	Any other construct, as for example general mental health, mental disorders other than anxiety during pregnancy, stress or postnatal anxiety.
Type of measurement	All studies with published psychometric data using self-report rating scales to assess anxiety symptoms in pregnancy and meeting the inclusion criteria for type of study, population, and construct of interest detailed above were eligible for inclusion in the review.	Anxiety symptoms assessed exclusively with methods other than a self-report rating scale (for example open questions by a health professional, physiological measures of anxiety). Scale designed to be completed by a health professional after observation (i.e. not self-report). Scale not developed to generate a total score or single-item measures.

References

1 Dennis CL, Falah-Hassani K, Shiri R. Prevalence of antenatal and postnatal anxiety: systematic review and meta-analysis. *Br J Psychiatry* 2017; **210**: 315–23.

2 Heron J, O'Connor TG, Evans J, Golding J, Glover V. The course of anxiety and depression through pregnancy and the postpartum in a community sample. *J Affect Disord* 2004; **80**: 65–73.

3 Dunkel-Schetter C, Lobel M. Pregnancy and birth: a multilevel analysis of stress and birth weight. In *Handbook of Health Psychology 2* (eds T Revenson, A Baum, J Singer): 427–53. Lawrence Erlbaum, 2011.

4 Stein A, Pearson RM, Goodman SH, Rapa E, Rahman A, McCallum M, et al. Effects of perinatal mental disorders on the fetus and child. *Lancet* 2014; **384**: 1800–19.

5 Rubertsson C, Hellström J, Cross M, Sydsjö G. Anxiety in early pregnancy: prevalence and contributing factors. *Arch Womens Ment Health* 2014; **17**: 221–8.

6 National Institute for Health and Care Excellence. *Antenatal and Postnatal Mental Health: Clinical Management and Service Guidance (CG 192)*. NICE, 2014.

7 Milgrom J, Gemmill A, Bilszta JL, Hayes B, Barnett B, Brooks J, et al. Antenatal risk factors for postnatal depression: a large prospective study. *J Affect Disord* 2008; **108**: 147–57.

8 Grant K, McMahon C, Austin MP. Maternal anxiety during the transition to parenthood: a prospective study. *J Affect Disord* 2008; **108**: 101–11.

9 Verreault N, DaCosta D, Marchand A, Ireland K, Dritsa M, Khalife S. Rates and risk factors associated with depressive symptoms during pregnancy and with postpartum onset. *J Psychosom Obstet Gynaecol* 2014; **35**: 84–91.

10 Sutter-Dallay AL, Giaconne-Marcesche V, Glatigny-Dallay E, Verdoux H. Women with anxiety disorders during pregnancy are at increased risk of intense postnatal depressive symptoms: a prospective survey of the MATQUID cohort. *Eur Psychiatr* 2004; **19**: 459–63.

11 Austin MP, Tully L, Parker G. Examining the relationship between antenatal anxiety and postnatal depression. *J Affect Disord* 2007; **101**: 169–74.

12 Field T, Diego M, Hernandez-Reif M, Figueiredo B, Deeds O, Ascencio A, et al. Comorbid depression and anxiety effects on pregnancy and neonatal outcome. *Infant Behav Dev* 2010; **33**: 23–9.

13 Diego MA, Jones NA, Field T, Hernandez-Reif M, Schanberg S, Kuhn C, et al. Maternal psychological distress, prenatal cortisol, and fetal weight. *Psychosom Med* 2006; **68**: 747–53.

14 Berle JO, Mykletun A, Daltveit AK, Rasmussen S, Holsten F, Dahl AA. Neonatal outcomes in offspring of women with anxiety and depression during pregnancy: a linkage study from The Nord-Trøndelag Health Study (HUNT) and Medical Birth Registry of Norway. *Arch Womens Ment Health* 2005; **8**: 181–9.

15 Ding XX, Wu YL, Xu SJ, Zhu RP, Jia XM, Zhang SF, et al. Maternal anxiety during pregnancy and adverse birth outcomes: a systematic review and meta-analysis of prospective cohort studies. *J Affect Disord* 2014; **159**: 103–10.

16 Talge NM, Neal C, Glover V. Antenatal maternal stress and long-term effects on child neurodevelopment: how and why? *J Child Psychol Psychiatry* 2007; **48**: 245–61.

17 Van Der Bergh BR, Mulder EJ, Mennes M, Glover V. Antenatal maternal anxiety and stress and the neurobehavioural development of the fetus and child: links and possible mechanisms. A review. *Neurosci Biobehav Rev* 2005; **29**: 237–58.

18 Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006; **166**: 1092–7.