Australian Perinatal Mental Health Guideline Evidence Review

Technical Report Part B

Psychosocial assessment and screening for depression or anxiety

Prepared by

hereco
clinical evidence for the real world

June 2017

TABLE OF CONTENTS

B1 INTRODUCTION 1

B2 CLINICAL QUESTIONS 3

B2.1 Psychosocial assessment 3

B2.2 Depression screening 4

B2.3 Anxiety screening 5

B3 SEARCH METHODS 6

B3.1 Overall approach to searches 6

B3.1.1 Screening 7

B3.1.2 Psychosocial assessment 10

B3.1.3 Acceptability, Effectiveness and Implementation 12

B3.1.4 Cost-effectiveness of perinatal mental health screening 12

B4 PSYCHOSOCIAL ASSESSMENT 13

B4.1 Relevant outcomes and critical appraisal methods 13

B4.2 Characteristics of studies of technical performance 13

B4.3 Evidence profile tables 18

B4.4 Non-technical characteristics of relevant tools 20

B4.5 Clinical usefulness of relevant tools 20

B4.5.1 ALPHA 20

B4.5.2 ANRQ 21

B4.5.3 PRQ 23

B4.6 General evidence of clinical usefulness 23

B4.7 Overall summary of findings 25

B5 SCREENING FOR DEPRESSION 27

B5.1 Relevant outcomes of technical performance 27

B5.2 Critical appraisal of technical performance 28

B5.2.1 Quality assessment of individual studies 28

B5.2.2 Overall certainty of the evidence by outcome 29

B5.3 Evidence of technical performance 29

B5.3.1 Characteristics of studies in foundation review 29

B5.3.2 Evidence summaries based on studies in foundation review 34

B5.3.3 Summary of findings regarding technical performance 41

B5.4 Non-technical characteristics of relevant tools 50

B5.5 Clinical usefulness of relevant tools 50

B5.5.1 Acceptability of depression screening 50

B5.5.2 Effectiveness of depression screening 51

B5.5.3 Implementation of depression screening 52

B5.6 Cost-effectiveness of perinatal depression screening 52

B5.7 Overall summary of findings 53

B6 SCREENING FOR ANXIETY 55

B6.1 Relevant outcomes of technical performance 55

B6.2 Critical appraisal of technical performance 55

B6.3 Evidence of technical performance 55

B6.3.1 Characteristics of individual studies included for anxiety screening 55

B6.3.2 Evidence summaries from included studies 58

B6.3.3 Summary of findings regarding technical performance 64

B6.4 Non-technical characteristics of relevant included tools 68

B6.5 Clinical usefulness of relevant tools 68

B6.6 Overall summary of findings 68

B7 REFERENCES 70

B8 APPENDICES 72

B8.1 Details of searches for psychosocial assessment and screening tools 72

B8.1.1 Search strings 72

B8.1.2 Exclusion of studies 77

B8.1.3 Excluded studies list 78

B8.2 Comparison of individual studies within relevant SRs of screening 78

B8.2.1 Depression screening SRs 78

B8.3 Lists of included studies from supplementary searches 82

B8.3.1 Anxiety screening 82

B8.3.2 Psychosocial assessment 82

B8.4 Details of QUADAS-2 assessments 82

B8.4.1 Questions for quality assessment of diagnostic studies 82

B8.4.2 Checklists for additional screening studies 83

TABLE OF TABLES

Table B3-1 Criteria for determining study eligibility by NICE 2015 for depression screening 8
Table B3-2 Criteria for determining study eligibility for anxiety screening 10
Table B3-3 Criteria for determining study eligibility: psychosocial assessment – technical characteristics 12
Table B4-1 Design and psychometric properties of individual studies included in Johnson 2012 15
Table B4-2 Design and psychometric properties of individual included studies published since Johnson 2012 18
Table B4-3 Evidence regarding the technical performance of the included psychometric instruments 18
Table B4-4 Non-technical characteristics of the relevant included tools 20
Table B4-5 Evidence from the supplementary literature search regarding the acceptability of the ALPHA 21
Table B4-6 Evidence from the supplementary literature search regarding the effectiveness of the ALPHA 21
Table B4-7 Evidence from the supplementary literature search related to the implementability of the ALPHA 21
Table B4-8 Evidence from the supplementary literature search regarding the acceptability of the ANRQ 22
Table B4-9 Evidence from the supplementary literature search regarding the effectiveness of the ANRQ 22
Table B4-10 Evidence from the supplementary literature search related to the implementability of the ANRQ 23
Table B4-11 Evidence from the supplementary literature search regarding the effectiveness of the PRQ 23
Table B4-12 Evidence from the supplementary literature search related to the implementability of the PRQ 23
Table B4-13 Overall Summary of Findings related to the use of perinatal psychosocial assessment tools 26
Table B5-1 Consequences of findings from psychometric screening tools administered during the perinatal period 28
Table B5-2 Key characteristics of studies using the EPDS, PHQ, Whooley Questions or K-10 to identify perinatal depression 31
Table B5-3 Evidence summary table for the EPDS for detection of depression in antenatal women 35
Table B5-4 Evidence summary table for the EPDS for detection of depression in postnatal women 36
Table B5-5 Evidence summary table for the PHQ for detection of depression in antenatal women 38
Table B5-6 Evidence summary table for the PHQ for detection of depression in postnatal women 38
Table B5-7 Evidence summary table for the 'Whooley questions' for detection of depression in antenatal women 39
Table B5-8 Evidence summary table for the 'Whooley questions' for detection of depression in postnatal women 39
Table B5-9 Evidence summary table for the K-10 for detection of depression in antenatal women 40
Table B5-10 Evidence summary table for the K-10 for detection of depression in postnatal women 40
Table B5-11 Summary of Findings table for the EPDS and identification of depression in the antenatal period (based on NICE 2015) 42
Table B5-12 Summary of Findings table for the PHQ and identification of depression in the antenatal period (based on NICE 2015) 43
Table B5-13 Summary of Findings table for the 'Whooley questions' and identification of depression in the antenatal period (based on NICE 2015) 43
Table B5-14, Summary of Findings table for the K-10 and identification of depression in the antenatal period (based on NICE 2015) 44
Table B5-15 Summary of Findings table for the EPDS and identification of depression in the postnatal period (based on NICE 2015) 44
Table B5-16 Summary of Findings table for the PHQ and identification of depression in the postnatal period (based on NICE 2015) 47
Table B5-17 Summary of Findings table for the 'Whooley questions' and identification of depression in the postnatal period (based on NICE 2015) 49
Table B5-18 Summary of Findings table for the K-10 and identification of depression in the postnatal period (based on NICE 2015) 49
Table B 5-19 Non-technical characteristics of the relevant included tools for depression screening 50
Table B5-20 Overall Summary of Findings related to the use of perinatal depression screening tools 54
Table B6-1 Key characteristics of studies that used the EPDS (full or 3A), GAD-7, GHQ (12, 28 or 30-item), HADS, K-10 or STAI to identify perinatal anxiety 57
Table B6-2 Evidence summary table for the EPDS for detection of anxiety in antenatal women 59
Table B6-3 Evidence summary table for the EPDS for detection of anxiety in postnatal women 59
Table B6-4 Evidence summary table for the EPDS for detection of anxiety in a mixed population of antenatal and postnatal women 59
Table B6-5 Evidence summary table for the GAD-7 for detection of anxiety in a mixed population of antenatal and postnatal women 60
Table B6-6 Evidence summary table for the GHQ for detection of anxiety in antenatal women 60
Table B6-7 Evidence summary table for the GHQ for detection of anxiety in postnatal women 61
Table B6-8 Evidence summary table for the HADS for detection of anxiety in antenatal women 62
Table B6-9 Evidence summary table for the K-10 for detection of anxiety in antenatal women 62
Table B6-10 Evidence summary table for the STAI for detection of anxiety in antenatal women 63

Table B6-11 Summary of Findings table for the K-10 for the detection of anxiety in antenatal women 65
Table B6-12 Summary of Findings table for the STAI for detection of trait anxiety in antenatal women 65
Table B6-13 Summary of Findings table for the full EPDS for detection of anxiety in a mixed population of antenatal and postnatal women 66
Table B6-14 Summary of Findings table for the EPDS-3A for detection of anxiety in a mixed population of antenatal and postnatal women 67
Table B6-15 Summary of Findings table for the GAD-7 for detection of anxiety in a mixed population of antenatal and postnatal women 67
Table B6-16 Non-technical characteristics of the relevant included tools for anxiety screening 68
Table B6-17 Overall Summary of Findings related to the use of perinatal anxiety screening tools 69
Table B 8-1 Individual included studies in published SRs – depression screening 79
Table B8-2 List of individual studies included for anxiety screening 82
Table B8-3 List of individual studies included for psychosocial assessment 82
Table B8-4 Key aspects of the review question 83
Table B8-5 Risk of bias and applicability judgments – Grant 2008 84
Table B8-6 Risk of bias and applicability judgments – Grigoriadis 2011 86
aTable B8-7 Risk of bias and applicability judgments – Simpson 2014 87
Table B8-8 Risk of bias and applicability judgments – Spies 2009 89
Table B8-9 Risk of bias and applicability judgments – Tran 2011 90

TABLE OF FIGURES

Figure B1-1 Evidence to Decision framework for assessing the evidence related to perinatal psychosocial assessment and screening for depression and anxiety....2
Figure B3-1 Summary of how different published SRs have been used within the current evidence review 6
Figure B5-1 Summary of ROC curve for the EPDS administered in the postnatal period at different cut-off points and diagnoses 46
Figure B5-2 Summary of ROC curve for the PHQ (2-, 8- and 9-item versions) at different timings, diagnoses and cutoffs 48
Figure B8-1 Patient flow diagram – Grant 2008 84
Figure B8-2 Patient flow diagram – Grigoriadis 2011 85
Figure B8-3 Patient flow diagram – Simpson 2014 87
Figure B8-4 Patient flow diagram – Spies 2009 89
Figure B8-5 Patient flow diagram – Tran 2011 90

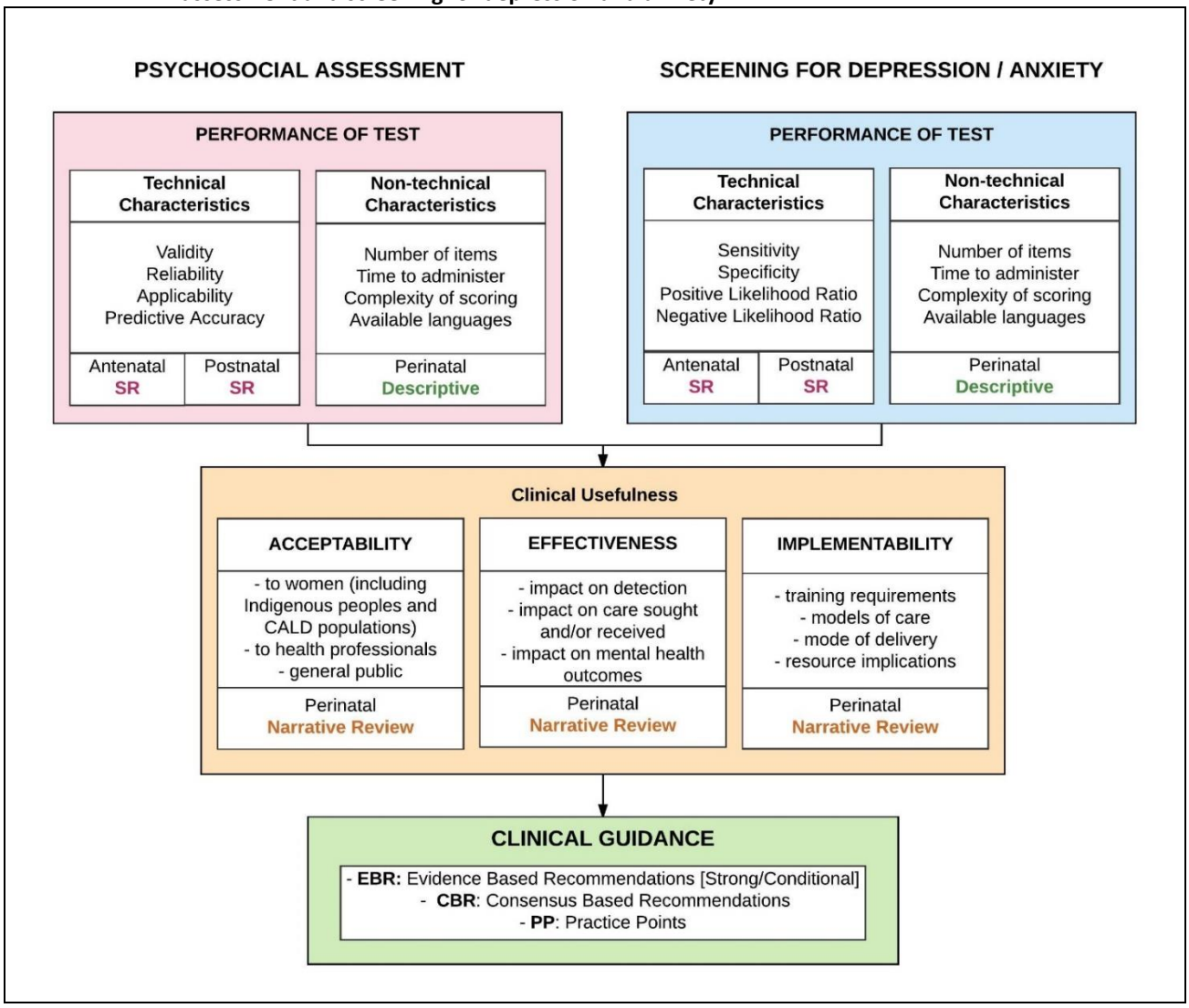
B1 INTRODUCTION

This part of the Technical Report covers evidence and information related to psychosocial assessment and to screening for depression and anxiety. A mixed methods approach has been employed to cover all aspects of care relevant to these two distinct, but closely related topics. The approach includes the use of systematic reviews of quantitative evidence (e.g. screening test performance appraised using QUADAS-2), descriptions of non-technical characteristics of the tests (e.g. time to administer, complexity of scoring), and narrative reviews of the effectiveness, and implementation and acceptability issues associated with perinatal mental health assessment (psychosocial assessment as well as depression/anxiety screening). Where possible, available evidence is presented separately for antenatal versus postnatal populations. Where mixed populations (i.e. pregnant and postpartum women) are included, these are referred to as 'perinatal' populations. The evidence in this Technical Report has then been used to develop Evidence Based Recommendations (EBR), Consensus Based Recommendations (CBR) and Practice Points (PPs), and the rationale for these is included within the Guideline itself.

An Evidence to Decision framework illustrating how the different evidence review methods have been used to inform the development of clinical guidance for psychosocial assessment and depression/anxiety screening is shown in **Figure B1-1**. The sections within this Technical Report follow the format of our Evidence to Decision framework, with evidence presented in the following order: psychosocial assessment; screening for depression; screening for anxiety; acceptability of psychosocial assessment and/or mental health screening; effectiveness of psychosocial assessment and/or mental health screening; implementation of psychosocial assessment and/or mental health screening.

This Technical Report includes an overview of the methods used to identify and appraise the evidence and key findings (presented as Summary of Findings tables, where appropriate). Details of the literature search strategies, critical appraisal methodology used, characteristics of included studies, and Evidence Profile tables (as appropriate) are included in the accompanying Appendices.

Figure B1-1 Evidence to Decision framework for assessing the evidence related to perinatal psychosocial assessment and screening for depression and anxiety



Abbreviations: SR, systematic review.

B2 CLINICAL QUESTIONS

As illustrated in **Figure B1-1** each of the sub-questions has been addressed using methods appropriate to the nature of the question: by systematic, narrative, or descriptive review. All the findings have then been considered by the EWG and their judgment applied to develop appropriate clinical guidance: Evidence-Based Recommendation (EBR; Strong or Conditional), Consensus-Based Recommendation (CBR), or Practice Point (PP). It was agreed *a priori* by the EWG that EBRs could only be derived for a specific tool where there is evidence of the technical performance of that tool.

B2.1 PSYCHOSOCIAL ASSESSMENT

The focus of psychosocial assessment in this evidence review is on validated tools that have been developed to identify a range of factors in a woman's current situation or past that might place her at increased risk of not coping with the pregnancy or newborn, or developing mental health issues. Thus, the clinical focus of the psychosocial assessment questions in this Guideline is the identification of multiple factors known to influence perinatal mental health. Instruments that examine only current mental health are not included here (although instruments for identifying depression and anxiety, such as the Edinburgh Postnatal Depression Scale (EPDS) and the Hospital Anxiety and Depression Scale (HADS), are included in subsequent sections of this Technical Report). For a more detailed discussion of the role of psychosocial assessment within a model of integrated perinatal care the reader is referred to Austin 2015.

Although some tools have been assessed for their ability to predict postnatal depression or anxiety, this is not their sole value. Rather, they provide a structured approach for health professionals to assess risk more broadly and to identify ways in which different kinds of services (not all of them clinical) can be activated to support the woman and her family through pregnancy and after birth. All of the tools included in the current evidence review have been developed to detect factors known to be associated with the onset of perinatal mental health issues. A detailed description of each tool is outside the scope of the current review.

The current review has sought to find evidence of the effectiveness of the included psychosocial assessment tools regarding impact on detection of risk factors, impact on help-seeking behavior (i.e. services sought or utilised), and impact on mental health outcomes. Given the sensitive and personal nature of the questions asked, acceptability to women and non-mental health professionals (such as midwives, child and family health nurses, GPs and obstetricians) is of paramount importance. Particular attention has been given to evidence of acceptability to women of culturally and linguistically diverse backgrounds. The training requirements for non-mental health professionals are also important, as are other implementation issues such as the mode of delivery of the psychosocial assessment tool (e.g. self-report versus healthcare professional administered; hard copy versus electronic copy), and broader models of care.

Based on the concerns described above, specific clinical questions have been asked regarding the effectiveness, acceptability and implementation of psychosocial assessment (see below).

Main question:

1. What is the most appropriate method for psychosocial assessment of women at risk of mental health problems in the perinatal period?

Sub-questions:

1a. What is the performance (defined as reliability, validity and predictive accuracy) of validated multidimensional tools for perinatal psychosocial assessment? [*addressed via systematic review*]

1b. What are the non-technical characteristics (defined as number of items, time to administer, complexity of scoring, and available languages) of validated multidimensional tools for perinatal psychosocial assessment? [*addressed via descriptive review*]

1c. What is the acceptability to pregnant or post-partum women, health professionals, and the general public of validated multidimensional tools for perinatal psychosocial assessment? [*addressed via narrative review*]

1d. What is the effectiveness (defined as impact on detection, care sought or received, and mental health outcomes) of perinatal psychosocial assessment with validated multidimensional tools? [*addressed via narrative review*]

1e. What are the implications (for resourcing, workforce, training requirements and models of care) of implementing perinatal psychosocial assessment (via different modes of delivery) with a validated multidimensional tool? [*addressed via narrative review*]

B2.2 DEPRESSION SCREENING

The focus of depression screening in this evidence review is on validated tools that have been developed or assessed in antenatal and/or postnatal women. As explained above, the current review is limited to instruments that examine current mental health. The specific tools included within scope of this review are detailed in later sections. It should be noted that some tools have been used to screen for depression, some to screen for anxiety, and some to screen for depression and/or anxiety. Consequently, there is some overlap in the questions and evidence included for depression screening and for anxiety screening. In general, the questions addressing technical performance are presented separately for depression and anxiety screening. However, the evidence regarding acceptability, effectiveness and implementation tend to relate to the tools themselves, not the mental health issue. Therefore, the evidence regarding Questions 2b-e and Questions 3b-e is presented together.

Main question:

2. What is the most appropriate method for screening women for depression in the perinatal period?

Sub-questions:

2a. What is the performance (defined as reliability, sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio) of validated tools for perinatal depression screening? [*addressed via systematic review*]

2b. What are the non-technical characteristics (defined as number of items, time to administer, complexity of scoring, and available languages) of validated tools for perinatal depression screening? [*addressed via descriptive review*]

2c. What is the acceptability to pregnant or post-partum women, health professionals, and the general public of screening for perinatal depression? [*addressed via narrative review*]

2d. What is the effectiveness (defined as impact on detection, care sought or received, and mental health outcomes) of screening for perinatal depression? [*addressed via narrative review*]

2e. What are the implications (for resourcing, workforce, training requirements, and models of care) of implementing perinatal depression screening (via different modes of delivery) with a validated tool? [*addressed via narrative review*]

B2.3 ANXIETY SCREENING

The focus of anxiety screening in this evidence review is on validated tools that have been developed or assessed in antenatal and/or postnatal women. As explained above, the current review is limited to instruments that examine current mental health. The specific tools included within scope of this review are detailed in later sections, presented according to mental health issue for Question 3a, and by tool for Questions 3b-3e.

Main question:

3. What is the most appropriate method for screening women for anxiety in the perinatal period?

Sub-questions:

3a. What is the performance (defined as reliability, sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio) of validated tools for perinatal anxiety screening? [*addressed via systematic review*]

3b. What are the non-technical characteristics (defined as number of items, time to administer, complexity of scoring, and available languages) of validated tools for perinatal anxiety screening? [*addressed via descriptive review*]

3c. What is the acceptability to pregnant or post-partum women, health professionals, and the general public of screening for perinatal anxiety? [*addressed via narrative review*]

3d. What is the effectiveness (defined as impact on detection, care sought or received, and mental health outcomes) of screening for perinatal anxiety? [*addressed via narrative review*]

3e. What are the implications (for resourcing, workforce, training requirements and models of care) of implementing perinatal anxiety screening (via different modes of delivery) with a validated tool? [*addressed via narrative review*]

B3 SEARCH METHODS

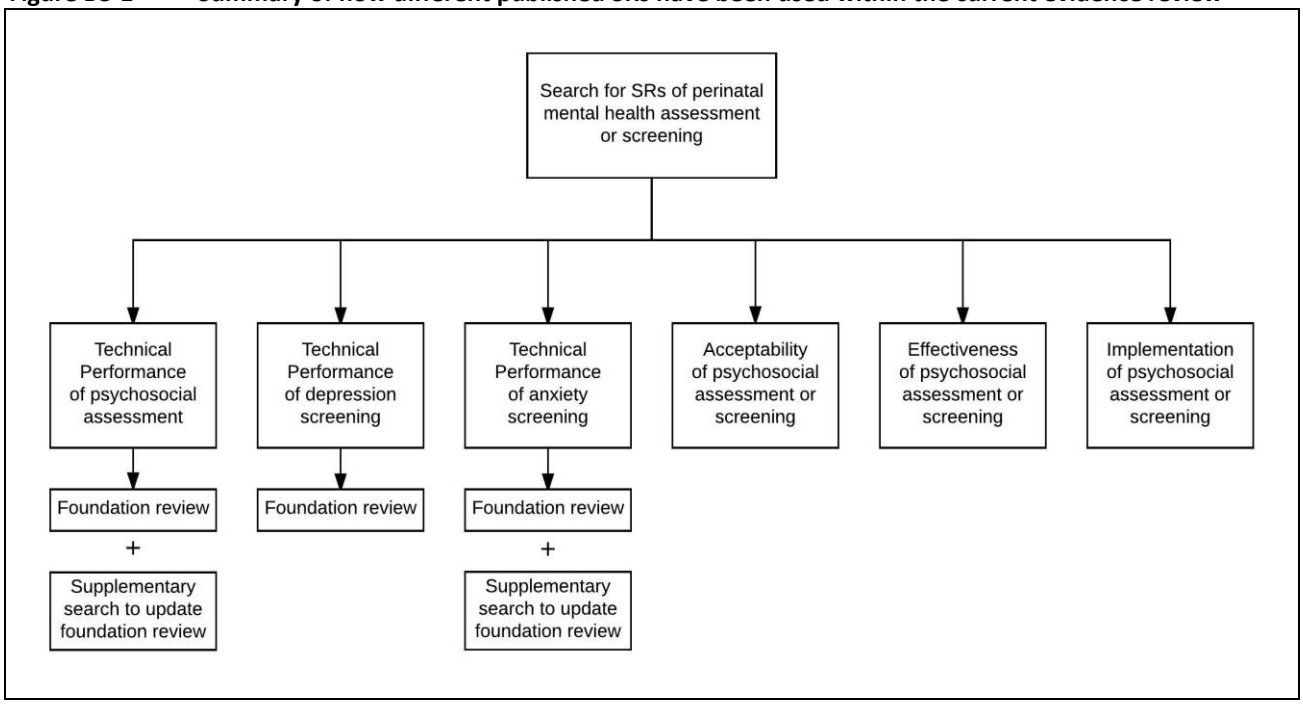
B3.1 OVERALL APPROACH TO SEARCHES

An initial search was undertaken to identify all published systematic reviews (SRs) of perinatal mental health screening and psychosocial assessment. One broad search was undertaken because there is significant variation in the literature regarding the terminology used to describe screening and psychosocial assessment. It was also recognised that some SRs would focus on screening or psychosocial assessment, whereas other others include both aspects of care. In addition, it was expected that some SRs would focus on technical performance, whilst others would focus on acceptability, effectiveness, and/or implementation issues. In other words, one search was undertaken to identify high level evidence across all aspects of the Evidence to Decision framework shown in Figure B1-1. Details of the search can found in **Section B8.1**.

Once the literature search was conducted, the included SRs were sorted (as per our definitions, see following sections for further details) according to whether they described psychosocial assessment, depression screening, or anxiety screening (i.e. the three 'topics').

The included SRs were reviewed and one SR selected as a 'foundation review' for technical performance of psychosocial assessment, depression screening, and anxiety screening. The rationale for selection of the foundation reviews is described in more detail below. Additional searches were then undertaken to update the foundation review and/or identify supplementary evidence. Additional searches were only undertaken after full data extraction and critical appraisal of the foundation reviews was complete. Results of supplementary searches are shown in the respective sections below.

Figure B3-1 Summary of how different published SRs have been used within the current evidence review



Abbreviations: SR, systematic review.

It should be noted that the mental health terms applied in the initial search related to depression and anxiety. This was based on advice from the EWG that although bipolar disorder, borderline personality disorder, schizophrenia and postpartum psychosis are all within scope for the current Guideline, these lower prevalence mental health conditions are not typically screened for in primary practice.

B3.1.1 Screening

To assess the technical performance of tools used for depression and anxiety screening, the current authors sought to identify published SRs that focused on diagnostic accuracy (which is also known as 'criterion validity' in the psychological literature) in the ante- and/or post-natal period. As the current Guideline sought to use GRADE methods for recommendation development, the EWG agreed that priority would be given to SRs that used GRADE or Cochrane risk of bias methods. The rationale for the selection of different foundation reviews for depression and anxiety screening are described below.

B3.1.1.1 Selection of foundation review for depression screening

The initial search identified seven SRs relevant to the technical performance of perinatal depression screening: Gibson 2009; Kozinzsky 2015; Mann 2011; Myers 2013; NICE 2015; O'Connor 2016; Thombs 2014. A comparison of the search dates and included study lists across the six SRs is provided in Section **B8.2.1**.

The SR by **Gibson 2009** was limited to studies of the EPDS and had a search date of July 2008. It was therefore excluded from further consideration.

The SR by **Kozinzsky 2015** is limited to consideration of the EPDS and was found to include only a subset of the studies included in NICE 2015, Myers 2013 and O'Connor 2016. It was therefore excluded from further consideration.

The SR by **Mann 2011** was limited to studies of the 'Whooley questions', and the single study included in this SR is also included in the SRs by NICE 2015, Myers 2013 and O'Connor 2016. It was therefore excluded from further consideration.

The SR by **Thombs 2014** was limited to studies of the EPDS and GHQ-12, had a search date of April 2013 but had almost no overlap of included studies with the SRs by NICE 2015, Myers 2013 or O'Connor 2016. Thombs 2014 was therefore excluded as evidence for technical performance of depression screening (but is included for consideration of screening effectiveness).

The three remaining SRs by **NICE 2016**, **Myers 2013** and **O'Connor 2016** are all judged to be recent, comprehensive, high quality reviews aligned with the clinical questions in the current Guideline. The stated NHMRC now has a stated preference for Australian guidelines to use GRADE methods. Consequently, given the range of tools included (EPDS, PHQ, Whooley questions, K-10), the use of Cochrane and QUADAS quality assessment methods, and the recent date of the literature search (April 2014), the EWG chose NICE 2015 as the foundation review for depression screening. A discussion of the approach and limitations of the NICE 2015 SR for screening for depression is provided below.

Although findings are presented for depression and anxiety within the NICE 2015 Guideline, the NICE literature search for anxiety was limited to studies reporting findings from the EPDS, PHQ, Whooley questions and the Kessler-10. As discussed in a separate Section, the EWG agreed to use a different published SR (Meades 2011) as the foundation review for anxiety screening.

B3.1.1.2 Supplementary search for individual studies of depression screening

The NICE 2015 Guideline included studies of diagnostic accuracy that met the following criteria:

- Conducted in a perinatal population,
- Reports on a psychometric instrument that includes 12 or fewer items,
- Reports sensitivity and specificity relative to a diagnostic interview for the relevant cut-off points,

- Published in English
- With no limits applied regarding study size or country/setting of studies.

For the studies of test accuracy to identify depression the current guideline relies on the studies identified by the NICE 2015 Guideline. Findings from studies of perinatal depression screening that use electronic modes of delivery are then considered in the context of screening implementation (see below).

Table B3-1 Criteria for determining study eligibility by NICE 2015 for depression screening

Study design	Diagnostic accuracy
Population	Pregnant and/or post-partum women
Intervention	EPDS (EDS), PHQ, K-10, 'Whooley questions'
Comparator	Any type of standardised diagnostic interview
Outcomes	Sensitivity, Specificity of detecting depression

Abbreviations: EPDS, Edinburgh Postnatal Depression Scale (also known as the Edinburgh Depression Scale, EDS); K-10, Kessler 10 item questionnaire; NICE, National Institute of Health and Care Excellence; PHQ; Patient Health Questionnaire.

The literature search for the NICE 2015 Guideline was conducted in April 2014. As noted above, two SRs on the topic of perinatal screening for depression have been published with more recent search dates: O'Connor 2016 (search date January 2015) and Shrestha 2016 (search date April 2015). The lists of included studies within each of these published SRs were reviewed to identify relevant individual studies that have been published since the NICE literature search was run. The majority of the additional individual studies published since NICE 2015 and included in O'Connor 0216 or Shrestha 2016 are validation studies of the EPDS in different languages. The EWG agreed that this aspect of the EPDS would be captured in the current evidence review under 'Available Languages' within the domain of Non-Technical characteristics of screening tools. Consequently, the EWG agreed that no update of the NICE 2015 SR was required.

That said, the EWG expressed their strong interest in the potential role of electronic modes of delivering perinatal screening (i.e. web-based, app-based, self-completed electronic forms etc). Thus, the current review highlights individual studies that describe screening for perinatal depression using electronic modes of delivery, in general, and where related to specific tools with evidence of adequate technical performance.

B3.1.1.3 Selection of foundation review for anxiety screening

Following the initial search for SRs of perinatal anxiety screening, four published systematic reviews on this topic were considered relevant by the EWG: Evans 2015 (search date September 2014), Meades 2011 (search date September 2010), Shrestha 2016 (search date April 2015) and NICE 2015. An additional narrative review was also considered potentially relevant (Matthey 2013b).

As described in more detail below the EWG considered the characteristics of each of the identified SRs (search span, psychometric instruments included, and comprehensiveness of studies included). Although the systematic review for the NICE 2015 guideline was chosen as the foundation review for depression screening, it has not been used as the foundation review for anxiety screening. This was because the EWG considered that there were a number of psychometric instruments that may represent credible perinatal anxiety screening tools that were not included in NICE 2015: DASS-21, EPDS (full and 3 question versions), GAD-2 or GAD-7, GHQ, HADS and HADS-A, K-10, and the STAI. Of these, NICE 2015 included only the EPDS and the K-10.

The review by **Evans 2015** describes the psychometric properties of self-report instruments to identify anxiety during pregnancy. The aim of the review was to identify optimal methods for clinicians and researchers. The research questions were (1) what instruments are available to identify anxiety during antenatal care? and (2) to what extent have the psychometric properties of the instruments been evaluated for use in a pregnant population? The review lists studies that reported on criterion validity, but does not present the findings considered important for the COPE guideline (i.e. sensitivity and specificity)

from those studies. Consequently, the EWG agreed this SR could not be the foundation review. However, the bibliography of the Evans 2015 review has been searched for individual studies as part of the literature search update for test performance the COPE guideline. In addition, the EWG noted that the discussion and analysis within Evans 2015 of the other psychometric properties of the instruments was relevant to the guideline. Findings from Evans 2015 are therefore discussed alongside conclusions from the evidence review update of test performance/criterion validity.

The second systematic review by **Meades 2011** focuses on a systematic examination of the validation of anxiety measures in perinatal populations. The review lists details of 30 studies, that each reported on one or more the following: criterion validity (relevant to the current ER), concurrent validity, and reliability. Across the studies that reported criterion validity findings, the following psychometric instruments were studied: HADS, GHQ and STAI. As the scope and approach of Meades 2011 aligns with the anxiety screening research question for the COPE guideline, the EWG agreed this SR would be the **foundation review** for this topic. The EWG agreed that the anxiety screening SR by Meades 2011 would be updated with a literature search from September 2010 to current dates.

The third systematic review by **Shrestha 2016** assessed the reliability and validity of the EPDS for detecting common mental disorders in the perinatal period among women in low-and lower-middle-income countries. The aim of this systematic review was to appraise formally validated local language versions of the EPDS from these resource-constrained settings. Because this SR was limited to the EPDS and developing countries, it did not meet the criteria for a foundation review for the purpose of systematically assessing test performance for the COPE guideline. However, the EWG agreed to the inclusion of Shrestha 2016 in the broader evidence base as it includes information potentially of relevance to implementation of recommendations in local culturally and linguistically diverse (CALD) populations.

The narrative review by **Matthey 2013b** discusses conceptual and methodological considerations related to the use of the EPDS to screen for anxiety disorders. The review does not present test performance measures within scope for the current evidence review of test performance, but it does provide a comprehensive discussion of issues associated with discriminating between cases of depression and/or anxiety. Although Matthey 2013b did not meet the criteria for a foundation review for this guideline, the EWG agreed that the aspects discussed by the authors would be of relevance during recommendation development. Consequently, the EWG agreed to include Matthey 2013b as part of the broader evidence base for this topic.

B3.1.1.4 Supplementary search for individual studies for anxiety screening

Although Meades 2011 was selected as the foundation review for the technical performance of anxiety screening tools, a number of issues were identified during data extraction from this SR:

- The quality ratings used by the foundation review (Meades 2011) are inconsistent with QUADAS-2 methods (i.e. the methods used by NICE 2015 and the current authors for depression screening)
- There is duplication of reporting in Meades 2011
- There is information relevant to our considerations that is missing in Meades
- There are significant differences across studies in their definition of 'cases' which needed to be known before decisions could be made regarding the appropriateness of pooling of data.

Consequently, a decision was made to use Meades 2011 as the means of identifying relevant individual studies, but with *de novo* data extraction and quality assessment from these studies, using QUADAS-2 methods. The supplementary search for anxiety screening was therefore focused on fully updating the Meades 2011 search.

The supplementary search sought to include only studies that met the following criteria:

- Conducted in a perinatal population

- Reports on one or more of the psychometric instruments nominated by the EWG (i.e. DASS-21, EPDS (full and 3 question versions), GAD-2 or GAD-7, GHQ, HADS and HADS-A, K-10, STAI),
- Reports sensitivity and specificity relative to a diagnostic interview for anxiety, using relevant cut-off points.

No limits were applied regarding study size or publication language (although English language abstracts were required for preliminary screening of search hits).

Studies included in the foundation review (Meades 2011) that met the above criteria were extracted. A literature search was then undertaken with a search start date of January 2010 which overlapped with the last search date of Meades 2011 (i.e. September 2010).

Table B3-2 Criteria for determining study eligibility for anxiety screening

Study design	Diagnostic accuracy
Population	Pregnant and/or post-partum women
Intervention	DASS-21, EPDS (full or 3 question), GAD-2 or GAD-7, GHQ, HADS or HADS-A, K-10, STAI
Comparator	Any type of standardised diagnostic interview
Outcomes	Sensitivity, Specificity of detecting anxiety

Abbreviations: DASS21, Depression Anxiety Stress Scales; EPDS, Edinburgh Postnatal Depression Scale; GAD-2 , Generalised Anxiety Disorder scale 2; GAD-7, Generalised Anxiety Disorder scale 7; GHQ, General Health Questionnaire; HADS, Hospital Anxiety and Depression Scale; HADS-A, Hospital Anxiety and Depression Scale for anxiety; K-10, Kessler 10 item questionnaire; STAI, State-Trait Anxiety Inventory

B3.1.2 Psychosocial assessment

B3.1.2.1 Selection of foundation review

The initial SR search identified five published SRs that were potentially relevant to the current Guideline: Walker 2015; Nilaweera 2014; Nast 2013; Myers 2013; Johnson 2012.

Based on the advice of the Expert Working Group (EWG) only specific, validated perinatal psychosocial assessment tools were to be included in the current review: antenatal psychosocial health assessment tool (ALPHA), antenatal risk questionnaire (ANRQ; also known as the perinatal risk questionnaire), Australian routine psychosocial assessment (ARPA), contextual assessment of maternity experience (CAME), Camberwell assessment of need – Mothers (CAN-M), pregnancy risk questionnaire (PRQ), and a perinatal specific risk factor assessment tool (RFA). As noted above, instruments that assess current mental health are not included within our definition of psychosocial assessment (but tools used to detect depression or anxiety are included in separate Sections herein).

The SR by **Nast 2013** sought to review psychometric instruments that have been used to assess 'psychosocial stress' during pregnancy. Specifically, they included studies on associations of maternal psychosocial stress during pregnancy with any behavioral fetal, infant or childhood outcome. The authors identified 58 instruments and extracted data on construct validity, concurrent and predictive validity, sensitivity, specificity and positive predictive values. However, none of the instruments included within the SR was in the list of multidimensional instruments specified by the EWG. Consequently, this SR was excluded from further analysis.

The SR by **Nilaweera 2014** addressed the prevalence, nature and determinants of postpartum mental health problems among women who have migrated from South Asian countries to high-income countries. The SR also considered the barriers and enablers to health care seeking among these women. The SR included 15 studies but none of these studies examined the performance of a multidimensional psychosocial instrument as pre-specified by the EWG. This SR is excluded from further analysis of the technical characteristics of tools, but is included in the narrative review of Acceptability [*see Section B.5*].

The SR by **Myers 2013** is a comparative effectiveness review undertaken by the Agency or Healthcare Research and Quality (AHRQ). The SR focused on screening instruments for postpartum depression but did

include one of the pre-specified multidimensional instruments (the ANRQ). However, the authors assessed the ANRQ as a screening tool for postpartum depression, not as a psychosocial assessment tool. Consequently, their assessment is not included in our current review (although the individual studies of the ANRQ are included, see below).

Although the focus of the Myers 2013 SR is the efficacy and safety of screening for postpartum depression, the authors did ask a question relevant to psychosocial assessment: "Are there individual factors (age, race, parity), history of mood disorders, history of intimate partner violence, perinatal outcomes, cultural factors) that affect baseline risk of postpartum depression and, therefore, the subsequent positive and negative predictive values of screening instruments?". For this question, the authors identified 15 studies that assessed associations between patient characteristics and the risk of postpartum depression. One or more studies found the following factors to be associated with an increased risk of postpartum depression: pre-term or low birthweight baby; past history of depression or anxiety; certain personality traits (neuroticism, vulnerability, low organisation); poor quality relationships; poor social support.

Walker 2015 undertook a SR of self-administered scales for measuring psychosocial and behavioural health that had been validated for postpartum women in the USA. The scales covered the domains of depression, body image, diet, physical activity, smoking and alcohol use. The authors examined the characteristics and psychometric properties of the 19 included scales. The majority of the scales focused on the detection of depression, and none of the psychosocial assessment tools pre-specified by the EWG was included. Consequently, this SR has been excluded from further consideration in the current review.

The aim of the SR by **Johnson 2012** was to critically analyse existing multidimensional tools that measure perinatal mental health risk and to report on the psychometric properties of these tools. The SR included most of the psychosocial assessment instruments pre-specified by the EWG, and provides an assessment of the reliability and validity of the tools, together with an overall rating of each instrument. This SR has been chosen as the **foundation review** for psychosocial assessment topic in the current review, and is described in more detail below.

B3.1.2.2 Supplementary search and inclusion criteria for individual studies

As noted above, the SR by Johnson 2012 was chosen as the foundation review for this topic as it completely aligns with the research questions for the current Guideline. However, given that this SR was more than two years old at the time of consideration by the EWG, a literature search was undertaken to update the studies included within Johnson 2012.

Specific terms for the relevant instruments (ALPHA, ANRQ, ARPA, CAME, CAN-M, PNRQ, and PRQ) as well as generic terms for perinatal psychosocial assessment were used in the literature search. The search was run in December 2016, with a search span from 1 January 2011 (to overlap with the search date of Johnson 2012). Full details of the literature search are included in **B8.1**.

As discussed below, the current review sought to find any relevant information on the technical characteristics (defined as reliability, validity and predictive accuracy) of the specified instruments in perinatal populations. It was recognised that different technical characteristics could be determined using different study designs, and so no limits were placed on study type as an inclusion criterion. Studies were included if they reported on at least one or more of the technical characteristics of interest.

Table B3-3 Criteria for determining study eligibility: psychosocial assessment – technical characteristics

Study design	Any type
Population	Pregnant and/or post-partum women
Test	ALPHA, ANRQ, ARPA, CAME, CAN-M, PNRQ, PRQ, 'Perinatal Risk Factor Assessment'
Technical characteristics	Reliability, validity, predictive accuracy

Abbreviations: ALPHA, Antenatal Psychosocial Health Assessment; ANRQ, Antenatal Risk Questionnaire; ARPA, Antenatal Routine Psychosocial Assessment; CAME, Contextual Assessment of Maternity Experience; CAN-M, Camberwell Assessment of Need – Mothers; PNRQ, Postnatal Risk Questionnaire; PRQ, Pregnancy Risk Questionnaire

Upon reviewing the literature search results, it became apparent that there were a number of studies that have used the EPDS in conjunction with structure psychosocial assessment. Some of these studies have been conducted with the psychosocial assessment tools specified above, but other have been conducted with unvalidated questionnaires. These studies have been excluded from consideration of the technical characteristics of the tools (i.e. reliability, validity, predictive accuracy), but have been included for considerations of acceptability, effectiveness and implementability (see below). This is because concurrent use of psychosocial assessment and the EPDS is already accepted as a model of care in Australia, and it was possible that these studies might provide useful contextual information within the Evidence to Decision framework.

B3.1.3 Acceptability, Effectiveness and Implementation

B3.1.3.1 Approach to evidence review

Each of these non-technical aspects of screening and psychosocial assessment has been addressed as a narrative review, based on the combined findings from the systematic review search for screening, and the targeted search for psychosocial assessment. Assessment of the technical performance of the included tools was completed first, and a judgment made by the EWG regarding the strength of the evidence for each tool. The consideration of evidence related to acceptability, effectiveness and implementation issues has then been limited to the EPDS (as this is the tool recommended by the EWG in the current guideline for depression screening) and those psychosocial assessment tools where there was moderate to high quality evidence of technical performance.

In practice, psychosocial assessment and screening for mental health issues occur at the same visit. Consequently, studies that evaluated the clinical usefulness of the EPDS together with any structured psychosocial assessment (with or without the use of validated tool) have been included (see table above).

B3.1.4 Cost-effectiveness of perinatal mental health screening

To address potential resourcing implications of screening, a separate search was undertaken to identify economic evaluations/cost-effectiveness analyses of perinatal screening for depression or anxiety. Full details of the literature search are included in **Section B8.1**.

B4 PSYCHOSOCIAL ASSESSMENT

B4.1 RELEVANT OUTCOMES AND CRITICAL APPRAISAL METHODS

The foundation review by **Johnson 2012** undertook a quality assessment that was consistent with published general methods for quality assessment of psychometric tests, but which was not entirely compatible with a GRADE approach. Consequently, the quality of every included study (as identified by Johnson 2012 and via the updated literature search) has been re-assessed by the current evidence review authors based on information in the primary source papers. Quality assessments have been based on published information regarding study design and the content validity, reliability and applicability of each instrument.

Standard methods for quality assessment of diagnostic tests (e.g. QUADAS-2) were not considered to be appropriate for psychometric tests used to identify psychosocial risk factors. The reasons for this are threefold: (1) the clinical value of the psychosocial assessment tools is not in the overall score, but in the responses to individual domains within the tool; (2) psychosocial assessment necessarily relies on a woman's self-report/recall of risk factors (some of which may have taken place during her childhood) and which are not readily verifiable (e.g. history of abuse, absence of caring relationship with her own mother). In other words, there is no reference standard; and (3) the value of psychosocial assessment is much greater than simply predicting the likelihood of depression or anxiety, so relying solely on the predictive accuracy fails to capture the full benefits associated with reducing risks to the woman, her infant, and her family.

Consequently, the critical appraisal of the included studies has been informed by the methods used in the foundation review by Johnson 2012, and adapted to provide 'GRADE-style' assessments of the quality of the evidence for each tool. Johnson 2012 relied on the criteria for critically analysing psychometric tests published by Hammill 1992. These authors proposed separate consideration of technical and non-technical characteristics of psychological instruments consistent with the key standards of the American Psychological Association. The technical characteristics were defined by Hamill 1992 as validity, reliability, sensitivity and specificity, and the availability of normative data. The authors proposed a scoring system based on the number of different measures reported/available for a particular instrument. However, the framework proposed by Hamill 1992 does not take account of the design of the studies used to generate the technical characteristics, or the broader applicability of the study population and setting to the current context (i.e. beyond the availability of normative data).

To adopt a 'GRADE-style' approach, the current evidence review considers the design of each included study, and then rates the quality of each study on the basis of study design, validity, reliability, and applicability. Validity includes face or construct validity but excludes criterion validity (this is because sensitivity and specificity are captured within the outcome of 'Predictive Accuracy'). Applicability has been defined as including three sub-domains of country, setting and availability of normative data. These domains are presented in Study Characteristics tables for each instrument, with an assessment of quality for each study. Findings regarding predictive accuracy are presented in Evidence Profile tables, and Overall Summary of Findings tables bring together evidence across all of the aspects of technical performance, non-technical characteristics and clinical usefulness.

B4.2 CHARACTERISTICS OF STUDIES OF TECHNICAL PERFORMANCE

Key characteristics of the studies included in Johnson 2012 are presented in **Table B4-1**. The supplementary literature search identified one additional study that met our inclusion criteria: Reilly 2015. The characteristics and quality assessment of this study are presented in **Table B4-2**.

To make a judgment on the quality of each included study the following decision rules have been applied:

- **High quality**: evidence from a prospective, controlled study (reference standard and/or concurrent controls) plus data on all three of the following psychometric properties: content validity (e.g. comprehensiveness of domains; description of the methods used to develop the instrument); reliability (e.g. inter-rater, test-retest, internal consistency), and applicability (e.g. normative data; consideration of relevant sociodemographic and psychological factors in an Australian population).
- **Moderate quality**: evidence from a prospective, controlled study (as above) plus data on two out of three of the following psychometric properties: content validity, reliability, and applicability information (all as defined above).
- **Low quality:** evidence from a prospective, controlled study (as above) plus data on one out of three of the following psychometric properties: content validity, reliability, and applicability information (all as defined above).
- **Very low quality**: evidence from a prospective, controlled study (as above) but no data on content validity, reliability or applicability; or any psychometric evidence from a non-prospective or uncontrolled study.

Upon further appraisal of the instruments included in Johnson 2012 it became apparent that the **CAN-M** has been designed for use in a population (pregnant women and mothers with current severe mental illness who are already receiving mental health care) that is very different to the target population for the current guideline (women under routine antenatal care with unknown past or current mental health status). Consequently, although the study of the CAN-M by Howard 2007 was assessed as being of high quality, the CAN-M has not been considered further in the current evidence review.

Similarly, the **CAME** has been developed and tested in women known to be at high risk, namely women with past or current major depressive disorder, and women living in poverty. Because women with a history of major depressive disorder and women living in poverty comprise a subset of the target population, the evidence for the CAME has been taken through to the Summary of Findings. However, there are issues regarding the generalisability of the evidence from Bernazzani 2005 to a general antenatal population.

Table B4-1 Design and psychometric properties of individual studies included in Johnson 2012

Study ID	Study characteristics (N)	Content Validity	Reliability	Applicability	Quality Notes
ALPHA					
Carroll 2005	*Study design:* Cluster randomised controlled trial *Study population(s):* Pregnant women undergoing routine prenatal care: ALPHA group (n=98 women; 21 providers), control group (n=129 women; 27 providers) N=227	*Included domains:* • Family factors • Maternal factors • Substance use • Family violence *Method of development:* Based on critical review of literature and expert consensus	*Reliability measures:* None reported	*Country:* Canada *Setting:* Variety of practice locations and antenatal care providers *Normative data:* Yes; describe relationship between family and maternal factors, substance use, family violence and identification of psychosocial concerns	**Moderate** (●●●○) Based on evidence of validity and applicability from a prospective controlled study
ANRQ					
Austin 2013	*Study design:* Prospective cohort with reference standard (CIDI to a subset of participants) to describe technical characteristics; subset of women included in study of PRQ by Austin 2005 (see below) Cross-sectional survey to ascertain acceptability *Study population(s):* Pregnant women (N=1,196) Pregnant women (n=378; subset of main cohort) and midwives (n=44)	*Included domains:* • Emotional support from subject's own mother in childhood • Past history of depressed mood or mental illness and treatment received. • Perceived level of support available following the birth of the baby. • Partner emotional support. • Life stresses in previous 12 months. • Personality style (anxious or perfectionistic traits). • History of abuse (emotional, physical and sexual). *Method of development:* 12 items extracted from the original 23 item PRQ. Developed by a panel of experts based on systematic literature reviews of postnatal depression risk factors, and on face and construct validity of these factors.	*Reliability measures:* None reported	*Country:* Australia *Setting:* Hospital-based maternity clinic *Normative data:* Yes; compares sociodemographic and clinical profiles of women in the analysis subset versus women not in the analysis subset	Moderate (●●●○) Based on evidence of validity and applicability from a prospective controlled study
ARPA					
Mathey 2004	*Study design:* Prospective case series *Study population(s):* Pregnant women (76% in second trimester) N=2,167	*Included domains:* • Support • Stressors • Personality • Mental health • Childhood abuse • Family violence • Current mood measured with the EPDS	*Reliability measures:* No reliability/stability testing of items was reported.	*Country:* Australia *Setting:* Hospital-based antenatal clinic *Normative data:* No	**Very low** (●○○○) Due to uncontrolled study design

Study ID	Study characteristics (N)	Content Validity	Reliability	Applicability	Quality Notes
		Method of development: Face/content validity of items derived from existing known psychosocial risk factors. 12 item version derived from analysis of initial 31 item instrument that was found to have redundancies and ceiling effects.			
CAN-M					
Howard 2007	*Study design:* Prospective controlled studies of inter-rater and test-retest reliability Prospective validity study comparing CAN-M assessment of needs with Global Assessment of Functioning (GAF) *Study population(s):* Pregnant women and mothers with severe mental illness (SMI; psychotic disorder or chronic non-psychotic disorder) Reliability studies: N=31 SMI women, N=34 health professionals Validity study: N=63 SMI women	*Included domains:* • Accommodation • Food • Looking after the home • Self-care • Daytime activities • General physical health • Pregnancy care • Sleep • Psychotic symptoms • Psychological distress • Information • Safety to self • Safety to child and others • Substance misuse • Company • Intimate relationships • Sexual health • Violence and abuse • Practical demands of childcare • Emotional demands of childcare • Basic education • Telephone • Transport • Budgeting • Benefits • Language, culture and religion *Method of development:* Based on structure, format and coding algorithm of CAN. Identification of new domains based on interviews with women with severe mental illness, and findings assessed by expert steering group. Three versions: CAN-M-S (short version for routine clinical use), CAN-M-R (full version for research), CAN-M-C (full version for broader clinical assessment).	*Test-Retest reliability:* For longer research version (CAN-M-R): 0.91 (service users), 0.85 (staff) *Inter-rater reliability:* For longer research version (CAN-M-R): 0.93 (service users), 0.83 (staff)	*Country:* UK *Setting:* Inpatient or community-based mental health services *Normative data:* Yes; description of sociodemographic and psychological factors for study cohort.	**High** (●●●●) Prospective controlled studies with information for all domains but note significant issues regarding generalisability of study population to target population

Study ID	Study characteristics (N)	Content Validity	Reliability	Applicability	Quality Notes
CAME					
Bernazzani 2005	*Study design:* Prospective study with reference standard (SCID) Retrospective study with comparison between CAME and other psychometric instruments *Study population(s):* Pregnant women with past or current major depressive disorder, N=85 Postnatal women living in poverty, N=60	*Included domains:* • Recent life adversity or stressors • Quality of social support including partner relationship • Maternal feelings towards pregnancy, motherhood and the baby *Method of development:* Used a vulnerability-stress theoretical model, together with semi-structured interviews to identify and test relevant domains	*Internal consistency:* Study 1: internal consistency for prenatal social support components: α=0.86 for partner scale; α =0.81 for other significant relationship.	*Country:* UK *Setting:* Not stated *Normative data:* Partial; some description of sociodemographic factors and recent life adversity.	**Moderate** (●●●○) Based on validity and reliability data from prospective controlled study, but only limited applicability data *but note significant issues regarding generalisability of study population to target population*
PRQ					
Austin 2005	*Study design:* Prospective cohort with reference standard (CIDI to a subset of participants) to describe technical characteristics *Study population(s):* Pregnant women (N=1,296)	*Included domains:* • Mother's attitude to her pregnancy. • Mother's experience of parenting in childhood. • History of physical or sexual abuse. • History of depression. • Impact of depression on psychosocial function. • Whether treatment was sought or recommended. • Presence of emotional support from partner and mother. • Presence of other support. • Presence of stressors during pregnancy. • Trait anxiety. • Obsessional traits. • Self-esteem. *Method of development:* Developed by a panel of experts, based on past reviews of postnatal depression risk factors, and on face and construct validity of these factors.	*Reliability measures:* None reported	*Country:* Australia *Setting:* Hospital-based maternity clinic *Normative data:* Yes, reports relationship between key sociodemographic and psychological variables with CIDI depression	Moderate (●●●○) Based on evidence of validity and applicability from a prospective controlled study
Footnotes					

Abbreviations: ALPHA, Antenatal Psychosocial Health Assessment; ARPA, Antenatal Routine Psychosocial Assessment; ANRQ, Antenatal Risk Questionnaire; CAME, Contextual Assessment of Maternity Experience; CAN-M, Contextual Assessment of Maternity Experience; EPDS, Edinburgh Postnatal Depression Scale; PRQ, Pregnancy Risk Questionnaire; CIDI, Composite International Diagnostic Interview; SCID, Structural Clinical Interview for DSM Disorders.

Table B4-2 Design and psychometric properties of individual included studies published since Johnson 2012

Study ID	Study characteristics (N)	Content Validity	Reliability	Applicability	Overall certainty Notes
ANRQ					
Reilly 2015	*Study design:* Prospective study of co-administration of the ANRQ and EPDS to predict cases of depression or anxiety Study population(s): Postnatal women N=220	*Included domains:* As above (Austin 2013) *Method of development:* As above (Austin 2013)	*Reliability measures:* None reported	*Country:* Australia *Setting:* Hospital maternity unit *Normative data:* Yes; describes sociodemographic factors and psychosocial profile relative to EPDS scores	Moderate (●●●○) Based on evidence of applicability from a prospective controlled study and known evidence of validity
Footnotes					

Abbreviations: ANRQ, Antenatal Risk Questionnaire; EPDS, Edinburgh Postnatal Depression Scale

B4.3 EVIDENCE PROFILE TABLES

Evidence regarding the predictive capacity of each of the instruments from the included studies is presented below.

Table B4-3 Evidence regarding the technical performance of the included psychometric instruments

Evidence base		Performance[3]		Overall assessment of performance Quality
k (N)	Study ID(s)	Predictive accuracy	Concurrent validity	
ALPHA				
1 (227)	Carroll 2005	After accounting for provider dropouts: OR of identifying a concern, 1.00 (95% CI: 0.6-1.7); OR of identifying a high level of concern, 2.8 (95% CI: 0.7-11.7); OR of identifying family violence, **2.7 (95% CI**: **1.1-6.9)**	Sens, NR Spec, NR PPV, NR NPV, NR	Limited **Moderate** (●●●○)
ANRQ				
2 (1,416)	Austin 2013 Reilly 2015	OR (ANRQ score ≥23 is also a depression case), **6.3 (95% CI**: **3.5-11.5)** *The cut-off (23 out of a possible 62) was based on 'known groups' using a diagnostic interview on women with high depression scores or items identifying distress*	Sens, 0.62 Spec, 0.64 PPV, 0.30 NPV, 0.87 AUROC 0.69 (95% CI; 0.61-0.77)	Acceptable **Moderate** (●●●○)

Evidence base		Performance[3]		Overall assessment of performance Quality
k (N)	**Study ID(s)**	**Predictive accuracy**	**Concurrent validity**	
ARPA				
1 (2,167)	Mathey 2004	Not reported	Sens, NR Spec, NR PPV, NR NPV, NR	Unknown **Very low** (●○○○)
CAME				
1 (85)	Bernazzani 2005	Relative risk ratio of exposure to severe adversity and subsequent development of depression, **1.57 (95% C:I 1.06-2.33)**	Sens, NR Spec, NR PPV, NR NPV, NR	Limited **Moderate** (●●●○)
PRQ				
1 (1,296)	Austin 2005	OR (PRQ score >46 is also a depression case), **9.18 (p <0.001)** (at the maximum κ).	Sens, 0.44 Spec, 0.92 PPV, 0.235 NPV, 0.968 AUROC 0.788 (95% CI 0.727-0.848) The AUC between the PRQ and the EDS were significantly different (0.788 and 0.659, respectively, p<0.001).	Acceptable **Moderate** (●●●○)

Evidence Statements:

The ALPHA is effective at identifying family violence (moderate quality evidence).

The ANRQ is effective at predicting cases of depression (moderate quality evidence).

The predictive performance of the ARPA is unknown (very low quality evidence).

The predictive performance of the CAME is unknown (moderate quality evidence).

The PRQ is effective at predicting cases of depression (moderate quality evidence).

Abbreviations: ALPHA, Antenatal Psychosocial Health Assessment; ANRQ, Antenatal Risk Questionnaire; ARPA, Antenatal Routine Psychosocial Assessment; AUC, area under the curve; AUROC, area under the receiver-operator curve; CAME, Contextual Assessment of Maternity Experience; CI, confidence interval; EDS, Edinburgh Depression Scale; NPV, negative predictive value; OR, odds ratio; PPV, positive predictive value; PRQ, Pregnancy Risk Questionnaire; Sens, Sensitivity; Spec, Specificity.

B4.4 NON-TECHNICAL CHARACTERISTICS OF RELEVANT TOOLS

The table below summaries the non-technical characteristics of the three psychosocial tools with high or moderate quality evidence of technical performance. The complexity of scoring for each tool has been assessed as Simple, Moderate or High on the basis of information in the published literature and the experience of the EWG.

Table B4-4 Non-technical characteristics of the relevant included tools

Tool	Number of items	Time to administer (mins)	Complexity of scoring	Available languages
ALPHA	35	>10 minutes[1]	Simple *Three-point scoring for each question*	English
ANRQ	12	5-10 minutes	Moderate *Combination of categorical and continuous data*	English Mandarin
PRQ	21	10-20 minutes	Moderate *Five-point Likert scale for each question*	English

Abbreviations: ALPHA, Antenatal Psychosocial Health Assessment; ANRQ, Antenatal Risk Questionnaire; PRQ, Pregnancy Risk Questionnaire

[1] assumed based on number of items and comparison with PRQ

B4.5 CLINICAL USEFULNESS OF RELEVANT TOOLS

As shown above, three psychosocial assessment tools met our criteria for high or moderate quality evidence of technical performance in a relevant population: **ALPHA, ANRQ** and **PRQ**. The evidence regarding the clinical usefulness of these tools (based on the studies identified in the supplementary literature search) is described below.

Additional evidence of the clinical usefulness of psychosocial assessment that is not tool-specific but is considered relevant to the Australian context is also described below. In particular, Australian studies that describe co-administration of the **EPDS plus structured psychosocial assessment** (with or without one of the included tools) are described. This part of the evidence review was undertaken after all of the evidence regarding the technical characteristics of tools for depression or anxiety screening had been completed, and recommendations for screening had been drafted by the EWG. Thus, it was clear that the EPDS was the recommended tool for depression screening, and hence why studies that considered the EPDS (but not other depression or anxiety screening tools) alongside psychosocial assessment were considered relevant for inclusion.

B4.5.1 ALPHA

The study included for technical performance of the ALPHA (**Carroll 2005**) also reported on the acceptability and effectiveness of the tool. Regarding acceptability, 73% of women interviewed felt comfortable discussing personal issues and 76% of women felt that this was part of their health providers' job. Of the 21 providers who administered the ALPHA, 14 completed the feedback form. Of these 86% (i.e. 12) agreed they would use the ALPHA if it was recommended as standard practice.

Overall, the detection of any concern was 1.17 concerns per woman with the ALPHA versus 0.74 concerns per woman in the control group (OR 1.8; 95% CI 1.1-3.0). The ALPHA was also associated with a higher rate of reporting 'high risk' concerns: OR 4.8, 95% CI 1.1-20.2). For individual concerns, there was a trend for the proportion of women identified with a psychosocial concern to be higher for the group administered the ALPHA than for the control group although the differences were only statistically significant for the concern 'experienced or witnessed abuse as a child' (14.3% versus 2.3%; OR 7.0 (99% CI 1.3-37.5).

The paper by Carroll 2015 notes that the majority of providers in the study were GPs, and note that the results might not be generalisable to midwives or obstetricians. Providers noted concerns regarding the time required to administer the ALPHA, and a self-report version has been developed (but not tested in this study).

Table B4-5 Evidence from the supplementary literature search regarding the acceptability of the ALPHA

Study ID	Acceptability		
	To women (pregnant/postpartum) (N)	To healthcare providers	To general public
Carroll 2015	• 73% of pregnant women felt comfortable discussing personal issues (N=98)	• 86% of providers would use the tool in standard practice (N=14)	None reported

Abbreviations: ALPHA, Antenatal Psychosocial Health Assessment.

Table B4-6 Evidence from the supplementary literature search regarding the effectiveness of the ALPHA

Study ID	Effectiveness		
	Detection rates	Impact on care sought or received	Impact on mental health outcomes
Carroll 2015	• ALPHA associated with higher rate of detection of any psychosocial concern: OR 1.8 (95% CI: 1.1-3.0). • ALPHA associated with higher rate of detection of high risk psychosocial concern: OR 4.8 (95% CI: 1.1-20.2) • ALPHA associated with higher rate of reporting family violence: 14.3% versus 2.3%	None reported	None reported

Abbreviations: ALPHA, Antenatal Psychosocial Health Assessment; CI, confidence interval; OR, odds ratio.

Table B4-7 Evidence from the supplementary literature search related to the implementability of the ALPHA

Study ID	Implementability		
	Training requirements	Implications for models of care	Resource implications
Carroll 2015	• A single 1-hour interactive workshop	• Majority of providers were GPs, generalisability to other health professionals unknown • Self-report version may address time constraints of GPs	Not reported

B4.5.2 ANRQ

The two studies included with evidence of the technical performance of the ANRQ (**Austin 2013** and **Reilly 2015**), also include evidence regarding acceptability, effectiveness or implementation. Both studies included the use of the ANRQ as an adjunct to the EPDS. The study by Reilly 2015 was a feasibility study of a new model of care that aimed to integrate psychosocial assessment and referral pathways within an Australian private hospital maternity setting. Based on evidence from a study outside the scope of the current review (Priest 2008), a score of 23 or more on the ANRQ is considered to be clinically significant i.e. consideration of mental health assessment is warranted. In addition, positive scores on the ANRQ items pertaining to history of depression, other psychiatric diagnosis, or abuse/emotional neglect in childhood, are considered to warrant referral to a mental health intake meeting.

Acceptability of the ANRQ was found to be high among pregnant women and midwives in both studies. In Austin 2013, acceptability to women was ascertained by asking the question “Was any aspect of this questionnaire distressing to you? If so which question(s)?”. Of the 379 women participants, 92% found the ANRQ “not at all” distressing, and 1% found it “much” or “very much” distressing. Acceptability to midwives was ascertained by asking the questions: “How comfortable are you about using the ANRQ?”, and “How useful have you found the ANRQ for identifying women at risk, and planning care?”. Of the 44 midwives

who participated in the study, 40 responded to the feedback survey. Of these 70% were "very comfortable", 25% were "somewhat comfortable", and 5% were "uncomfortable" using the ANRQ. Of the midwives, 70% found the ANRQ "very useful" and 28% found the ANRQ "moderately useful" for identifying women at risk and planning care. In Reilly 2015, less than 1% of women reported that they were not comfortable completing the ANRQ, with the majority reporting they felt 'comfortable' or 'very comfortable' completing the questionnaire. There was no significant difference in the acceptability ratings of women who scored above or below the recommended cut-off of 23 for the ANRQ, or above or below the recommended cut-off of 13 or more for the EPDS.

It should be noted that the ANRQ can also be used in a postnatal setting (as all of the domains included within the ANRQ remain relevant postnatally). When the ANRQ is used postnatally three additional items are added relating to the baby and the delivery.

Table B4-8 Evidence from the supplementary literature search regarding the acceptability of the ANRQ

Study ID	Acceptability		
	To women (pregnant/postpartum) (N)	To healthcare providers	To general public
Austin 2013	• 92% of pregnant women found the ANRQ "not at all distressing" (N=379)	• 95% of midwives were very or somewhat comfortable administering the ANRQ (N=40) • 98% of midwives found the ANRQ very or moderately useful for identifying women with risk factors, and planning care	None reported
Reilly 2015	• 97% of pregnant women felt 'comfortable' or 'very comfortable' completing the ANRQ (N=220)	None reported	None reported

Abbreviations: ANRQ, Antenatal Risk Questionnaire.

Table B4-9 Evidence from the supplementary literature search regarding the effectiveness of the ANRQ

Study ID	Effectiveness		
	Detection rates	Impact on care sought or received	Impact on mental health outcomes

Reilly 2015	• 5% of pregnant women scored above the recommended EPDS cut-off of 12
	• 32% of pregnant women scored above the recommended ANRQ cut-off of 23
	• Proportion of pregnant women with psychosocial risk factors:
	  • No risk factors, 45%
	  • 1 risk factor, 24%
	  • 2 risk factors, 18%
	  • 3 or more risk factors, 13%
		• 11% of women were referred for additional support or treatment:
		  • 6% referred to GP
		  • 3% referred to private psychologist or psychiatrist
		  • 2% referred to community perinatal mental healthcare team
			None reported

Abbreviations: ANRQ, Antenatal Risk Questionnaire; EPDS, Edinburgh Postnatal Depression Scale.

Table B4-10 Evidence from the supplementary literature search related to the implementability of the ANRQ

Study ID	Implementability		
	Training requirements	**Implications for models of care**	**Resource implications**
Austin 2013	• Tool is brief and user-friendly and was designed in collaboration with midwives and mental health care professionals • *Training limited to the administration of the ANRQ takes approximately 1 hour (as a workshop)*[1] • *Training that includes use of the ANRQ, use of the EPDS, and guidance around scoring and referral decisions based on findings takes approximately 3-4 hours (as a workshop)*[1]	• A total score ≥23 is considered to be clinically significant and warranting consideration • ANRQ administered as adjunct to EPDS identified more women than EPDS alone as requiring further assessment, monitoring or referral • Can be used as self-report questionnaire given by midwives	• Limited costs: questionnaire is freely available and takes 5 minutes to complete with extra time to explore significant items as they arise; can be included within existing antenatal visits
Reilly 2015	• One-day training workshop for all midwives delivered by a psychiatrist	• Aim of study was to ascertain feasibility of a new model of care in a private maternity practice that integrates psychosocial assessment and referral pathways to inpatient and community mental health services. • Model of care remains embedded at research site, demonstrating ease of use, clinical relevance and appropriateness	• Continuous availability of midwives trained in the use of the ANRQ + EPDS is required to ensure all women undergo assessment • Availability of mental health care team to midwives was seen as critical for implementation success
[1] Personal communication from authors			

Abbreviations: ANRQ, Antenatal Risk Questionnaire; EPDS, Edinburgh Postnatal Depression Scale.

B4.5.3 PRQ

No additional individual studies were identified for the PRQ, beyond the study (**Austin 2005**) included in the foundation review (Johnson 2012). Limited information on effectiveness and implementability were included in that paper, and is presented in the tables below.

Table B4-11 Evidence from the supplementary literature search regarding the effectiveness of the PRQ

Study ID	Effectiveness		
	Detection rates	**Impact on care sought or received**	**Impact on mental health outcomes**
Austin 2005	• 17/18 items on questionnaire found to be statistically significantly associated with presence of CIDI-diagnosed depression	None reported	None reported

Abbreviations: CIDI, Composite International Diagnostic Interview; PRQ; Pregnancy Risk Questionnaire

Table B4-12 Evidence from the supplementary literature search related to the implementability of the PRQ

Study ID	Implementability		
	Training requirements	**Implications for models of care**	**Resource implications**
Austin 2005	None reported	• PRQ assessed by authors as not meeting the necessary criteria for routine screening in a public health setting due to its length	None reported

Abbreviations: PRQ; Pregnancy Risk Questionnaire

B4.6 GENERAL EVIDENCE OF CLINICAL USEFULNESS

Four studies were identified in the supplementary search that describe the use of the EPDS in conjunction with structured psychosocial assessment (but not with the ALPHA, ANRQ or PRQ): Kohlhoff 2016; Matthey 2016; Rollans 2013; Quispel 2012. One of the studies included in the foundation review (Matthey 2004)

also describes the use of the EPDS in conjunction with structured psychosocial assessment. The design of and findings from each of these four studies is described below.

The study by **Kohlhoff 2016** describes referral rates associated with the introduction of an antenatal psychosocial assessment and depression screening program at a private obstetric hospital in Australia. A total of 993 pregnant women participated in a structured psychosocial assessment interview and completed the EPDS. Midwives were trained to deliver the psychosocial assessment interviews, and referrals were made on the basis of an EPDS score of 13 or more, or the presence of significant psychosocial risk. A total of 94 women (9.5%) were identified to the obstetrician who then made appropriate referrals for clinical support, and a further 90 women (9.0%) were identified by the midwives and referred for other types of support (e.g. to a social worker). Overall, 6% of the cohort was found to have an EPDS score of 13 or more, and a further 14% had an EPDS score of 10-12. The authors note that higher socioeconomic status does not protect against psychosocial risk factors or mental health issues, and emphasise the importance of establishing routine psychosocial assessment and depression screening in private hospitals (given that 30% of women in Australia deliver their babies in this setting).

Matthey 2004 recruited a consecutive sample of pregnant women (N=2173) who presented to an Australian public hospital antenatal clinic. The study presents relevant information on the rates of detection of different psychosocial risk factors using the ARPA in conjunction with the EPDS: of the women in the main sample 10% reported an abusive childhood, 5% reported domestic violence within the family, 23% reported personality traits associated with higher risk of mental health issues, 24% reported recent stressors, and 19% reported a history of depression or anxiety. Overall, more than 50% of the sample reported at least one psychosocial risk factor: 40% had 1-2 risks, 10% had 3-4 risks, and 2% had 5-7 risks. A sub-sample of women (N=1050) were followed longitudinally to measure their use of referral services. Of the 294 women from this sub-sample who were offered a referral, 131 accepted the referral, and of these 33% participated in phone support with a clinician, and 31% participated in one or more face-to-face counselling sessions. The authors discuss the usefulness of psychosocial assessment tools as providing an opportunity for exploring risk responses, so that the likelihood of coping difficulties (antenatally and postnatally) is reduced.

The article by **Matthey 2016** describes the experience of an Australian public hospital antenatal clinic with routine antenatal screening for mental health issues. Specifically, the study describes the impact of changing the threshold for referral to a 'Safe Start' Meeting (a weekly multidisciplinary meeting to discuss assessment of a woman's needs and referral to mental health services if required). When the Safe Start program was first introduced at the hospital, all pregnant women with an EPDS score of 10 or more were referred to a Safe Start meeting. However, the impression of the health professionals was that women who scored 10-12 on the EPDS rarely needed the specialist triaging provided by the Safe Start service. Based on a review of the evidence, the hospital revised its referral threshold so that all women with an EPDS score of 13 or more were still referred, but women scoring 10-12 on the EPDS were only referred is at least one psychosocial risk factor was assessed as being present. Women who scored 10-12 on the EPDS with no psychosocial factors were provided with a Letter from the clinic that provided details of the hospital's social work department should she wish to discuss any psychological or social issues. The impact of this change was to reduce the number of referrals to Safe Start meetings by 20%. None of the women who scored 10-12 on the EPDS subsequently sought services from the hospital's social work department. The authors concluded that the change in referral threshold did not appear to result in any women in need being missed, and allowed the re-allocation of time within safe Start meetings to women with high psychosocial needs.

The paper by **Quispel 2012** describes an observational study in which pregnant women in the Netherlands were asked to self-complete a single questionnaire comprised of the EPDS plus questions about psychosocial problems. The psychosocial assessment component of the questionnaire was developed for the study and the items include psychiatric history, substance use, financial or housing problems, past or

current physical or sexual abuse, and presence of relational problems. The novel aspect of the screening model was that the results of the psychosocial assessment are not disclosed to the health professional administering the assessment, but the tool instead suggests tailored intervention advice. This advice can then be discussed during the antenatal booking visit. Time to complete the full questionnaire was typically 10 minutes, and it could be delivered by non-mental health professionals. No data were collected on the acceptability of the tool, but the authors note that digital surveys have been advocated for the collection of sensitive data.

Finally, **Rollans 2013** describes an Australian perspective of women's experience of combined psychosocial assessment and screening during pregnancy and following birth. This was a qualitative study that found that most participants found it acceptable to be asked the psychosocial questions although they felt unprepared for the sensitive nature of the questions. Of note, women with a history of trauma or loss were distressed by retelling their experiences. The authors concluded that it is crucial that health professionals are educationally prepared for this work and receive ongoing training and support in order to deliver care that is sensitive and empathetic.

B4.7 OVERALL SUMMARY OF FINDINGS

The table below shows the overall summary of findings regarding all relevant aspects of perinatal psychosocial assessment: technical characteristics/performance, non-technical characteristics and clinical usefulness.

As shown in the table below, the tool that scores highest across all domains of interest is the ANRQ/PNRQ.: there is moderate quality evidence that this tool has acceptable technical performance, that it is easy to administer in practice, that it has high acceptability among pregnant women and midwives, and that it impacts positively on the rates of referral for further mental health assessment. The ANRQ is currently the only tool available in a language other than English. No published evidence has been identified describing the use of any of these tools in culturally and linguistically diverse populations of women. It should also be noted that all of the available evidence included in this review has been derived from studies of tools administered in the antenatal period.

Table B4-13 Overall Summary of Findings related to the use of perinatal psychosocial assessment tools

Tool(s)	Technical characteristics		Non-technical characteristics		Clinical usefulness		
	Performance[1]	**Certainty [2]**	**Ease of Administration[3]**	**Language availability[4] & cultural sensitivity[5]**	**Acceptability[6]**	**Effectiveness[7]**	**Implementability[8]**
ALPHA	Limited	Moderate (●●●○)	Moderate	English only; Cultural sensitivity unknown	Moderate	Limited	Limited
ANRQ	Acceptable	Moderate (●●●○)	High	English & Mandarin; Cultural sensitivity unknown	High	Good	High
PRQ	Acceptable	Moderate (●●●○)	Moderate	English only; Cultural sensitivity unknown	Unknown	Unknown	Limited

Footnotes

[1] Performance defined as predictive accuracy, sensitivity, specificity, positive predictive value and/or negative predictive value (defined as Acceptable, Limited, or Unknown).

[2] Certainty assessed on the basis of study design and evidence of validity, reliability and applicability (defined as High, Moderate, Low or very Low).

[3] Ease of administration was based on judgment regarding the number of items, and the time and complexity of administering and scoring the tool (rated as High, Moderate, or Low)

[4] Language availability based on information from the included literature and the awareness of the EWG

[5] Cultural sensitivity was based on information from the included literature of any use in culturally and linguistically diverse populations

[6] Acceptability was based on the overall judgement of the EWG of the acceptability of each tool to women, health care professionals and/or the general public (rated as High, Moderate, Low or Unknown)

[7] Effectiveness was defined as positive impact on the number of psychosocial risk factors identified, services referred to or utilized, and impact on a woman's mental health (rated as High, Good, Limited, or Unknown)

[8] Implementability was based on the overall judgement of the EWG based on available information regarding the training requirements for use of the tool and implications for current models of care and staff and service availability

Abbreviations: ALPHA, Antenatal Psychosocial Health Assessment; ANRQ, Antenatal Risk Questionnaire; PRQ, Pregnancy Risk Questionnaire; EWG, Expert Working Group

B5 SCREENING FOR DEPRESSION

B5.1 RELEVANT OUTCOMES OF TECHNICAL PERFORMANCE

The meta-analyses presented in the NICE 2015 review only included those studies that had reported sufficient information to calculate true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN).

Although TP, TN, FP and FN were calculated by the NICE 2015 Guideline authors and included in Appendix 19, they were not presented within the body of the NICE 2015 Guideline and do not appear to have been used to inform the develop of recommendations. The steering group for the NICE 2915 Guideline specified sensitivity and specificity as the 'critical outcomes'. They specified Positive Predictive value (PPV), Negative Predictive Value (NPV) and Area Under the Curve (AUC) as important but not critical outcomes.

By contrast the Expert Working Group for the current Guideline have identified **positive and negative likelihood ratios**, and the **AUC** (for different cut-offs) as the **critical outcomes**, and sensitivity and specificity as important outcomes. The rationale for this is described below.

In agreeing on the relative importance of different test performance measures, the EWG considered the place of a perinatal depression screening tool in the Australian setting. A previous guideline by beyondblue recommend in favour of the use of the EPDS by all 'first line' health professionals who come in contact with pregnant women or women in the first year postpartum: midwives, child and family health nurses, maternal child health nurses, general practitioners and obstetricians.

It is recognised that there may be some issues associated with the suitability of the EPDS for all of the first line health professionals named above. Such issues relate to the time required to administer the instrument, confidence of the health professionals to administer the tools and respond appropriately, and knowledge of referral pathways. These issues have been considered during the development of recommendations within the current guideline.

The EWG agreed that the primary goal of a tool for screening for perinatal depression is to identify women at increased risk of mental health issues to facilitate referral to appropriate services to allow further assessment and intervention if required. The EWG discussed the clinical consequences of different test results, and these are summarized in **Table B5-1**. Given the place of depression screening in perinatal care, and recognition that under-reporting is more likely than over-reporting, there was unanimous agreement that it was most important to minimize false negatives, even if that is associated with an over-representation of False Positives.

In discussing the consequences of testing with psychometric instruments the EWG noted that appraisal of the technical performance of these tools using QUADAS-2 methods does not fully capture the patient-relevant benefits of testing. It is recognised that women regularly present in crisis who had participated in screening and who had chosen to provide false responses due to a range of factors such as shame, lack of safety in the relationship with health professionals, fear of consequences, or belief that nothing can help.

Health professionals witness the frequent trajectory from mild or moderate to severe perinatal mental illness and the potentially devastating outcomes of delayed help-seeking on the individual woman, the infant and family unit. With high prevalence disorders such as depression and anxiety it is often a lack of early validation, support and treatment that has led to delayed help-seeking for distressing symptoms (e.g., suicide thoughts, intrusive thoughts of harm to baby, inability to care for infant). The EWG agree that these issues could be resolved with both a tool and cultural change in administration of the tool.

Thus, whilst the evidence base for this guideline necessarily focuses on the diagnostic accuracy of screening for depression or anxiety, the most important aspect of screening (and psychosocial assessment) from a

clinical perspective is the establishment of trust between a woman and the health professional caring for her, together with the ability of that health professional to respond effectively. At the end of the day, the tool is less important than the conversation and response.

Table B5-1 Consequences of findings from psychometric screening tools administered during the perinatal period

Test result	Possible consequences
True positive	**Benefits:** appropriate referral and management; opportunity for education, reduction of stigma and validation of experience. **Harms:** unnecessary anxiety if no effective treatment available.
False positive	**Benefits:** Opportunity for education on importance of emotional and mental wellbeing and seeking help early. **Harms:** unnecessary anxiety, stigma; 'unnecessary' further consultations and/or tests.
True negative	**Benefits:** reassurance; increased awareness of the importance of emotional and mental wellbeing and normalizing of challenges. **Harms:** none specified
False negative	**Benefits:** increased awareness of the importance of emotional and mental wellbeing and normalizing of challenges. **Harms:** delayed diagnosis and treatment resulting in unnecessary morbidity Delayed

The EWG recognised the importance of sensitivity and specificity as test measures, and emphasized how comparing these measures at different cut-off thresholds is important for clinical interpretation of results. The EWG agreed that the Area Under the Curve (AUC) of the Receiver-Operator Curve (ROC) is a valuable global measure of test performance.

However, although discrimination properties of the depression screening instrument are important, the EWG felt that the most useful test performance measures are those that predict the probability of the condition in an individual. The EWG agreed that the Positive Likelihood Ratio (LR+) and the Negative Likelihood Ratio (LR-) have greater clinical utility than the Positive Predictive Value (PPV) or Negative Predictive Value (NPV). This is because LR+ and LR- are independent of prevalence, whereas PPV and NPV are not.

B5.2 CRITICAL APPRAISAL OF TECHNICAL PERFORMANCE

B5.2.1 Quality assessment of individual studies

Cochrane methods were used by NICE 2015 to assess the quality of the included studies of test accuracy. Specifically, the QUADAS-2 checklist was used to assess the risk of bias and applicability concerns for each included study. These are detailed in Appendix 17 of the NICE 2015 Guideline, and the assessments undertaken by NICE for depression screening are reproduced herein. The QUADAS-2 questions are listed in Appendix **B8.4.1**.

Although details of the risk of bias and applicability judgements conducted by NICE 2015 are included in Appendix 17 of their Technical Report the NICE 2015 Guideline do not present an assessment of the overall quality of each study. Furthermore, the quality assessments of the individual studies within NICE 2015 do not appear to have influenced which studies have been taken through to the meta-analyses, or the development of the NICE 2015 recommendations.

To enable the development of GRADE-style recommendations, the current COPE guideline presents an overall quality for each screening study. These quality assessments are then included within the assessment

of the overall certainty of the evidence (see below). The overall quality of each study has been determined using the following framework:

- **High quality** when all seven sub-domains are assessed as low risk or low concern according to the QUADAS-2 checklist
- **Moderate quality** when one or two sub-domains of the QUADAS-2 checklist are assessed as unclear but no domains are assessed as high risk or high concern, or when only one domain is assessed as high risk or high concern and all other domains are low risk or low concern.
- **Low quality** when two QUADAS-2 sub-domains are assessed as high risk or high concern, and all five other sub-domains are assessed as low risk or low concern.
- **Very Low quality** when four or fewer sub-domains of the QUADAS-2 checklist are rated as low risk or low concern, regardless of the whether the remaining three sub-domains are assessed as high risk or high concern, or are unclear.

B5.2.2 Overall certainty of the evidence by outcome

Once the results across studies are pooled by type of tool, cut-off threshold and type of mental health issue (jn this case, depression), the overall certainty of the evidence was determined by the EWG with reference to:

- The number of studies (k)
- The total number of participants across all studies (N)
- The point estimates and confidence intervals for the pooled results (or individual study results if there is only one study or two or more studies that have not been meta-analysed)
- The overall quality of each study (taking account of risk of bias and applicability related to country and/or setting of the study)
- The generalisability of the study populations to the Guideline context (i.e. community versus psychological sample – see text below for further explanation)

The **overall certainty** for each outcome is then ranked as per the GRADE approach as High (●●●●), Moderate (●●●○), Low (●●○○) or Very Low (●○○○).

Whilst LR+ and LR- are independent of prevalence, they are still influenced by the spectrum of disease within a study population. To determine the generalisability of the included studies to the guideline question it was considered important to identify whether each study recruited a ‘community’ (i.e. a general perinatal population with no known mental health issues) or a ‘psychological’ sample (i.e. women already identified as having mental health symptoms who have been referred for further assessment). These determinations have been undertaken for the current COPE guideline, relying on information extracted in the NICE 2015 guideline and review of individual study abstracts when these were readily available.

B5.3 EVIDENCE OF TECHNICAL PERFORMANCE

B5.3.1 Characteristics of studies in foundation review

The characteristics of the included studies considered key to development of recommendations within the current COPE guideline are presented in Table B5-2. This table lists the 58 studies that were used to derive estimates of sensitivity and specific within the NICE 2015 guideline. The table presents the quality assessments reported by NICE 2015 together with the overall quality ratings determined by the EWG for the current COPE guideline. A number of discrepancies were identified within the NICE 2015 documents (e.g. between Appendix 17 and Tables 11 to 14 of the Guideline). As a general rule, we have given preference to information in Appendix 17 or information taken directly by us from the title and/or abstract

of the included study. Where information for a study was extracted from multiple publications by NICE 2015, we have continued to use the same Study ID and combined presentation of information.

The number of studies (k) that reported on each instrument is as follows: EPDS, 52 studies; PHQ, four (4) studies; K-10, three (3) studies; Whooley questions, two (2) studies. Three studies reported on two instruments.

Table B5-2 Key characteristics of studies using the EPDS, PHQ, Whooley Questions or K-10 to identify perinatal depression

Study ID	Tool(s)	Country; Setting (Population sample [2])	Ref in App 17 NICE 2015	Patient Selection [1]		Index test(s) [1]		Ref. Standard [1]		Flow & Timing [1]	Study Quality [3]
				Risk of bias	Applicability concerns	Risk of bias	Applicability concerns	Risk of bias	Applicability concerns	Risk of bias	
Adewuya 2005	EPDS	Nigeria; Postnatal clinics (Community)	1.1.1	Low	High	Unclear	Low	*Low*	*Low*	Unclear	Very low
Adewuya 2006	EPDS	Nigeria; Antenatal clinic (Community)	1.1.2	High	High	High	High	Low	Low	High	Very low
Agoub 200	EPDS	*Morocco*; Mother-baby unit (Community)	1.1.3	Low	Low	Unclear	Low	Unclear	Low	Unclear	Very low
Alvarado-Esquivel 2006	EPDS	Mexico; Postnatal clinic (Community – low SES)	1.1.4	Low	High	Low	Low	Low	Low	Low	Moderate
Ascaso 2003	EPDS	*Spain*; Antenatal clinic (Community)	1.1.5	*Unclear*	*Unclear*	*Unclear*	*Unclear*	*Unclear*	*Unclear*	*Unclear*	Very low
Aydin 2004	EPDS	*Turkey;* Primary care clinic (Community)	1.1.6	Low	Low	Low	Low	Low	Low	Low	High
Baggaley 2007	K-10	Burkina Faso; NR (Community – selected)	1.1.7	Low	Low	Low	Low	*Unclear*	*Low*	*Unclear*	Moderate
Barnett 1999	EPDS	Australia; NR (Community)	1.1.8	Low	Low	Unclear	High	Unclear	Low	Low	Very low
Beck 2001	EPDS	USA; Childbirth classes (Community)	1.1.9	Low	Low	Low	Low	Low	Low	Unclear	Moderate
Benvenuti 1999	EPDS	Italy; Obstetric clinic (Community)	1.1.10	Low	Low	Unclear	Low	Low	Low	Unclear	Moderate
Bergink 2011	EPDS	Netherlands; Midwife practices (Community)	1.1.11	Low	Low	Unclear	Low	Low	Low	High	Low
Berle 2003	EPDS	Norway; Postnatal visits (Mixed sample)	1.1.12	High	Low	Low	Low	Low	Low	High	Low
Boyce 1993	EPDS	Australia; Postnatal clinics and outpatient psychiatric referrals (Mixed)	1.1.13	High	High	Low	Low	Unclear	Low	Unclear	Very low
Bunevicius 2009	EPDS	Lithuania; *NR (Community*)	1.1.14	Low	Low	Low	Low	Low	Low	Unclear	Moderate
Carpiniello 1997	EPDS	Italy; Obstetrics clinic (Community)	1.1.15	Low	Low	Low	Low	Unclear	Low	Low	Moderate
Chaudron 2010	EPDS	USA; Postnatal child health visits (Community – low SES)	1.1.16	Low	Low	Unclear	Low	Low	Low	High	Low
Chibanda 2010	EPDS	Zimbabwe; Postnatal visits (Community)	1.1.17	Low	Low	Low	Low	Low	Low	Low	High
Clarke 2008	EPDS	Canada; Postnatal and parenting groups (Community)	1.1.18	Low	Low	Unclear	Low	Unclear	Low	Unclear	Very low
Cox 1987	EPDS	UK; Health visitors (Mixed – mostly psych)	1.1.19	High	Low	High	Low	Low	Low	Unclear	Very low
Eberhard-Gran 2001	EPDS	Norway; Child health clinics (Mixed – case control)	1.1.20	High	Low	Unclear	Low	Low	Low	High	Very low

Study ID	Tool(s)	Country; Setting (Population sample [2])	Ref in App 17 NICE 2015	Patient Selection [1]		Index test(s) [1]		Ref. Standard [1]		Flow & Timing [1]	Study Quality [3]
				Risk of bias	Applicability concerns	Risk of bias	Applicability concerns	Risk of bias	Applicability concerns	Risk of bias	
Ekeroma 2012	EPDS	*New Zealand;* Antenatal clinic (Community)	1.1.21	Low	Low	Low	Low	Low	Low	High	Moderate
Felice 2006	EPDS	Malta; Antenatal clinic (*Community*)	1.1.22	Low	Low	Low	Low	Low	Low	Low	High
Fernandes 2011	EPDS K-10	India; Antenatal clinic (*Community*)	1.1.23	Low	High	Unclear	High	Unclear	Low	Low	Very low
Flynn 2011	EPDS PHQ	USA; Psychiatry services (Psychological)	1.1.24	Low	Low	Low	Low	Unclear	*Low*	High	Low
Garcia-Esteve 2003	EPDS	Spain; Postnatal visits (*Community*)	1.1.25	High	Low	Low	Low	Low	Low	High	Very low
Gausia 2007	EPDS	Bangladesh; Child health clinic (Community)	1.1.26	Low	Low	Low	Unclear	Low	Low	Low	Moderate
Ghubash 1997	EPDS	United Arab Emirates; Hospital clinic (Community)	1.1.27	Low	Low	Unclear	Low	Unclear	Low	Unclear	Very low
Gjerdincjen 2009	PHQ WQ	USA; Perinatal clinics (Community)	1.1.28	Low	Low	Unclear	Low	Unclear	Low	Unclear	Very low
Guedeney 1998	EPDS	France; Mother-baby nurse visits (Community)	1.1.29	High	Low	Low	Low	Low	Low	High	Low
Harris 1989	EPDS	UK; Antenatal clinic (Community)	1.1.30	Low	Low	Unclear	Low	Low	Low	Low	Moderate
Jadresic 1995	EPDS	Chile; Antenatal clinic (Community)	1.1.31	*Unclear*	*Unclear*	*Unclear*	*Unclear*	*Unclear*	*Unclear*	*Unclear*	Very low
Kadir 2005	EPDS	Malaysia; Postnatal clinic (Community)	1.1.33	Low	Low	Unclear	Low	Unclear	Low	Unclear	Very low
Lau 2010	EPDS	China; Outpatient clinics (Community)	1.1.34	Low	Low	Unclear	Low	Low	Low	Low	Moderate
Lee 1998	EPDS	Hong Kong; Outpatient clinics (Community)	1.1.32	Low	Low	Low	Low	Low	Low	High	Moderate
Leonardou 2009	EPDS	Greece; Maternity clinic (Community)	1.1.35	Low	Low	Unclear	Low	Low	Low	Unclear	Moderate
Leverton 2000	EPDS	UK; Antenatal clinic (Community)	1.1.36	High	Low	Low	Low	Low	Low	High	Low
Mahmud 2003	EPDS	Malaysia; Perinatal clinic (Community)	1.1.37	Low	Low	Low	Low	Low	Low	Low	High
Mann 2012	WQ	UK; Maternity unit (Community)	1.1.38	Low	Low	Low	Low	Low	Low	High	Moderate
Mazhari 2007	EPDS	Iran Child health visits (Community)	1.1.40	High	Low	Low	High	Low	Low	High	Very low
Milgrom 2005A	EPDS	Australia; Mother & baby clinics (Community)	1.1.41	High	Low	Low	Low	Unclear	Low	High	Very low
Murray 1990B	EPDS	UK; Antenatal clinic (Community)	1.1.42	Low	Low	Unclear	Low	Low	Low	Unclear	Moderate
Muzik 2000	EPDS	Austria; NR (Psychological (women at risk of MDD)	1.1.43	High	Low	Unclear	Unclear	Unclear	Low	High	Very low

Study ID	Tool(s)	Country; Setting (Population sample [2])	Ref in App 17 NICE 2015	Patient Selection [1]		Index test(s) [1]		Ref. Standard [1]		Flow & Timing [1]	Study Quality [3]
				Risk of bias	Applicability concerns	Risk of bias	Applicability concerns	Risk of bias	Applicability concerns	Risk of bias	
Phillips 2009	EPDS	Australia; Parent-infant unit (Community – at risk)	1.1.44	Low	Low	Unclear	Low	Low	Low	High	Low
Pitanupong 2007	EPDS	Thailand; Hospital clinic (Community)	1.1.45	Low	Low	Unclear	Low	Low	Low	High	Low
Regmi 2002	EPDS	Nepal; Postnatal clinic (Community)	1.1.46	High	Low	Unclear	Unclear	Unclear	Unclear	High	Very low
Rubertsson 2011	EPDS	Sweden; Antenatal clinics (Community)	1.1.47	Low	Low	Unclear	Low	Unclear	Low	High	Very low
Santos 2007	EPDS	Brazil NR (Community?)	1.1.48	High	Low	Unclear	Low	Low	Low	High	Very low
Sidebottom 2012	PHQ	USA; Community clinics (Community)	1.1.49	Low	Low	Low	Low	Unclear	Unclear	High	Very low
Smith 2010	PHQ	USA; Prenatal clinics (Mixed)	1.1.50	High	Low	Low	Low	Unclear	Low	High	Very low
Spies 2009	K-10	South Africa; Midwife clinics (Community)	1.1.51	Low	Low	Unclear	High	Unclear	Low	Unclear	Very low
Tandon 2012	EPDS	USA; home visits (Community – low SES)	1.1.52	Low	Low	Low	High	High	Low	Low	Low
Teng 2005	EPDS	Taiwan; Maternity wards (Community)	1.1.53	Low	Low	Unclear	Low	Low	Low	High	Moderate
Thiagayson 2013	EPDS	Singapore; Maternity wards (Community – high risk pregnancy)	1.1.54	Low	*Low*	Unclear	Low	Low	Low	Low	Moderate
Toreki 2013	EPDS	Hungary; Antenatal clinics (Community)	1.1.55	Low	Low	Low	Low	Low	Low	Low	High
Uwakwe 2003	EPDS	Nigeria; Maternity wards and postnatal clinics (Community)	1.1.57	Low	Low	Low	High	Unclear	Low	High	Very low
Werrett 2006	EPDS	UK (Punjabi); Postnatal clinics (Community)	1.1.58	Low	Low	Low	Low	Low	Low	Low	High
Wickberg 1996	EPDS	Sweden; Child health clinics (Community)	1.1.59	High	Low	Low	Low	Low	Low	High	Low
Yoshida 2001	EPDS	UK/Japan; Antenatal classes and advertisement (Community)	1.1.60	Low	Low	Low	Unclear	Unclear	Low	Unclear	Very low

Abbreviations: App, appendix; EPDS, Edinburgh Postnatal Depression Scale; K-10, Kessler-10; NR, not reported; PHQ, Primary Health Questionnaire; SES, socioeconomic status; UK, United Kingdom; USA, United States of America; WQ, Whooley questions

[1] Compiled from assessments presented in the methodology checklists of Appendix 17 and Appendix 18 of the NICE 2015 Guideline, limited to studies subsequently included in the meta-analyses presented in the NICE 2015 Guideline.

[2] Compiled from assessments presented in the study characteristics tables in Tables 11-14 and Appendix 18 of the NICE 2015 Guideline, titles/abstracts provided additional information where needed;

[3] Determined by the Expert Working Group for the current Guideline

Notes: Where obvious discrepancies were identified within the NICE documents the current authors made a judgement regarding the information most likely to be correct, with priority given to title and abstract of the source study (if readily available), Appendix 17, then Table 11-14 (highlighted *in italics*). Full cross-checking of all characteristics of all studies included by the NICE 2015 authors was not undertaken. Where information from a study has been extracted from multiple publications the current authors have used the same Study ID as the NICE 2015 Guideline.

B5.3.2 Evidence summaries based on studies in foundation review

The tables below present evidence as extracted by NICE 2015 for TP, FP, FN, TN, sensitivity and specificity. Although AUC was defined as a critical outcome by the EWG for the current COPE guideline, this measure was not extracted from individual studies by NICE. The study quality presented in the tables below are based on the overall ratings undertaken for the current guideline.

A total of eight (8) separate tables are presented for each instrument and each perinatal time period: i.e. an antenatal table and a postnatal table for each of the EPDS, PHQ, Whooley Questions, and K-10. Results are reported by condition and cut-off threshold.

B5.3.2.1 EPDS

Table B5-3 Evidence summary table for the EPDS for detection of depression in antenatal women

Condition; cut-off	TP [1]	FP [1]	FN [1]	TN [1]	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC [2]	Study Quality [3]
Mixed depression; 9/10								
Adewuya 2006	13	6	2	65	0.87 (0.60, 0.98)	0.92 (0.83, 0.97)	NR	Very low
Felice 2006	29	38	3	153	0.91 (0.75, 0.98)	0.80 (0.74, 0,86)	NR	High
Thiagayson 2013	25	38	11	126	0.69 (0.52, 0.84)	0.77 (0.70, 0.83)	NR	Moderate
Toreki 2013	11	6	11	191	0.50 (0.28, 0.72)	0.97 (0.93, 0.99)	NR	High
Mixed depression; 12/13								
Bunevicius 2009	12	9	2	207	0.86 (0.57, 0.98)	0.96 (0.92, 0.98)	NR	Moderate
Felice 2006	25	20	7	171	0.78 (0.60, 0.91)	0.90 (0.84, 0.93)	NR	High
Murray 1990B	9	9	5	77	0.64 (0.35, 0.87)	0.90 (0.81, 0.95)	NR	Moderate
Toreki 2013	4	1	18	196	0.18 (0.05, 0.40)	0.99 (0.97, 1.00)	NR	High
Mixed depression; 14/15								
Felice 2006	21	6	11	185	0.66 (0.47, 0.81)	0.97 (0.93, 0.99)	NR	High
Murray 1990B	8	2	6	84	0.57 (0.29, 0.82)	0.98 (0.92, 1.00)	NR	Moderate
Toreki 2013	3	0	19	197	0.14 (0.03, 0.35)	1.00 (0.98, 1.00)	NR	High
Major depression; 9/10								
Bergink 2011	41	40	6	758	0.87 (0.74, 0.95)	0.95 (0.93, 0.96)	NR	Low
Fernandes 2011	28	86	0	80	1.00 (0.88, 1.00)	0.48 (0.40, 0.56)	NR	Very low
Toreki 2013	3	15	4	197	0.43 (0.10, 0.82)	0.93 (0.89, 0.96)	NR	High
Major depression; 12/13								
Adewuya 2006	9	3	0	74	1.00 (0.66, 1.00)	0.96 (0.89, 0.99)	NR	Very low
Bergink 2011	11	11	1	207	0.92 (0.62, 1.00)	0.95 (0.91, 0.97)	NR	Low
Fernandes 2011	28	25	0	141	1.00 (0.88, 1.00)	0.85 (0.79, 0.90)	NR	Very low
Flynn 2011	46	3	12	8	0.79 (0.67, 0.89)	0.73 (0.39, 0.94)	NR	Low
Murray 1990B	6	12	0	82	1.00 (0.54, 1.00)	0.87 (0.79, 0.93)	NR	Moderate
Rubertsson 2011	7	7	2	105	0.78 (0.40, 0.97)	0.94 (0.88, 0.97)	NR	Very low
Thiagayson 2013	16	46	6	132	0.73 (0.50, 0.89)	0.74 (0.67, 0.80)	NR	Moderate
Toreki 2013	2	2	5	210	0.29 (0.04, 0.71)	0.99 (0.97, 1.00)	NR	High
Major depression; 14/15								
Adewuya 2006	7	1	2	76	0.78 (0.40, 0.97)	0.99 (0.93, 1.00)	NR	Very low
Fernandes 2011	21	12	7	154	0.75 (0.55, 0.89)	0.93 (0.88, 0.96)	NR	Very low
Murray 1990B	6	4	0	90	1.00 (0.54, 1.00)	0.96 (0.89, 0.99)	NR	Moderate
Toreki 2013	2	1	5	211	0.29 (0.04, 0.71)	1.00 (0.97, 1.00)	NR	High

Abbreviations: AUC, area under the receiver-operator curve; CI, confidence interval; FN, false negative; FP, false positive; NR, not reported; TN, true negative; TP, true positive.

[1] Reproduced from corresponding forest plots in Appendix 19 of the NICE 2015 Guideline

[2] As reported in the NICE 2015 Guideline or Appendices

[3] Applied from Table B5-2 in this document.

Table B5-4 Evidence summary table for the EPDS for detection of depression in postnatal women

Condition; cut-off	TP [1]	FP [1]	FN [1]	TN [1]	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC [2]	Study Quality [3]
Mixed depression; 9/10								
Adewuya 2005	113	7	15	756	0.88 (0.81, 0.93)	0.99 (0.98, 1.00)	NR	Very low
Agoub 2005	27	14	0	103	1.00 (0.87, 1.00)	0.88 (0.81, 0.93)	NR	Very low
Alvarado-Esquivel 2006	3	6	1	39	0.75 (0.19, 0.99)	0.87 (0.73, 0.95)	NR	Moderate
Ascaso 2003	30	87	4	213	0.88 (0.73, 0.97)	0.71 (0.66, 0.76)	NR	Very low
Aydin 2004	47	155	2	137	0.96 (0.86, 1.00)	0.47 (0.41, 0.53)	NR	High
Beck 2001	27	15	19	89	0.59 (0.43, 0.73)	0.86 (0.77, 0.92)	NR	Moderate
Benvenuti 1999	15	10	3	85	0.83 (0.59, 0.96)	0.89 (0.81, 0.95)	NR	Moderate
Berle 2003	37	17	4	42	0.90 (0.77, 0.97)	0.71 (0.58, 0.82)	NR	Low
Carpiniello 1997	9	9	0	43	1.00 (0.66, 1.00)	0.83 (0.70, 0.92)	NR	Moderate
Chaudron 2010	68	6	43	81	0.61 (0.52, 0.70)	0.93 (0.86, 0.97)	NR	Low
Felice 2006	15	17	3	188	0.83 (0.59, 0.96)	0.92 (0.87, 0.95)	NR	High
Garcia-Esteve 2003	89	72	11	951	0.89 (0.81, 0.94)	0.93 (0.91, 0.94)	NR	Very low
Gausia 2007	8	12	1	79	0.89 (0.52, 1.00)	0.87 (0.78, 0.93)	NR	Moderate
Ghubash 1997	12	13	1	69	0.92 (0.64, 1.00)	0.84 (0.74, 0.91)	NR	Very low
Guedeney 1998	38	9	7	33	0.84 (0.71, 0.94)	0.79 (0.63, 0.90)	NR	Low
Jadresic 1995	11	19	0	78	1.00 (0.72, 1.00)	0.80 (0.71, 0.88)	NR	Very low
Kadir 2005	8	4	3	37	0.73 (0.39, 0.94)	0.90 (0.77, 0.97)	NR	Very low
Lau 2010	12	62	4	264	0.75 (0.48, 0.93)	0.81 (0.76, 0.85)	NR	Moderate
Lee 1998	14	18	3	110	0.82 (0.57, 0.96)	0.86 (0.79, 0.91)	NR	Moderate
Leonardou 2009	10	10	0	61	1.00 (0.69, 1.00)	0.86 (0.76, 0.93)	NR	Moderate
Leverton 2000	9	30	1	159	0.90 (0.55, 1.00)	0.84 (0.78, 0.89)	NR	Low
Mahmud 2003	9	4	0	51	1.00 (0.66, 1.00)	0.93 (0.82, 0.98)	NR	High
Mazhari 2007	62	24	7	107	0.90 (0.80, 0.96)	0.82 (0.74, 0.88)	NR	Very low
Pitanupong 2007	23	31	15	282	0.61 (0.43, 0.76)	0.90 (0.86, 0.93)	NR	Low
Santos 2007	96	123	9	150	0.91 (0.84, 0.96)	0.55 (0.49, 0.61)	NR	Very low
Tandon 2012	27	12	5	51	0.84 (0.67, 0.95)	0.81 (0.69, 0.90)	NR	Low
Uwakwe 2003	18	6	6	195	0.75 (0.53, 0.90)	0.97 (0.94, 0.99)	NR	Very low
Werrett 2006	7	6	0	10	1.00 (0.59, 1.00)	0.63 (0.35, 0.85)	NR	High
Yoshida 2001	12	4	3	69	0.80 (0.52, 0.96)	0.95 (0.87, 0.98)	NR	Very low
Mixed depression; 12/13								
Adewuya 2005	63	0	65	748	0.49 (0.40, 0.58)	1.00 (1.00, 1.00)	NR	Very low
Agoub 2005	25	5	2	112	0.93 (0.76, 0.99)	0.96 (0.90, 0.99)	NR	Very low
Alvarado-Esquivel 2006	2	2	2	43	0.50 (0.07, 0.93)	0.96 (0.85, 0.99)	NR	Moderate
Ascaso 2003	21	24	13	276	0.62 (0.44, 0.78)	0.92 (0.88, 0.95)	NR	Very low
Aydin 2004	37	83	12	209	0.76 (0.61, 0.87)	0.72 (0.66, 0.77)	NR	High
Berle 2003	20	3	21	56	0.49 (0.33, 0.65)	0.95 (0.86, 0.99)	NR	Low
Carpiniello 1997	6	0	3	52	0.67 (0.30, 0.93)	1.00 (0.93, 1.00)	NR	Moderate
Clarke 2008	14	10	3	76	0.82 (0.57, 0.96)	0.88 (0.80, 0.94)	NR	Very low
Cox 1987	30	11	5	38	0.86 (0.70, 0.95)	0.78 (0.63, 0.88)	NR	Very low
Felice 2006	14	4	4	201	0.78 (0.52, 0.94)	0.98 (0.95, 0.99)	NR	High
Garcia-Esteve 2003	62	20	38	1003	0.62 (0.52, 0.72)	0.98 (0.97, 0.99)	NR	Very low
Gausia 2007	6	6	3	85	0.67 (0.30, 0.93)	0.93 (0.86, 0.98)	NR	Moderate

Condition; cut-off	TP [1]	FP [1]	FN [1]	TN [1]	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC [2]	Study Quality [3]
Ghubash 1997	9	8	4	74	0.69 (0.39, 0.91)	0.90 (0.82, 0.96)	NR	Very low
Guedeney 1998	27	1	18	41	0.60 (0.44, 0.74)	0.98 (0.87, 1.00)	NR	Low
Jadresic 1995	6	6	5	91	0.55 (0.23, 0.83)	0.94 (0.87, 0.98)	NR	Very low
Kadir 2005	6	0	5	41	0.55 (0.23, 0.83)	1.00 (0.91, 1.00)	NR	Very low
Lau 2010	14	82	2	245	0.88 (0.62, 0.98)	0.75 (0.70, 0.80)	NR	Moderate
Lee 1998	7	6	10	122	0.41 (0.18, 0.67)	0.95 (0.90, 0.98)	NR	Moderate
Leonardou 2009	8	1	2	70	0.80 (0.44, 0.97)	0.99 (0.92, 1.00)	NR	Moderate
Leverton 2000	7	13	3	176	0.70 (0.35, 0.93)	0.93 (0.89, 0.96)	NR	Low
Mahmud 2003	7	1	2	54	0.78 (0.40, 0.97)	0.98 (0.90, 1.00)	NR	High
Mazhari 2007	51	9	18	122	0.74 (0.62, 0.84)	0.93 (0.87, 0.97)	NR	Very low
Milgrom 2005A	222	24	38	60	0.85 (0.80, 0.89)	0.71 (0.61, 0.81)	NR	Very low
Pitanupong 2007	13	9	25	304	0.34 (0.20, 0.51)	0.97 (0.95, 0.99)	NR	Low
Santos 2007	63	33	42	240	0.60 (0.50, 0.69)	0.88 (0.83, 0.92)	NR	Very low
Teng 2005	23	27	1	152	0.96 (0.79, 1.00)	0.85 (0.79, 0.90)	NR	Moderate
Uwakwe 2003	12	4	12	197	0.50 (0.29, 0.71)	0.98 (0.95, 0.99)	NR	Very low
Werrett 2006	5	3	2	13	0.71 (0.29, 0.96)	0.81 (0.54, 0.96)	NR	High
Yoshida 2001	8	1	7	72	0.53 (0.27, 0.79)	0.99 (0.93, 1.00)	NR	Very low
Major depression; 9/10								
Barnett 1999(A)	7	17	2	69	0.78 (0.40, 0.97)	0.80 (0.70, 0.88)	NR	Very low
Barnett 1999(AC)	6	16	1	82	0.86 (0.42, 1.00)	0.84 (0.75, 0.90)	NR	Very low
Barnett 1999(V)	5	33	0	75	1.00 (0.48, 1.00)	0.69 (0.60, 0.78)	NR	Very low
Berle 2003	26	28	1	45	0.96 (0.81, 1.00)	0.62 (0.50, 0.73)	NR	Low
Boyce 1993	9	10	0	84	1.00 (0.66, 1.00)	0.89 (0.81, 0.95)	NR	Very low
Chibanda 2010	64	39	0	107	1.00 (0.94, 1.00)	0.73 (0.65, 0.80)	NR	High
Eberhard-Gran 2001	9	6	0	41	1.00 (0.66, 1.00)	0.87 (0.74, 0.95)	NR	Very low
Ekeroma 2012(s)	13	19	2	51	0.87 (0.60, 0.98)	0.73 (0.61, 0.83)	NR	Moderate
Ekeroma 2012(t)	10	15	4	56	0.71 (0.42, 0.92)	0.79 (0.68, 0.88)	NR	Moderate
Garcia-Esteve 2003	36	120	0	967	1.00 (0.90, 1.00)	0.89 (0.87, 0.91)	NR	Very low
Kadir 2005	11	7	0	34	1.00 (0.72, 1.00)	0.83 (0.68, 0.93)	NR	Very low
Mazhari 2007	42	44	1	113	0.98 (0.88, 1.00)	0.72 (0.64, 0.79)	NR	Very low
Muzik 2000	7	10	2	31	0.78 (0.40, 0.97)	0.76 (0.60, 0.88)	NR	Very low
Major depression; 12/13								
Adewuya 2005	48	17	0	811	1.00 (0.93, 1.00)	0.98 (0.97, 0.99)	NR	Very low
Barnett 1999(A)	5	8	4	78	0.56 (0.21, 0.86)	0.91 (0.82, 0.96)	NR	Very low
Barnett 1999(AC)	4	7	3	91	0.57 (0.18, 0.90)	0.93 (0.86, 0.97)	NR	Very low
Barnett 1999(V)	5	12	0	96	1.00 (0.48, 1.00)	0.89 (0.81, 0.94)	NR	Very low
Beck 2001	14	1	4	131	0.78 (0.52, 0.94)	0.99 (0.96, 1.00)	NR	Moderate
Benvenuti 1999	10	1	8	94	0.56 (0.31, 0.78)	0.99 (0.94, 1.00)	NR	Moderate
Berle 2003	15	8	12	65	0.56 (0.35, 0.75)	0.89 (0.80, 0.95)	NR	Low
Boyce 1993	9	4	0	90	1.00 (0.66, 1.00)	0.96 (0.89, 0.99)	NR	Very low
Chaudron 2010	40	11	33	114	0.55 (0.43, 0.66)	0.91 (0.85, 0.96)	NR	Low
Chibanda 2010	52	18	12	128	0.81 (0.70, 0.90)	0.88 (0.81, 0.93)	NR	High
Ekeroma 2012(s)	10	10	5	60	0.67 (0.38, 0.88)	0.86 (0.75, 0.93)	NR	Moderate
Ekeroma 2012(t)	8	6	6	65	0.57 (0.29, 0.82)	0.92 (0.83, 0.97)	NR	Moderate

Condition; cut-off	TP [1]	FP [1]	FN [1]	TN [1]	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC [2]	Study Quality [3]
Flynn 2011	70	11	6	12	0.92 (0.84, 0.97)	0.52 (0.31, 0.73)	NR	Low
Garcia-Esteve 2003	31	54	5	1033	0.86 (0.71, 0.95)	0.95 (0.94, 0.96)	NR	Very low
Harris 1989	21	7	1	97	0.95 (0.77, 1.00)	0.93 (0.87, 0.97)	NR	Moderate
Kadir 2005	3	3	1	45	0.75 (0.19, 0.99)	0.94 (0.83, 0.99)	NR	Very low
Mazhari 2007	41	19	2	138	0.95 (0.84, 0.99)	0.88 (0.82, 0.93)	NR	Very low
Muzik 2000	6	2	3	39	0.67 (0.30, 0.93)	0.95 (0.83, 0.99)	NR	Very low
Phillips 2009	30	23	12	100	0.71 (0.55, 0.84)	0.81 (0.73, 0.88)	NR	Low
Regmi 2002	5	7	0	88	1.00 (0.48, 1.00)	0.93 (0.85, 0.97)	NR	Very low
Tandon 2012	22	3	5	65	0.81 (0.62, 0.94)	0.96 (0.88, 0.99)	NR	Low
Wickberg 1996	48	27	8	45	0.86 (0.74, 0.94)	0.63 (0.50, 0.74)	NR	Low

Abbreviations: AUC, area under the receiver-operator curve; CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; FN, false negative; FP, false positive; NR, not reported; TN, true negative; TP, true positive.

[1] Reproduced from corresponding forest plots in Appendix 19 of the NICE 2015 Guideline

[2] As reported in the NICE 2015 Guideline or Appendices

[3] Applied from Table B5-2 in this document.

B5.3.2.2 PHQ

Table B5-5 Evidence summary table for the PHQ for detection of depression in antenatal women

Condition; cut-off	TP [1]	FP [1]	FN [1]	TN [1]	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC [2]	Study Quality [3]
PHQ-9 (simple) – Mixed depression; 10								
Sidebottom 2012	59	80	20	586	0.75 (0.64, 0.84)	0.88 (0.85, 0.90)	NR	Very low
PHQ-9 (simple) – Major depression; 10								
Flynn 2011	43	3	15	8	0.74 (0.61, 0.85)	0.73 (0.39, 0.94)	NR	Low
Sidebottom 2012	23	115	4	603	0.85 (0.66, 0.96)	0.84 (0.81, 0.87)	NR	Very low

Abbreviations: AUC, area under the receiver-operator curve; CI, confidence interval; FN, false negative; FP, false positive; NR, not reported; PHQ, Patient Health Questionnaire; TN, true negative; TP, true positive.

[1] Reproduced from corresponding forest plots in Appendix 19 of the NICE 2015 Guideline

[2] As reported in the NICE 2015 Guideline or Appendices

[3] Applied from Table B5-2 in this document.

Table B5-6 Evidence summary table for the PHQ for detection of depression in postnatal women

Condition; cut-off	TP [1]	FP [1]	FN [1]	TN [1]	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC [2]	Study Quality [3]
PHQ-9 (simple) – Major depression; 10								
Flynn 2011	68	8	8	15	0.89 (0.80, 0.95)	0.65 (0.43, 0.84)	NR	Low
Gjerdincjen 2009	37	74	8	387	0.82 (0.68, 0.92)	0.84 (0.80, 0.87)	NR	Very low
PHQ-9 (complex) – Major depression; 10								
Gjerdincjen 2009	30	37	15	424	0.67 (0.51, 0.80)	0.92 (0.89, 0.94)	NR	Very low
PHQ-2 – Major depression; 3								
Gjerdincjen 2009	38	97	7	364	0.84 (0.71, 0.94)	0.79 (0.75, 0.83)	NR	Very low
Smith 2010	10	82	3	118	0.77 (0.46, 0.95)	0.59 (0.52, 0.66)	NR	Very low

Condition; cut-off	TP [1]	FP [1]	FN [1]	TN [1]	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC [2]	Study Quality [3]
PHQ-2 – Major depression; 4								
Smith 2010	8	42	5	158	0.62 (0.32, 0.86)	0.79 (0.73, 0.84)	NR	Very low
PHQ-8 – Major depression; 10								
Smith 2010	10	76	3	124	0.77 (0.46, 0.95)	0.62 (0.55, 0.69)	NR	Very low

Abbreviations: AUC, area under the receiver-operator curve; CI, confidence interval; FN, false negative; FP, false positive; NR, not reported; PHQ, Patient Health Questionnaire; TN, true negative; TP, true positive.

[1] Reproduced from corresponding forest plots in Appendix 19 of the NICE 2015 Guideline

[2] As reported in the NICE 2015 Guideline or Appendices

[3] Applied from Table B5-2 in this document.

B5.3.2.3 Whooley questions

Table B5-7 Evidence summary table for the 'Whooley questions' for detection of depression in antenatal women

Condition; cut-off	TP [1]	FP [1]	FN [1]	TN [1]	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC [2]	Study Quality [3]
Whooley questions – Mixed depression								
Mann 2012	17	35	0	74	1.00 (0.80, 1.00)	0.68 (0.58, 0.77)	NR	Moderate
Whooley questions (+ help qn) – Mixed depression								
Mann 2012	10	3	7	32	0.59 (0.33, 0.82)	0.91 (0.77, 0.98)	NR	Moderate

Abbreviations: AUC, area under the receiver-operator curve; CI, confidence interval; FN, false negative; FP, false positive; NR, not reported; TN, true negative; TP, true positive.

[1] Reproduced from corresponding forest plots in Appendix 19 of the NICE 2015 Guideline

[2] As reported in the NICE 2015 Guideline or Appendices

[3] Applied from Table B5-2 in this document.

Table B5-8 Evidence summary table for the 'Whooley questions' for detection of depression in postnatal women

Condition; cut-off	TP [1]	FP [1]	FN [1]	TN [1]	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC [2]	Study Quality [3]
Whooley questions – Mixed depression								
Mann 2012	18	27	0	49	1.00 (0.81, 1.00)	0.64 (0.53, 0.75)	NR	Moderate
Whooley questions (+ help qn) – Mixed depression								
Mann 2012	7	0	11	27	0.39 (0.17, 0.64)	1.00 (0.87, 1.00)	NR	Moderate
Whooley questions – Major depression								
Gjerdincjen 2009	45	258	0	203	1.00 (0.92, 1.00)	0.44 (0.39, 0.49)	NR	Very low

Abbreviations: AUC, area under the receiver-operator curve; CI, confidence interval; FN, false negative; FP, false positive; NR, not reported; TN, true negative; TP, true positive.

[1] Reproduced from corresponding forest plots in Appendix 19 of the NICE 2015 Guideline

[2] As reported in the NICE 2015 Guideline or Appendices

[3] Applied from Table B5-2 in this document.

B5.3.2.4 K-10

Table B5-9 Evidence summary table for the K-10 for detection of depression in antenatal women

Condition; cut-off	TP [1]	FP [1]	FN [1]	TN [1]	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC [2]	Study Quality [3]
Major depression; 6								
Fernandes 2011	28	32	0	134	1.00 (0.88, 1.00)	0.81 (0.74, 0.86)	NR	Very low
Spies 2009	12	52	4	61	0.75 (0.48, 0.93)	0.54 (0.44, 0.63)	NR	Very low

Abbreviations: AUC, area under the receiver-operator curve; CI, confidence interval; FN, false negative; FP, false positive; NR, not reported; TN, true negative; TP, true positive.

[1] Reproduced from corresponding forest plots in Appendix 19 of the NICE 2015 Guideline

[2] As reported in the NICE 2015 Guideline or Appendices

[3] Applied from Table B5-2 in this document.

Table B5-10 Evidence summary table for the K-10 for detection of depression in postnatal women

Condition; cut-off	TP [1]	FP [1]	FN [1]	TN [1]	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC [2]	Study Quality [3]
Mixed depression; 6								
Baggaley 2007	23	20	4	14	0.85 (0.66, 0.96)	0.41 (0.25, 0.59)	NR	Moderate

Abbreviations: AUC, area under the receiver-operator curve; CI, confidence interval; FN, false negative; FP, false positive; K-10, Kessler 10 item questionnaire; NR, not reported; TN, true negative; TP, true positive.

[1] Reproduced from corresponding forest plots in Appendix 19 of the NICE 2015 Guideline

[2] As reported in the NICE 2015 Guideline or Appendices

[3] Applied from Table B5-2 in this document.

B5.3.3 Summary of findings regarding technical performance

The Summary of Findings (SOF) tables present a summary of the important and critical outcomes, as determined by the EWG. The pooled sensitivity and specificity measures have been extracted from NICE 2015. Where results were not pooled by NICE 2015, the unpooled sensitivity and specificity results are presented. The LR+ and LR- values have been calculated by the current authors, based on the corresponding pooled or unpooled sensitivity and specificity results. The EWG defined the 'goodness' of sensitivity and specificity as follows: >0.90, high; 0.70 – 0.90, moderate; <0.70, low (keeping in mind that <0.5 is non-discriminating).

Results are grouped together below according to the population in the studies: antenatal women only, or postnatal women only. No included studies of depression screening were conducted in a mixed population of antenatal and postnatal women.

Where Receiver-Operator curves (ROC) are available from the NICE 2015 Guideline, these are reproduced following the corresponding SOF tables.

B5.3.3.1 Depression screening in the antenatal period

Table B5-11 Summary of Findings table for the EPDS and identification of depression in the antenatal period (based on NICE 2015)

Tool; Condition; Cut-off	No of studies[1] (participants)	Important outcomes		Critical outcomes			Overall certainty
		Pooled sensitivity (95% CI) [1]	Pooled specificity (95% CI) [1]	Positive Likelihood Ratio [2]	Negative Likelihood Ratio [2]	Pooled AUC	
EPDS; minor depression[4]; 9/10	4 (728)	0.74 (0.65, 0.82)	0.86 (0.83, 0.89)	5.29	0.30	NR [3]	●●●○ Moderate[5]
EPDS; mixed depression[4]; 12/13	4 (722)	0.61 (0.5, 0.72)	0.94 (0.92, 0.96)	10.17	0.41	NR [3]	●●●○ Moderate[5]
EPDS; mixed depression[4]; 14/15	3 (542)	0.47 (0.35, 0.60)	0.98 (0.97, 0.99)	23.50	0.54	NR [3]	●●●○ Moderate[5]
EPDS; major depression[4]; 9/10	3 (1,258)	0.88 (0.89, 0.94)	0.88 (0.86, 0.90)	7.33	0.14	NR [3]	●●●● High[6]
EPDS; major depression[4]; 12/13	8 (1,219)	0.83 (0.76, 0.88)	0.90 (0.88, 0.92)	8.30	0.19	NR [3]	●●●● High[6]
EPDS; major depression[4]; 14/15	4 (599)	0.72 (0.58, 0.84)	0.97 (0.95, 0.98)	24.00	0.29	NR [3]	●●●○ Moderate[5]

Evidence statements:

A score of 13 or more on the EPDS has moderate sensitivity and high specificity for detecting possible major depression in pregnant women (high certainty evidence).

A score of 10 or above on the EPDS has moderate sensitivity and moderate specificity for detecting possible depression in pregnant women (moderate certainty evidence).

Footnotes:

[1] Reproduced from Table 15 of the NICE 2015 Guideline.

[2] Calculated from the pooled sensitivity and specificity using the following formulas: LR+ = sensitivity/(1-specificity); LR- = (1-sensitivity)/specificity.

[3] Pooled AUC not reported but ROC included herein.

[4] Definitions of 'minor', 'major' and 'mixed' depression as per NICE 2015.

[5] Multiple studies with a range of quality.

[6] Multiple studies with a range of quality from a large total sample (greater than 1,000).

Abbreviations: AUC, area under the curve; CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; NICE, National Institute of Health and Care Excellence; NR, not reported; LR, likelihood ratio; ROC, receiver operating characteristics.

Table B5-12 Summary of Findings table for the PHQ and identification of depression in the antenatal period (based on NICE 2015)

Tool; Condition; Cut-off	No of studies[1] (participants)	Important outcomes		Critical outcomes			Overall certainty
		Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	Positive Likelihood Ratio [2]	Negative Likelihood Ratio [2]	Pooled AUC	
PHQ-9 (simple[3]); mixed depression[4]; 9/10	1 (745)	0.75 (0.64-0.84)	0.88 (0.85, 0.90)	6.25	0.28	NR	●○○○ Very low[5]
PHQ-9 (simple[4]); major depression[4]; 9/10	2 (814)	0.74 (0.61, 0.85) 0.85 (0.66, 0.96)	0.73 (0.38, 0.94) 0.84 (0.81, 0.87)	2.74 5.31	0.36 0.18	NR	●●○○ Low[6]
Evidence statement: *It is uncertain if the PHQ has adequate sensitivity or specificity to detect possible depressive disorders in pregnant women (very low to low certainty evidence)*							
Footnotes: [1] Reproduced from Table 17 of the NICE 2015 Guideline [2] Calculated from the pooled sensitivity and specificity using the following formulas: LR+ = sensitivity/(1-specificity); LR- = (1-sensitivity)/specificity [3] Simple scoring as reported by NICE 2015: result is positive if sum of numbered responses is ≥10. [4] Definitions of 'minor', 'major' and 'mixed' depression as per NICE 2015 [5] Single study of very low quality [6] Two studies of low to very low quality							

Abbreviations: AUC, area under the curve; CI, confidence interval; NICE, National Institute of Health and Care Excellence; NR, not reported; PHQ, Patient Health Questionnaire; LR, likelihood ratio.

Table B5-13 Summary of Findings table for the 'Whooley questions' and identification of depression in the antenatal period (based on NICE 2015)

Tool; Condition; Cut-off	No of studies[1] (participants)	Important outcomes		Critical outcomes			Overall certainty
		Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	Positive Likelihood Ratio [2]	Negative Likelihood Ratio [2]	Pooled AUC	
Whooley questions; mixed depression[3]	1 (126)	1.00 (0.80, 1.00)	0.68 (0.58, 0.77)	3.09	0.01	NR	●●○○ Low
Whooley questions (+ help question); mixed depression[3]	1 (52)	0.59 (0.33, 0.82)	0.91 (0.77-0.98)	6.56	0.45	NR	●●○○ Low
Evidence statement: *It is uncertain if the 'Whooley questions' have adequate sensitivity or specificity to detect possible minor or major depression in pregnant (low certainty evidence).*							
Footnotes: [1] Reproduced from Table 18 of the NICE 2015 Guideline. [2] Calculated from the pooled sensitivity and specificity using the following formulas: LR+ = sensitivity/(1-specificity); LR- = (1-sensitivity)/specificity, with rounding of values of 1.00 to 0.99 to avoid Div/0 error. [3] Definitions of 'minor', 'major' and 'mixed' depression as per NICE 2015. [5] Single study of moderate quality.							

Abbreviations: AUC, area under the curve; CI, confidence interval; LR, likelihood ratio; NICE, National Institute of Health and Care Excellence.

Table B5-14, Summary of Findings table for the K-10 and identification of depression in the antenatal period (based on NICE 2015)

Tool; Condition; Cut-off	No of studies[1] (participants)	Important outcomes		Critical outcomes			Overall certainty
		Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	Positive Likelihood Ratio [2]	Negative Likelihood Ratio [2]	Pooled AUC	
K-10; major depression[3]; 6	2 (323)	1.00 (0.88, 1.00) 0.75 (0.48, 0.93)	0.81 (0.74, 0.86) 0.54 (0.44, 0.63)	5.21 1.63	0.01 0.46	NR	●●○○ Low[4]
Evidence statement: *It is uncertain if the K-10 has adequate sensitivity or specificity to detect possible major depression in pregnant women (low certainty evidence).*							
Footnotes: [1] Reproduced from Table 19 of the NICE 2015 Guideline. [2] Calculated from the pooled sensitivity and specificity using the following formulas: LR+ = sensitivity/(1-specificity); LR- = (1-sensitivity)/specificity, with rounding of values of 1.00 to 0.99 to avoid Div/0 error. [3] Definitions of 'minor', 'major' and 'mixed' depression as per NICE 2015. [4] Two studies of very low quality.							

Abbreviations: CI, confidence interval; K-10, Kessler 10 item questionnaire; LR, likelihood ratio; NICE, National Institute of Health and Care Excellence; NR, not reported.

B5.3.3.2 Depression screening in the postnatal period

Table B5-15 Summary of Findings table for the EPDS and identification of depression in the postnatal period (based on NICE 2015)

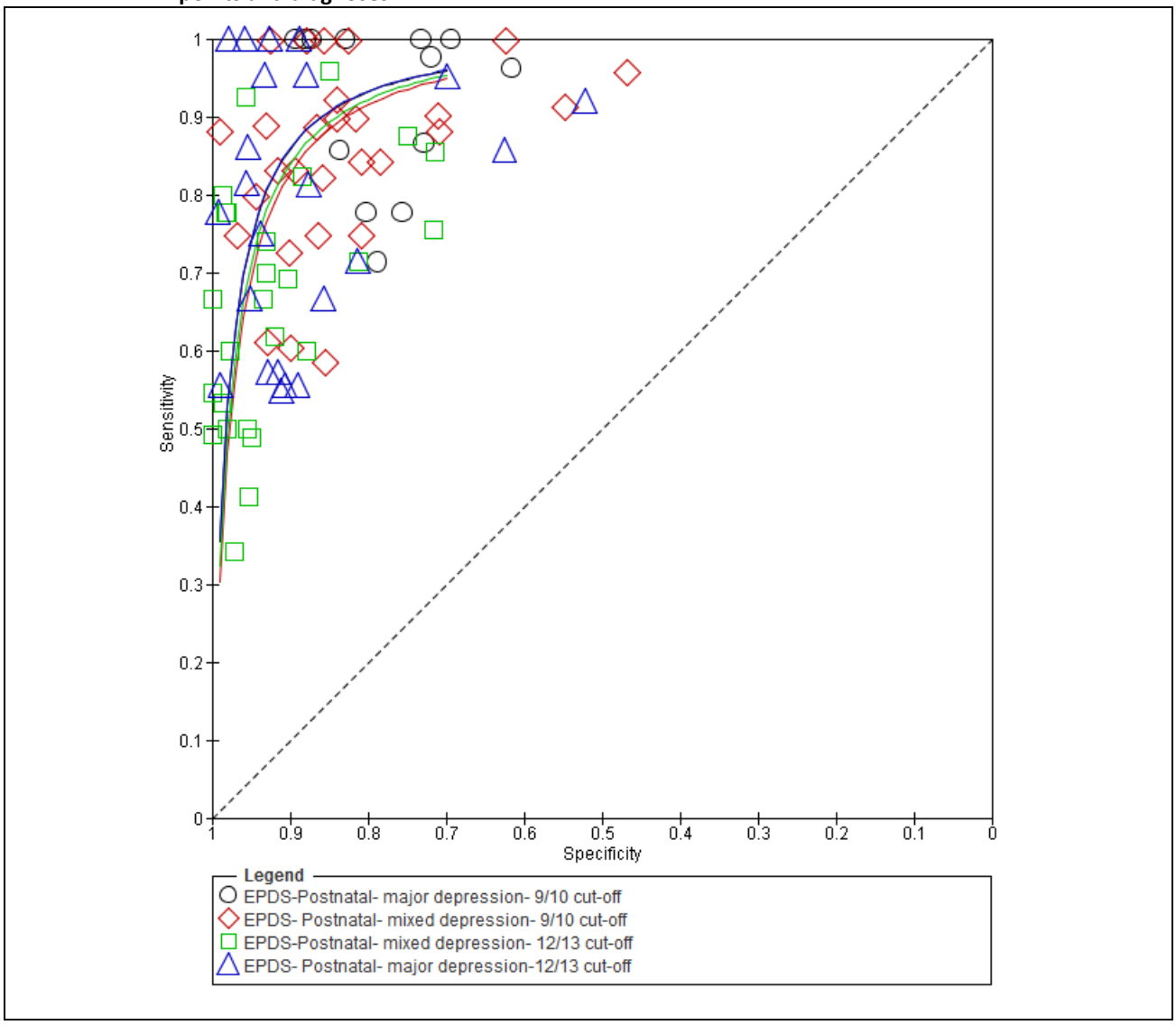
Tool; Condition; Cut-off	No of studies[1] (participants)	Important outcomes		Critical outcomes			Overall certainty
		Pooled sensitivity (95% CI) [1]	Pooled specificity (95% CI) [1]	Positive Likelihood Ratio [2]	Negative Likelihood Ratio [2]	Pooled AUC	
EPDS; mixed depression[3]; 9/10	29 (5,463)	0.83 (0.81, 0.86)	0.85 (0.84, 0.86)	5.53	0.20	NR [4]	●●●● High[5]
EPDS; mixed depression[3]; 12/13	29 (5,209)	0.68 (0.66, 0.71)	0.92 (0.92, 0.93)	8.50	0.35	NR [4]	●●●● High[5]
EPDS; major depression[3]; 9/10	13 (2,277)	0.95 (0.92, 0.97)	0.82 (0.80, 0.84)	5.28	0.06	NR [4]	●●●● High[5]
EPDS; major depression[3]; 12/13	22 (4,355)	0.80 (0.77, 0.83)	0.93 (0.92, 0.94)	11.43	0.22	NR [4]	●●●● High[5]
Evidence statements: *A score of 13 or more on the EPDS has moderate sensitivity and high specificity for detecting possible major depression in postpartum women (high certainty evidence)* *A score of 10 or above on the EPDS has moderate sensitivity and moderate specificity for detecting possible depressive disorders (minor and major depression) in postpartum women (high certainty evidence).*							

Tool; Condition; Cut-off	No of studies[1] (participants)	Important outcomes		Critical outcomes			Overall certainty
		Pooled sensitivity (95% CI) [1]	Pooled specificity (95% CI) [1]	Positive Likelihood Ratio [2]	Negative Likelihood Ratio [2]	Pooled AUC	

Footnotes:
[1] Reproduced from Table 16 of the NICE 2015 Guideline
[2] Calculated from the pooled sensitivity and specificity using the following formulas: LR+ = sensitivity/(1-specificity); LR- = (1-sensitivity)/specificity
[3] Definitions of 'minor', 'major' and 'mixed' depression as per NICE 2015
[4] Pooled AUC not reported but ROC included herein
[5] Multiple studies with a range of quality from a large total sample (greater than 1,000)

Abbreviations: AUC, area under the curve; CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; LR, likelihood ratio; NICE, National Institute of Health and Care Excellence; NR, not reported; ROC, receiver operating characteristics.

Figure B5-1 Summary of ROC curve for the EPDS administered in the postnatal period at different cut-off points and diagnoses



Reproduced from Figure 4 (page 104) of NICE 2015 Guideline.,

Abbreviations: EPDS, Edinburgh Postnatal Depression Scale; ROC, receiver operating characteristics.

Table B5-16 Summary of Findings table for the PHQ and identification of depression in the postnatal period (based on NICE 2015)

Tool; Condition; Cut-off	No of studies[1] (participants)	Important outcomes		Critical outcomes			Overall certainty
		Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	Positive Likelihood Ratio [2]	Negative Likelihood Ratio [2]	Pooled AUC	
PHQ-2; major depression [3]; 2/3	2 (719)	0.84 (0.71, 0.94) 0.77 (0.46, 0.95)	0.79 (0.75, 0.83) 0.59 (0.53, 0.66)	4.00 1.88	0.20 0.39	NR	●●○○ Low [6]
PHQ-2; major depression [3]; 3/4	1 (213)	0.63 (0.32, 0.86)	0.79 (0.73, 0.84)	3.00	0.47	NR	●○○○ Very low [7]
PHQ-8; major depression [3]; 9/10	1 (213)	0.77 (0.46, 0.95)	0.62 (0.55, 0.69)	2.03	0.37	NR	●○○○ Very low [7]
PHQ-9 (simple scoring[4]); major depression [3]; 9/10	2 (605)	0.89 (0.80, 0.95) 0.82 (0.68, 0.92)	0.65 (0.43, 0.84) 0.84 (0.80, 0.87)	2.54 5.13	0.17 0.21	NR	●●○○ Low [8]
PHQ-9 (complex scoring[5]); major depression [3]; 9/10	1 (506)	0.67 (0.51, 0.80)	0.92 (0.89, 0.94)	8.38	0.36	NR	●○○○ Very low [7]
Evidence statement: *It is uncertain if the PHQ has adequate sensitivity or specificity to detect possible depressive disorders in postpartum women (very low to low certainty evidence).*							

Footnotes:

[1] Reproduced from Table 17 of the NICE 2015 Guideline

[2] Calculated from the reported sensitivity and specificity using the following formulas: LR+ = sensitivity/(1-specificity); LR- = (1-sensitivity)/specificity

[3] Definitions of 'minor', 'major' and 'mixed' depression as per NICE 2015

[4] Simple scoring as reported by NICE 2015: result is positive if sum of numbered responses is ≥10.

[5] Complex scoring as reported by NICE 2015: result is positive if at least 5 symptoms are present, including symptom 1, symptom 2, or both, and each symptom present has a response score of 2 to 3, except for symptom 9, for which a response score of 1 to 3 was acceptable.

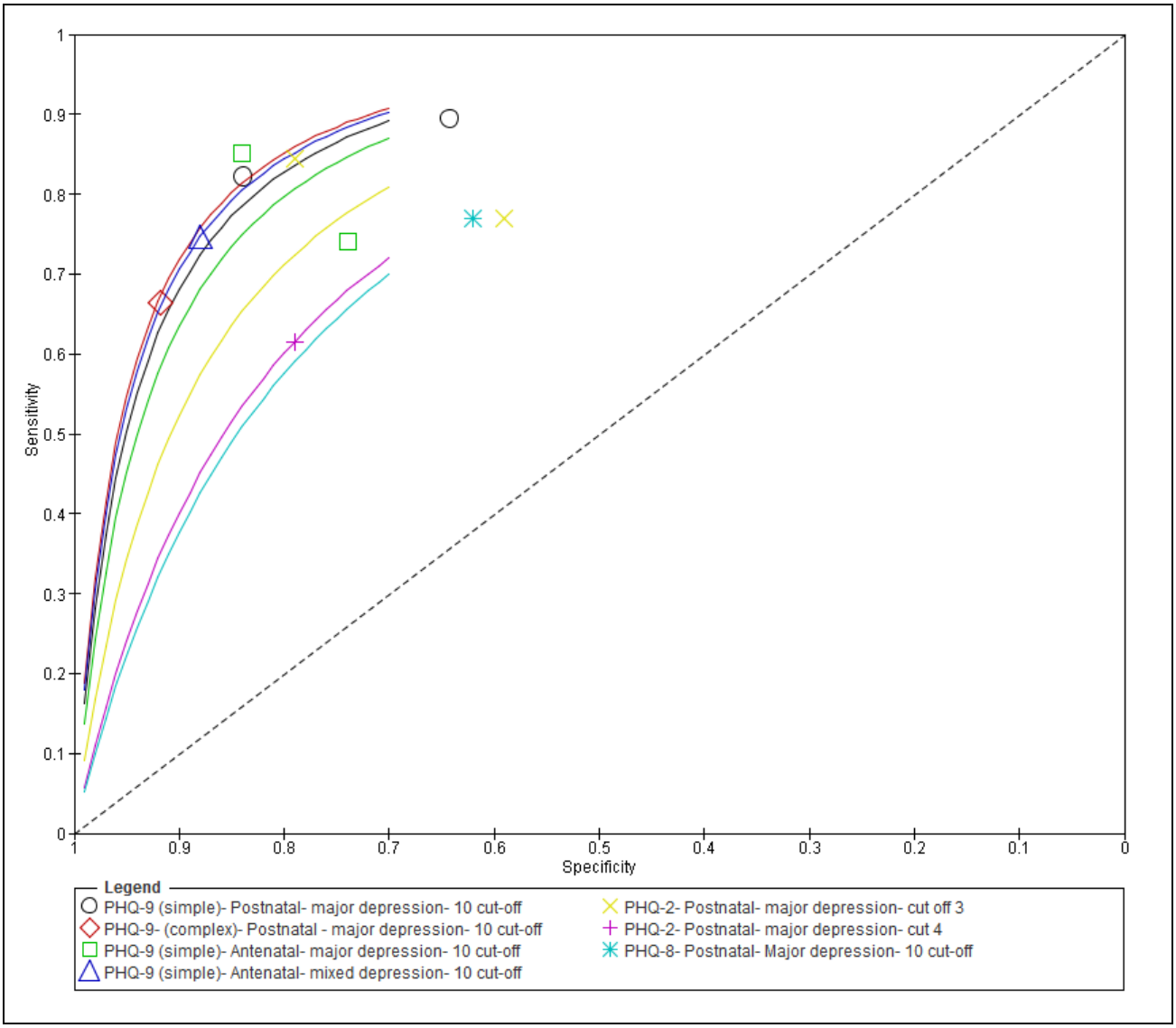
[6] Two studies of very low quality

[7] One study of very low quality

[8] Two studies of low to very low quality

Abbreviations: AUC, area under the curve; CI, confidence interval; LR, likelihood ratio; NICE, National Institute of Health and Care Excellence; NR, not reported; PHQ, Patient Health Questionnaire

Figure B5-2 Summary of ROC curve for the PHQ (2-, 8- and 9-item versions) at different timings, diagnoses and cutoffs



Reproduced from Figure 6 (page 107) of NICE 2015 Guideline.

Abbreviations: NICE, National Institute of Health and Care Excellence; PHQ, Patient Health Questionnaire; ROC, receiver operating characteristics.

Table B5-17 Summary of Findings table for the 'Whooley questions' and identification of depression in the postnatal period (based on NICE 2015)

Tool; Condition; Cut-off	No of studies[1] (participants)	Important outcomes		Critical outcomes			Overall certainty
		Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	Positive Likelihood Ratio [2]	Negative Likelihood Ratio [2]	Pooled AUC	
Whooley questions; mixed depression[3]	1 (94)	1.00 (0.81, 1.00)	0.64 (0.53-0.75)	2.75	0.02	NR	●○○○ Very low [4]
Whooley questions (+ help question); mixed depression[3]	1 (45)	0.39 (0.17, 0.64)	1.00 (0.87, 1.00)	39.00	0.62	NR	●○○○ Very low [4]
Whooley questions; major depression[3]	1 (506)	1.00 (0.92, 1.00)	0.44 (0.39, 0.49)	1.77	0.02	NR	●○○○ Very low [4]
Evidence statement: *It is uncertain if the 'Whooley questions' have adequate sensitivity or specificity to detect possible depression in postpartum women (very low certainty evidence)*							
Footnotes: [1] Reproduced from Table 18 of the NICE 2015 Guideline. [2] Calculated from the reported sensitivity and specificity using the following formulas: LR+ = sensitivity/(1-specificity); LR- = (1-sensitivity)/specificity, with rounding of values of 1.00 to 0.99 to avoid Div/0 error. [3] Definitions of 'minor', 'major' and 'mixed' depression as per NICE 2015. [4] One study of moderate quality.							

Abbreviations: AUC, area under the curve; CI, confidence interval; LR, likelihood ratio; NICE, National Institute of Health and Care Excellence; NR, not reported.

Table B5-18 Summary of Findings table for the K-10 and identification of depression in the postnatal period (based on NICE 2015)

Tool; Condition; Cut-off	No of studies[1] (participants)	Important outcomes		Critical outcomes			Overall certainty
		Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	Positive Likelihood Ratio [2]	Negative Likelihood Ratio [2]	Pooled AUC	
K-10; mixed depression[3]; 6	1 (61)	0.85 (0.66, 0.96)	0.41 (0.25, 0.59)	1.44	0.37	NR	●○○○ Very low [4]
Evidence statement: *It is uncertain if the K-10 has adequate sensitivity or specificity to detect possible depression in postpartum women (very low certainty evidence)*							
Footnotes: [1] Reproduced from Table 19 of the NICE 2015 Guideline [2] Calculated from the reported sensitivity and specificity using the following formulas: LR+ = sensitivity/(1-specificity); LR- = (1-sensitivity)/specificity [3] Definitions of 'minor', 'major' and 'mixed' depression as per NICE 2015 [4] One study of moderate quality							

Abbreviations: AUC, area under the curve; CI, confidence interval; K-10, Kessler 10 item questionnaire; NICE, National Institute of Health and Care Excellence; NR, not reported.

B5.4 NON-TECHNICAL CHARACTERISTICS OF RELEVANT TOOLS

The table below summaries the non-technical characteristics of the included depression screening tools for which there was evidence of technical performance (regardless of the certainty of that evidence). The complexity of scoring for each tool has been assessed as Simple, Moderate or High on the basis of information in the published literature and the experience of the EWG.

Table B 5-19 Non-technical characteristics of the relevant included tools for depression screening

Tool	Number of items	Time to administer (mins)	Complexity of scoring	Available languages
EPDS	10	5-10 mins	Simple	Developed in English and validated for depression screening in >20 languages Translated into >50 languages
PHQ-9	9	5-10 mins	Simple	English
Whooley questions	2	<2 mins	Simple	English
K-10	10	5-10 mins	Simple	English

Abbreviations: EPDS, Edinburgh Postnatal Depression Scale; PHQ, Patient Health Questionnaire; K-10 Kessler 10

B5.5 CLINICAL USEFULNESS OF RELEVANT TOOLS

B5.5.1 Acceptability of depression screening

Two of the SRs identified in the main search describe the acceptability of perinatal depression screening: Brealey 2010, El-Den 2015.

El-Den 2015 undertook a SR of studies that explored the acceptability of postnatal depression screening. The included studies encompassed a range of tools including the ANRQ and the EPDS, with the EPDS being the tool most commonly assessed for acceptability. Acceptability was measured in a range of participants: pregnant women, midwives, maternal child and family health nurses, health visitors, and primary health nurses. The 29 included studies used a variety of different qualitative and quantitative methods, and a wide range of terms, questions and statements. The SR authors highlight the need for uniform, psychometrically tested methods to measure acceptability. Nonetheless, they conclude that postnatal depression screening is generally acceptable to most perinatal women, healthcare professionals and the general public. They also draw attention to the importance of setting and highlighted a study by Drake 2014 that found that online postnatal depression screening at home can be easy, straightforward and personalised, and can help to overcome the challenges of fear and stigma associated with postnatal depression screening.

The earlier SR by **Brealey 2010** also sought to identify studies that assessed the acceptability of postnatal depression screening, and there is overlap in the studies included in the two SRs. However, the inclusion criteria for Brealey 2010 were more restricted: studies were only included if they assessed acceptability in the prenatal and postnatal periods. Fifteen of the 16 included studies focused on the EPDS. The studies found depression screening was generally acceptable to women and healthcare professionals, but emphasise the importance of ensuring a woman feels comfortable for her to answer screening questions honestly. They also discuss the importance of health professional awareness of differing cultural attitudes towards the topics in the screening questions, and the ambiguity of the question in the EPDS about self-harm.

B5.5.2 Effectiveness of depression screening

The main literature search identified four SRs describing the effectiveness of perinatal depression screening: Hewitt 2009a; Hewitt 2009b; Myers 2013; and Thombs 2014. Three of these SRs (Hewitt 2009a, 2009b and NICE 2015) also describe the cost-effectiveness of screening, which is addressed separately below. None of the SRs describe the effectiveness of co-administration of a screening tool with psychosocial assessment.

The aim of the SR by **Hewitt 2009a** was to identify studies that reported on the clinical effectiveness and cost-effectiveness of antenatal and postnatal identification of depressive symptoms. Four individual studies (one of them Australian) were identified that compared a formal method to identify perinatal depression (with or without enhancement of care) versus not using a formal method. All the included studies used the EPDS. The pooled results of the studies demonstrated beneficial effects of using the EPDS in reducing EPDS scores (OR, 0.61; 95% CI 0.48-0.76). The authors note that is was not possible to 'disentangle' the effects of screening from enhanced care interventions linked to positive screens.

The review by **Hewitt 2009b** describes a comprehensive, integrated evidence synthesis and value of information analysis undertaken from a UK perspective. The authors undertook multiple reviews of: identification of the methods to identify postnatal depression; the validity of methods to identify postnatal depression; the acceptability to women and health professionals of methods to identify postnatal depression; and the clinical effectiveness of methods to identify postnatal depression. The findings from these reviews were then combined to identify research priorities and determine if postnatal depression screening met the UK National Screening Committee criteria. The clinical effectiveness review in this publication is essentially the same as that reported in Hewitt 2009b, with a slight difference in the reported pooled OR of reducing EPDS scores (0.64, 95% CI 0.52-0.78).

The SR by the US Agency for Healthcare Research and Quality (**Myers 2013**) describes a comparative effectiveness review of the diagnostic accuracy, benefits and harms of available screening instruments available for postnatal depression. These authors describe five studies showing evidence of reduced number of symptoms with screening plus an intervention, one study showing improved scores with screening plus intervention, four studies showing no improvement in parental stress with screening plus intervention, and one study showing increased number of visits for infants of screened women. The included studies used a variety of screening tools. The SR concludes that there is evidence of screening effectiveness when staff-assisted depression care supports are in place, but not without these supports.

The aim of the SR by **Thombs 2014** was to determine whether depression screening improves depression outcomes among women during pregnancy or postpartum. The authors identified a single study in postpartum women, and no studies in pregnant women. The study in postpartum women reported a standardized mean difference for symptoms of depression at 6 months of 0.34 (95% CI 0.15-0.52), but it should be noted that this study was assessed by the authors as having a high risk of bias. Thombs provide a discussion of Hewitt 2009a and 2009b and Myers 2013, and draw attention to the fact that there was no overlap in the studies included in those to SRs. Thombs and colleagues suggest this might be due to differences in study inclusion criteria related to whether or not studies used pre-defined cut-off scores and whether or not they recruited only women without a pre-existing diagnosis of depression. Thombs and colleagues assert that an effective universal screening program should minimise false positives, and that the paucity of available evidence of effectiveness precludes recommendations for universal screening.

A detailed analysis of the discrepancies between all the SRs included here is outside the scope of the current review, however, we do note that the stated screening aim of Thombs 2014 aim is at odds with the stated aim of the EWG for the current Guideline.

B5.5.3 Implementation of depression screening

Four SRs were identified that describe different aspects of implementation of perinatal mental health screening: Goldin Evans 2015; Hewitt 2010; Reuland 2009; and Shrestha 2016. One of these SRs focuses on the validation of the EPDS in postnatal women, in different languages, and different settings (**Hewitt 2010**). Another SR describes the reliability and validity of the EPDS doe detecting depression and other common mental health conditions among women in low- and lower-middle-income countries (**Shrestha 2016**). A third SR describes a range of different depression screening tools (including the EPDS) that have been validated in Spanish (**Reuland 2009**), and the fourth SR describes a review of screening practices for postnatal depression (**Goldin Evans 2015**).

Collectively, these SRs provide evidence of the translation and validation of the EPDS in more than twenty languages, in antenatal clinics, postnatal hospital wards, and follow-up clinics. Whilst the EPDS appears to have been satisfactorily translated and validated in developed countries, there are some concerns regarding the absence of culturally sensitive translations in lower income countries. The SR by Shrestha 2016 found that the local language versions of the EPDS they identified in their search reported lower precision in a general perinatal population than original reports of the English version. These authors note that even in studies where a diagnostic interview was conducted as a reference standard, screening questions had not been culturally adapted and may not have been well-understood by the women being screened. The authors recommend that effective and culturally sensitive translation and validation processes require explicit consideration of local terminology for psychological distress, mental disorders, emotional literacy and general literacy. In addition, empirically validated cut-off scores are required for women from lower income countries.

The SR by Goldin Evans 2015 summarises evidence from studies of the screening practices of physicians (paediatricians, obstetricians, and family physicians (GPs)) in the USA. The studies reported that rates of postnatal depression screening were low in practice, and that paediatricians were the least likely to screen women. Only one in four physicians reported ever using a screening tool, and the majority of physicians did not feel confident in their skills to recognise postnatal depression. Almost two-thirds of physicians reported time constraints as a significant barrier to screening. Inadequate training or skills were also seen as a barrier, although most respondents reported they would be willing to using screening tools. The authors also discuss how government funding initiatives in the USA were associated with increases in the rates of postnatal depression screening.

B5.6 COST-EFFECTIVENESS OF PERINATAL DEPRESSION SCREENING

No cost-effectiveness data were identified that are directly relevant to the Australian context, due to differences in approach to screening, pathways to care, and differences in input costs.

An analysis conducted for the UK National Institute for Health Research (NIHR)-Health Technology Assessment Programme (Hewitt et al 2009; Paulden et al 2009) concluded that formal identification of postnatal depression (PND) using the EPDS (with cut points ranging 12–16) do not represent value for money for the UK National Health Service, mainly due to the potential additional costs of managing women incorrectly diagnosed as depressed.

In contrast, a more recent cost-effectiveness analysis for NICE found that the use of a brief case identification tool (that is, the Whooley questions), followed by the use of a more formal method (such as the EPDS or PHQ-9), appears to be the most cost-effective approach in the identification of depression in the postnatal period (NICE 2015).

Likewise, a recent study from a Medicaid payer perspective (Wilkinson 2017) assessed the cost-effectiveness of a two-stage approach to screening, whereby all women were screened with the short-form EPDS and then only those women who were positive received further screening with the 10-item long-

form. The analysis found that routine screening and treatment of PND is a cost-effective intervention under a wide range of willingness-to-pay thresholds and should be considered as part of usual postnatal care.

In Canada, a large randomised controlled trial (RCT) is underway to assess the clinical and cost-effectiveness of usual prenatal care plus an integrated intervention comprising online psychosocial assessment, referral and online cognitive behavioural therapy (CBT) (Kingston 2014). The integrated care model incorporates online screening for prenatal depression using the EPDS, together with online psychosocial risk assessment using the Antenatal Risk Questionnaire (ANRQ-R). Women who meet the criteria for CBT based on ANRQ-R and EPDS scores are then referred to online CBT, involving six, 30-minute interactive modules over 6 to 8 weeks. An early feasibility study found that women were very receptive to online screening (Kingston 2015). According to the study protocol, the economic evaluation will involve a within-trial cost-effectiveness analysis comparing the integrated intervention 'package' with usual prenatal care. The perspective of the primary analysis will be that of the Canadian health and social care budget, with a secondary analysis that adopts a societal perspective incorporating personal and productivity costs.

B5.7 OVERALL SUMMARY OF FINDINGS

An overall assessment of the technical performance, non-technical characteristics and clinical usefulness of depression screening tools is presented in **Table B5-20**.

Table B5-20 Overall Summary of Findings related to the use of perinatal depression screening tools

Tool(s)	Technical characteristics		Non-technical characteristics		Clinical usefulness		
	Performance[1]	**Certainty[2]**	**Ease of Administration[3]**	**Language availability[4] & cultural sensitivity[5]**	**Acceptability[6]**	**Effectiveness[7]**	**Implementability[8]**
EPDS	Antenatal: Acceptable	●●●● High	High	Multiple languages Multiple populations	High	Good	High
	Postnatal: Acceptable	●●●● High					
PHQ-9	Antenatal: Uncertain	●●○○ Low	High	English Western populations	Unknown but likely to be Good	Unknown	High
	Postnatal: Uncertain	●●○○ Low					
Whooley questions	Antenatal: Uncertain	●●○○ Low	High	English Western populations	Unknown but likely to be Good	Limited	High
	Postnatal: Uncertain	●○○○ Very low					
K-10	Antenatal: Uncertain	●●○○ Low	High	English Western populations	Unknown but likely to be Good	Unknown	High
	Postnatal: Uncertain	●○○○ Very low					

Footnotes:

[1] Performance defined as sensitivity, specificity, positive likelihood ratio, negative likelihood ratio (defined as Acceptable, Limited, or Uncertain).

[2] Certainty assessed according to GRADE and QUADAS-2 criteria (defined as High, Moderate, Low or very Low).

[3] Ease of administration was based on judgment regarding the number of items, and the time and complexity of administering and scoring the tool (rated as High, Moderate, or Low)

[4] Language availability based on information from the included literature and the awareness of the EWG

[5] Cultural sensitivity was based on information from the included literature of any use in culturally and linguistically diverse populations

[6] Acceptability was based on the overall judgement of the EWG of the acceptability of each tool to women, health care professionals and/or the general public (rated as High, Moderate, Low or Unknown)

[7] Effectiveness was defined as positive impact on depressive symptoms, services referred to or utilized, and impact on a woman's mental health (rated as High, Good, Limited, or Unknown)

[8] Implementability was based on the overall judgement of the EWG based on available information regarding the training requirements for use of the tool and implications for current models of care and staff and service availability

Abbreviations: EPDS, Edinburgh Postnatal Depression Scale; EWG, Expert Working Group; PHQ, Patient Health Questionnaire; K-10 Kessler 10.

B6 SCREENING FOR ANXIETY

B6.1 RELEVANT OUTCOMES OF TECHNICAL PERFORMANCE

As for depression screening, the EWG agreed that positive and negative likelihood ratios (**LR+** and **LR-**, respectively), and the **AUROC** (for different cut-offs) are the critical outcomes for assessing test performance, and **sensitivity** and **specificity** as important outcomes. The clinical consequences of different test results for perinatal anxiety screening are considered to be identical to those for perinatal depression screening.

B6.2 CRITICAL APPRAISAL OF TECHNICAL PERFORMANCE

Meades 2011 used the original version of the QUADAS checklist to assess the quality of the included studies of test accuracy. The authors assessed 11 criteria as present or absent in each included study, and summed the number of criteria present to derive an overall quality rating out of 11. The eleven criteria were as follows:

- Explicit study aims
- Adequate sample size
- Sample described in sufficient detail
- Sample representative of population receiving test in practice
- Clear inclusion and exclusion criteria
- Use of appropriate reference standard
- Reliability of measure reported
- Validity of measure reported
- Specification of drop-outs and withdrawal of participants
- Adequate description of data
- Discussion of generalisability

The Meades 2011 authors state that *"most studies were of good quality… having a score of 8 or more"* (Section 2.3, page 3). No other assessment of the overall degree of quality was given by Meades 2011.

To ensure consistency of approach between the current evidence reviews of anxiety screening and depression screening, individual studies included within Meades 2011 have been re-appraised according to the QUADAS-2 checklist. However, this re-appraisal has been limited to studies that clearly defined anxiety 'caseness' as a reference standard (see below for further details).

Three additional studies identified via the literature search update (Grigoriadis 2011, Simpson 2014, and Tran 2011) have also been critically appraised using the QUADAS-2 checklist. For all studies, the full text versions of the relevant study publications have been used to extract study characteristics. The quality of each study and the overall certainty of the evidence have been determined as for depression screening.

B6.3 EVIDENCE OF TECHNICAL PERFORMANCE

B6.3.1 Characteristics of individual studies included for anxiety screening

The table below presents the key characteristics of the studies included in the current evidence review of perinatal anxiety screening, comprising nine (9) individual studies cited in Meades 2011 and three (3) individual studies published subsequently. It should be noted that Meades 2011 appears to include 12 studies relevant to our research question. However, during the re-appraisal of these studies it became apparent that there was duplicate publication data from the same study population: Kitamura 1989, Kitamura 1994a and Kitamura 1994b. The demographic characteristics and reported findings are identical

across these three papers. Consequently, in the current evidence review they are presented as one study 'Kitamura 1994', with data extracted from each publication, as required. The specific reference used for data is footnoted in the relevant tables.

The number of studies (k) that reported on each instrument is as follows: EPDS, four (4) studies; GAD, one (1) study; GHQ, seven (7) studies; HADS, one (1) study; K-10, one (1) study; and STAI, one (1) study. Six (6) studies reported on two instruments.

A number of characteristics were considered to be fundamental to the ascertainment of certainty of the evidence, and these have been extracted from Meades 2011 and from the individual study publications. The country in which each study was performed has relevance to the applicability of the results to the Australian healthcare system. The setting and population in each study is most important for determining the generalizability of the study results to the proposed use of a psychometric screening instrument in the Australian context. In particular, it was considered important to identify whether each study recruited a 'primary care sample' (i.e. a general perinatal population with no known mental health issues) or a 'referral sample' (i.e. women already identified as having mental health symptoms who have been referred for further psychological or psychiatric assessment).

Another important characteristic of each study (which is not evident in the Meades 2011 review) is the variability across perinatal anxiety screening studies in the definition of a 'case'. Some examples of definitions are (1) generalized anxiety disorder alone, (2) generalized anxiety disorder with co-morbid depression, and (3) the presence of generalized anxiety or depression. This characteristic is important because it may be inappropriate to pool results from studies that have used different definitions of a 'case'.

Finally, only four individual studies were considered relevant to the assessment of anxiety screening tools to detect cases of anxiety (and assessed using QUADAS-2 methods): **Grant 2008** (high quality), **Grigoriadis 2011** (very low quality), **Simpson 2014** (very low quality), and **Spies 1988** (low quality). The remaining studies were rated by Meades 2011 as being of 'good' (6 studies), or 'not good' (1 study) quality. These studies have not been re-appraised using QUADAS-2.

Finally, the timing of testing for anxiety symptoms is important and has been extracted for each included study. The reasons for feelings of anxiousness could be reasonably expected to fluctuate during pregnancy (e.g. fear of miscarriage during the first trimester vs fear of childbirth in the third trimester) and in the post-partum period (e.g. fear related to the safety of the infant). Consequently, it may not be appropriate to pool results from women at different timepoints across the perinatal period

Table B6-1 Key characteristics of studies that used the EPDS (full or 3A), GAD-7, GHQ (12, 28 or 30-item), HADS, K-10 or STAI to identify perinatal anxiety

Study ID	Tool(s)	Country; Setting (Population sample)	Case Definition [2]	Patient Selection		Index test(s)		Ref. Standard		Flow & Timing	Study Quality
				Risk of bias	Applicability concerns	Risk of bias	Applicability concerns	Risk of bias	Applicability concerns	Risk of bias	
Studies in Meades 2011 that reported criterion validity											
Abiodun 1993	GHQ-30	Nigeria [1]; Non-psychiatric & community clinics [2] (Primary care [2])	Psychiatric morbidity	NA	*Low*	NA	NA	NA	NA	NA	9/11 [1]
Abiodun 1994	GHQ-12 HADS	Nigeria [1]; Non-psychiatric & community clinics [2] (Primary care [2])	Psychiatric morbidity, anxiety or depression	NA	*Low*	NA	NA	NA	NA	NA	6/11 [1]
Aderibigbe 1992	GHQ-28	Nigeria [1]; Antenatal clinic [2] (Primary care [2])	Any DSM-III diagnosis	NA	*Low*	NA	NA	NA	NA	NA	8/11 [1]
Grant 2008	STAI EPDS	Australia [1]; Antenatal clinic [2] (Primary care [2])	Anxiety	Low	Low	Low	Low	Low	Low	Low	High [4] (11/11 [1])
Kitamura 1994b	GHQ-30	Japan [1]; Perinatal clinics [2] (Primary care [2])	Anxiety, depression or psychiatric disorder	NA	*Low*	NA	NA	NA	NA	NA	8 or 9/11 [1]
Navarro 2007	GHQ-12 EPDS	Spain [1]; Postnatal clinic [2] (Primary care [2])	Anxiety, depression or adjustment disorder	NA	Low	NA	NA	NA	NA	NA	10/11 [1]
Nott 1982	GHQ-30	UK [1]; Postnatal home visits [2] (Primary care [2])	Psychiatric disorder	NA	Low	NA	NA	NA	NA	NA	9/11 [1]
Sharp 1988	GHQ-30	UK [1]; Antenatal clinic [2] (Primary care [2])	Psychiatric disturbance	NA	Low	NA	NA	NA	NA	NA	9/11 [1]
Spies 1988	K-10	South Africa [1]; Antenatal clinic [2] (Primary care [2])	PTSD; panic disorder; social phobia	Low	Low	Unclear	Unclear	Low	Low	High	Low [4] (9/11 [1])
Studies identified in literature search update											
Grigoriadis 2011	EPDS EPDS-3	Canada [2]; Psychiatric clinic (Referral sample)	GAD	Unclear	High	Unclear	Low	Unclear	Low	Unclear	Very low [4]
Simpson 2014	EPDS GAD-7	Canada [2]; Psychiatric clinic (Referral sample)	GAD; GAD + MDD	High	High	High	Unclear	High	Low	High	Very low [4]
Tran 2011	EPDS GHQ-12	Vietnam [2]; Perinatal clinics (Primary care [2])	Depression or GA or panic disorder	Unclear	Low	Low	Low	Low	Low	Unclear	Moderate [4]

Footnotes:

Italics indicate divergence from data reported in foundation review

[1] Compiled from information presented in Meades 2011.

[2] Compiled from information from the full text versions of the source articles.

[3] During data extraction it became apparent that the same data from the same study population has been reported in Kitamura 1989, Kitamura 1994a and Kitamura 1994b. Consequently these three publications are presented here as one study.

[4] Determined by the Expert Working Group for the current Guideline based on QUADAS-2 methods.

Abbreviations: DSM, Diagnostic and Statistical Manual; EPDS, Edinburgh Postnatal Depression Scale; GA, generalized anxiety; GAD, Generalised Anxiety Disorder; GAD-7, Generalised Anxiety Disorder scale – 7-item scale; GHQ, General Health Questionnaire; HADS, Hospital Anxiety and Depression Scale; K-10, Kessler 10 item questionnaire; MDD, major depressive disorder; NA, not applicable; PTSD, post-traumatic stress disorder; STAI, State-Trait Anxiety Inventory; UK, United Kingdom.

B6.3.2 Evidence summaries from included studies

The tables below present evidence from all the studies included in Meades 2011 and identified in the supplementary literature search. All data have been extracted from the source publication for each study, as not all of the information relevant to the current review was extracted by Meades 2011.

As noted above, only four studies met all of our inclusion criteria: only data from these four studies has been taken through to the Summary of Findings tables.

A total of nine (9) separate tables are presented for each instrument and each perinatal time period: EPDS (antenatal, postnatal, and perinatal), GAD-7 (perinatal), GHQ (antenatal, postnatal), HADS (antenatal only), K-10 (antenatal only), and STAI (antenatal only). Results are reported by version of each tool, condition and cut-off threshold. The timing of testing (e.g. pregnancy trimester, or weeks/months postpartum) is also noted in the tables.

B6.3.2.1 EPDS

Table B6-2 Evidence summary table for the EPDS for detection of anxiety in antenatal women

Version; Condition; Cut-off	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC (95% CI) [2]	LR+	LR-	Study Quality [3]
EPDS-full; Depression, generalized anxiety, panic disorder; 3/4						
Tran 2011 (third trimester; n=199)	0.66 (NR)	0.72 (NR)	0.75 (0.67, 0.873)	NR	NR	Moderate

Abbreviations: AUC, area under the curve; CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; LR, likelihood ratio; NR, not reported.

[3] Based on QUADAS-2 assessment for this review.

Table B6-3 Evidence summary table for the EPDS for detection of anxiety in postnatal women

Version; Condition; Cut-off	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC (95% CI) [2]	LR+	LR-	Study Quality [3]
EPDS-full; Depression, generalized anxiety, or panic disorder; 3/4						
Tran 2011 (4-6 weeks postpartum; n=165)	0.75 (NR)	0.75 (NR)	0.79 (0.71, 0.87)	NR	NR	Moderate
EPDS-full; Depression, anxiety or adjustment disorder; 9/10						
Navarro 2007 (6 weeks postpartum)	0.86 (0.80, 0.90)	0.85 (0.80, 0.90)	0.93 (0.91, 0.96)	NR	NR	*Not assessed*

Abbreviations: AUC, area under the curve; CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; LR, likelihood ratio; NR, not reported.

[3] Based on QUADAS-2 assessment for this review.

Table B6-4 Evidence summary table for the EPDS for detection of anxiety in a mixed population of antenatal and postnatal women

Version; Condition; Cut-off	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC (95% CI) [2]	LR+	LR-	Study Quality [3]
EPDS-full; Generalised anxiety disorder; 12/13						
Grigoriadis 2011 (any trimester, n=62; postpartum, n=29)	0.70 (NR)	0.82 (NR)	NR	NR	NR	Very low
Simpson 2014 (any trimester of pregnancy, n=155; or any time postpartum, n=85)	0.89 (NR)	0.40 (NR)	0.62 (NR)	NR	NR	Very low
EPDS-full; Co-morbid generalised anxiety disorder and MDD; 17						
Simpson 2014 (any trimester of pregnancy, n=155; or any time postpartum, n=85)	0.67 (NR)	0.65 (NR)	0.68 (NR)	NR	NR	Very low
EPDS-3A; Generalised anxiety disorder; >4						
Grigoriadis 2011 (any trimester, n=62; or postpartum, n=29)	0.88 (NR)	0.49 (NR)	NR	NR	NR	Very low

Version; Condition; Cut-off	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC (95% CI) [2]	LR+	LR-	Study Quality [3]
EPDS-3A; Generalised anxiety disorder; 7						
Simpson 2014 (any trimester of pregnancy, n=155; or any time postpartum, n=85)	0.68 (NR)	0.64 (NR)	0.69 (NR)	NR	NR	Very low
EPDS-3A; Co-morbid generalised anxiety disorder and MDD; 7						
Simpson 2014 (any trimester of pregnancy, n=155; or any time postpartum, n=85)	0.69 (NR)	0.59 (NR)	0.67 (NR)	NR	NR	Very low

Abbreviations: Abbreviations: AUC, area under the curve; CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; LR, likelihood ratio; NR, not reported.
[3] Based on QUADAS-2 assessment for this review.

B6.3.2.2 GAD-7

Table B6-5 Evidence summary table for the GAD-7 for detection of anxiety in a mixed population of antenatal and postnatal women

Version; Condition; Cut-off	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC (95% CI) [2]	LR+	LR-	Study Quality [3]
GAD-7; Generalized Anxiety Disorder alone; 10						
Simpson 2014 (any trimester of pregnancy, n=155; or any time postpartum, n=85)	0.76 (NR)	0.52 (NR)	0.71 (NR)	NR	NR	Very low
GAD-7; Generalized Anxiety Disorder alone; 13						
Simpson 2014 (any trimester of pregnancy, n=155; or any time postpartum, n=85)	0.61 (NR)	0.73 (NR)	0.71 (NR)	NR	NR	Very low
GAD-7; Generalized Anxiety Disorder with Major Depressive Disorder; 13						
Simpson 2014 (any trimester of pregnancy, n=155; or any time postpartum, n=85)	0.67 (NR)	0.69 (NR)	0.74 (NR)	NR	NR	Very low

Abbreviations: AUC, area under the receiver-operator curve; CI, confidence interval; FN, false negative; FP, false positive; NR, not reported; TN, true negative; TP, true positive.
[3] Based on QUADAS-2 assessment for this review.

B6.3.2.3 GHQ

Table B6-6 Evidence summary table for the GHQ for detection of anxiety in antenatal women

Version; Condition; Cut-off	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC (95% CI) [2]	LR+	LR-	Study Quality [3]
GHQ-30; Psychiatric morbidity or disorder; 4/5						
Abiodun 1993 (all trimesters)	0.80 (NR)	0.81 (NR)	NR	NR	NR	*Not assessed*
Kitamura 1994a (first trimester) [4]	0.89 (NR)	0.48 (NR)	NR	NR	NR	*Not assessed*

Version; Condition; Cut-off	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC (95% CI) [2]	LR+	LR-	Study Quality [3]
Kitamura 1994a (third trimester)	0.39 (NR)	0.65 (NR)	NR	NR	NR	*Not assessed*
GHQ-30; Psychiatric disturbance; 5/6						
Sharp 1988 (first & second trimesters)	0.77	0.78	0.84 (SD 0.04)	NR	NR	*Not assessed*
GHQ-30; Psychiatric morbidity or disorder; 7/8						
Kitamura 1994a (first trimester) [4]	0.83 (NR)	0.71 (NR)	NR	NR	NR	*Not assessed*
Kitamura 1994a (third trimester)	0.39 (NR)	0.82 (NR)	NR	NR	NR	*Not assessed*
GHQ-28; Any psychiatric diagnosis; 3/4						
Aderibigbe 1992 (third trimester - 'conventional scoring')	0.75 (NR)	0.83 (NR)	NR	NR	NR	*Not assessed*
GHQ-28; Any psychiatric diagnosis; 7/8						
Aderibigbe 1992 (third trimester - 'revised scoring')	0.82 (NR)	0.85 (NR)	NR	NR	NR	*Not assessed*
GHQ-12; Psychiatric morbidity; 3						
Abiodun 1994 (all trimesters)	0.83 (NR)	0.81 (NR)	NR	NR	NR	*Not assessed*
GHQ-12; Depression, generalized anxiety, or panic disorder; 0/1						
Tran 2011 (third trimester; n=199)	0.81 (NR)	0.58 (NR)	0.76 (0.69, 0.83)	NR	NR	Moderate

Abbreviations: AUC, area under the curve; CI, confidence interval; GHQ, General Health Questionnaire; LR, likelihood ratio; NR, not reported.

[3] Based on QUADAS-2 assessment for this review.

Duplicate results reported in Kitamura 1989 and Kitamura 1994a, only most recent publication cited here.

Table B6-7 Evidence summary table for the GHQ for detection of anxiety in postnatal women

Version; Condition; Cut-off	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC (95% CI) [2]	LR+	LR-	Study Quality [3]
GHQ-30; Psychiatric morbidity or disorder; 4/5						
Kitamura 1994a (5 days postpartum)	0.44 (NR)	0.64 (NR)	NR	NR	NR	*Not assessed*
Kitamura 1994a (1 month postpartum)	0.69 (NR)	0.61 (NR)	NR	NR	NR	*Not assessed*
GHQ-30; Psychiatric disturbance; 4/5						
Nott 1982 (8-14 weeks postpartum)	0.92 (NR)	0.66 (NR)	NR	NR	NR	*Not assessed*
GHQ-30; Psychiatric morbidity or disorder; 7/8						
Kitamura 1994a (5 days postpartum)	0.28 (NR)	0.79 (NR)	NR	NR	NR	*Not assessed*
Kitamura 1994a (1 month postpartum)	0.50 (NR)	0.84 (NR)	NR	NR	NR	*Not assessed*
GHQ-12; Depression, anxiety or adjustment disorder; 4/5						
Navarro 2007 (6 weeks postpartum)	0.81 (0.74, 0.86)	0.80 (0.74, 0.85)	0.90 (0.88, 0.93)	NR	NR	*Not assessed*

Version; Condition; Cut-off	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC (95% CI) [2]	LR+	LR-	Study Quality [3]
GHQ-12; Psychiatric disturbance; 1/2						
Nott 1982 (8-14 weeks postpartum)	0.68 (NR)	0.97 (NR)	NR	NR	NR	*Not assessed*
GHQ-12; Depression, generalized anxiety, panic disorder; 0/1						
Tran 2011 (4-6 weeks postpartum; n=165)	0.73 (NR)	0.54 (NR)	0.69 (0.60, 0.78)	NR	NR	Moderate

Abbreviations: AUC, area under the curve; CI, confidence interval; GHQ, General Health Questionnaire; LR, likelihood ratio; NR, not reported.
[3] Based on QUADAS-2 assessment for this review.

B6.3.2.4 HADS

Table B6-8 Evidence summary table for the HADS for detection of anxiety in antenatal women

Version; Condition; Cut-off	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC (95% CI) [2]	LR+	LR-	Study Quality [3]
HADS-A; Psychiatric morbidity; 8						
Abiodun 1994 (all trimesters)	0.93 (NR)	0.90 (NR)	NR	NR	NR	*Not assessed*

Abbreviations: Abbreviations: AUC, area under the curve; CI, confidence interval; HADS, Hospital Anxiety and Depression Scale; LR, likelihood ratio; NR, not reported.
[3] Based on QUADAS-2 assessment for this review.

B6.3.2.5 K-10

Table B6-9 Evidence summary table for the K-10 for detection of anxiety in antenatal women

Version; Condition; Cut-off	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC (95% CI) [2]	LR+	LR-	Study Quality [3]
K-10; Current panic disorder; 38.5						
Spies 2009 (first & second trimesters)	0.50 (NR)	0.98 (NR)	0.71 (NR)	21.2	0.50	Low
K-10; Social anxiety disorder; 26.5						
Spies 2009 (first & second trimesters)	1.00 (NR)	0.75 (NR)	0.76 (NR)	4	0	Low
K-10; Current post-traumatic stress disorder; 28.5						
Spies 2009 (first & second trimesters)	0.50 (NR)	0.80 (NR)	0.69 (NR)	2.5	0.6	Low

Abbreviations: AUC, area under the receiver-operator curve; CI, confidence interval; K-10, Kessler 10 item questionnaire; LR, likelihood ratio; NR, not reported.
[3] Based on QUADAS-2 assessment for this review

B6.3.2.6 STAI

Table B6-10 Evidence summary table for the STAI for detection of anxiety in antenatal women

Version; Condition; Cut-off	Sensitivity (95% CI) [1]	Specificity (95% CI) [1]	AUC (95% CI) [2]	LR+	LR-	Study Quality
STAI; Anxiety; >30						
Grant 2008 (third trimester - state anxiety)	0.90 (NR)	0.44 (NR)	0.87 (0.78, 0.93)	NR	NR	High
Grant 2008 (third trimester - trait anxiety)	1.00 (NR)	0.49 (NR)	0.89 (0.81, 0.95)	NR	NR	High
STAI; Anxiety; >34						
Grant 2008 (third trimester - state anxiety)	0.90 (NR)	0.59 (NR)	0.87 (0.78, 0.93)	NR	NR	High
Grant 2008 (third trimester - trait anxiety)	0.95 (NR)	0.63 (NR)	0.89 (0.81, 0.95)	NR	NR	High
STAI; Anxiety; >38						
Grant 2008 (third trimester - state anxiety)	0.81 (NR)	0.76 (NR)	0.87 (0.78, 0.93)	NR	NR	High
Grant 2008 (third trimester - trait anxiety)	0.86 (NR)	0.71 (NR)	0.89 (0.81, 0.95)	NR	NR	High
STAI; Anxiety; >40						
Grant 2008 (third trimester - state anxiety)	0.81 (NR)	0.80 (NR)	0.87 (0.78, 0.93)	NR	NR	High
Grant 2008 (third trimester - trait anxiety)	0.81 (NR)	0.80 (NR)	0.89 (0.81, 0.95)	NR	NR	High
STAI; Anxiety; >42						
Grant 2008 (third trimester - state anxiety)	0.71 (NR)	0.89 (NR)	0.87 (0.78, 0.93)	NR	NR	High
Grant 2008 (third trimester - trait anxiety)	0.71 (NR)	0.84 (NR)	0.89 (0.81, 0.95)	NR	NR	High
STAI; Anxiety; >44						
Grant 2008 (third trimester - state anxiety)	0.62 (NR)	0.91 (NR)	0.87 (0.78, 0.93)	NR	NR	High
Grant 2008 (third trimester - trait anxiety)	0.71 (NR)	0.87 (NR)	0.89 (0.81, 0.95)	NR	NR	High

Abbreviations: AUC, area under the receiver-operator curve; CI, confidence interval; LR, likelihood ratio; NR, not reported; STAI, State-Trait Anxiety Inventory.

[3] Based on QUADAS-2 assessment for this review.

B6.3.3 Summary of findings regarding technical performance

Due to the paucity of data regarding the technical performance of anxiety screening tools in perinatal women, Summary findings are presented for all of the tools from the studies that met our inclusion criteria regardless of the quality of the correpsonding included studies: EPDS, EPDS-3, GAD-7, K-10 and STAI.

Results are grouped together below according to the population in the studies: a mixed population of antenatal and postnatal women, and only antenatal women. No included studies of anxiety screening were conducted in postnatal women only. Pooling of values has not been undertaken due to heterogeneity in study characteristics and cut-off values used.

B6.3.3.1 Anxiety screening in antenatal women

Table B6-11 Summary of Findings table for the K-10 for the detection of anxiety in antenatal women

Tool; Condition; Cut-off	No of studies[1] (participants)	Important outcomes		Critical outcomes			Overall certainty
		Sensitivity (95% CI)	Specificity (95% CI)	Positive Likelihood Ratio [1]	Negative Likelihood Ratio [1]	Pooled AUC	
K-10; panic disorder; 38.5	1 (129)	0.50 (NR)	0.98 (NR)	25.0	0.51	0.71 (NR)	●○○○ Very low [2]
K-10; social anxiety disorder; 26.5	1 (129)	1.00 (NR)	0.75 (NR)	3.96	0.01	0.76 (NR)	●○○○ Very low [2]
K-10; current PTSD; 28.5	1 (129)	0.50 (NR)	0.80 (NR)	2.50	0.63	0.69 (NR)	●○○○ Very low [2]
Evidence statement: *It is uncertain if the K-10 has adequate sensitivity or specificity to detect panic disorder, social anxiety disorder or current post-traumatic stress disorder in pregnant women (very low certainty evidence).*							
Footnotes: [1] Calculated from the pooled sensitivity and specificity using the following formulas: LR+ = sensitivity/(1-specificity); LR- = (1-sensitivity)/specificity, with rounding of values of 1.00 to 0.99 to avoid Div/0 error [2] Single study of low quality							

Abbreviations: AUC, area under the curve; CI, confidence interval; K-10, Kessler 10 item questionnaire; LR, likelihood ratio; NR, not reported.

Table B6-12 Summary of Findings table for the STAI for detection of trait anxiety in antenatal women

Tool; Condition; Cut-off	No of studies[1] (participants)	Important outcomes		Critical outcomes			Overall certainty
		Sensitivity (95% CI)	Specificity (95% CI)	Positive Likelihood Ratio [1]	Negative Likelihood Ratio [1]	AUC	
STAI-trait; any anxiety disorder; >30	1 (100)	1.00 (NR)	0.49 (NR)	1.94	0.02	0.89 (0.81, 0.95)	●●●○ Moderate [2]
STAI; trait anxiety; >34	1 (100)	0.95 (NR)	0.63 (NR)	2.57	0.08	0.89 (0.81, 0.95)	●●●○ Moderate [2]
STAI; trait anxiety; >38	1 (100)	0.86 (NR)	0.71 (NR)	2.97	0.20	0.89 (0.81, 0.95)	●●●○ Moderate [2]
STAI; trait anxiety; >40	1 (100)	0.81 (NR)	0.80 (NR)	4.05	0.24	0.89 (0.81, 0.95)	●●●○ Moderate [2]
STAI; trait anxiety; >42	1 (100)	0.71 (NR)	0.84 (NR)	4.44	0.35	0.89 (0.81, 0.95)	●●●○ Moderate [2]
STAI; trait anxiety; >44	1 (100)	0.71 (NR)	0.87 (NR)	5.46	0.33	0.89 (0.81, 0.95)	●●●○ Moderate [2]

Tool; Condition; Cut-off	No of studies[1] (participants)	Important outcomes		Critical outcomes			Overall certainty
		Sensitivity (95% CI)	Specificity (95% CI)	Positive Likelihood Ratio [1]	Negative Likelihood Ratio [1]	AUC	
Evidence statement: *A score of 40 or above on the STAI (trait version) has moderate sensitivity and moderate specificity to detect trait anxiety in pregnant women (moderate certainty evidence).*							
Footnotes: [1] Calculated from the pooled sensitivity and specificity using the following formulas: LR+ = sensitivity/(1-specificity); LR- = (1-sensitivity)/specificity, with rounding of values of 1.00 to 0.99 to avoid Div/0 error [2] Single study of high quality							

Abbreviations: AUC, area under the curve; CI, confidence interval; LR, likelihood ratio; NR, not reported; STAI, State-Trait Anxiety Inventory.

B6.3.3.2 Anxiety screening in a mixed population of antenatal and postnatal women

Table B6-13 Summary of Findings table for the full EPDS for detection of anxiety in a mixed population of antenatal and postnatal women

Tool; Condition; Cut-off	No of studies[1] (participants)	Important outcomes		Critical outcomes			Overall certainty
		Sensitivity (95% CI)	Specificity (95% CI)	Positive Likelihood Ratio [1]	Negative Likelihood Ratio [1]	Pooled AUC	
EPDS-full; GAD; 12/13	2 (331)	0.70 (NR) 0.89 (NR)	0.82 (NR) 0.40 (NR)	3.89 1.48	0.37 0.28	NR 0.62 (NR)	●●○○ Low [2]
EPDS-full; GAD and MDD; 17	1 (240)	0.67 (NR)	0.65 (NR)	1.91	0.51	0.68 (NR)	●○○○ Very low [3]
Evidence statement: *It is uncertain if the EPDS-full version has adequate sensitivity or specificity to detect anxiety disorder in pregnant or postpartum women (low to very low certainty evidence).*							
Footnotes: [1] Calculated from the pooled sensitivity and specificity using the following formulas: LR+ = sensitivity/(1-specificity); LR- = (1-sensitivity)/specificity [2] Two studies of very low quality [3] One study of very low quality							

Abbreviations: AUC, area under the curve; CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; LR, likelihood ratio; NR, not reported.

Table B6-14 Summary of Findings table for the EPDS-3A for detection of anxiety in a mixed population of antenatal and postnatal women

Tool; Condition; Cut-off	No of studies[1] (participants)	Important outcomes		Critical outcomes			Overall certainty
		Sensitivity (95% CI)	Specificity (95% CI)	Positive Likelihood Ratio [1]	Negative Likelihood Ratio [1]	Pooled AUC	
EPDS-3A; GAD; >4	1 (91)	0.88 (NR)	0.49 (NR)	1.73	0.24	NR	●○○○ Very low [2]
EPDS-3A; GAD; 7	1 (240)	0.68 (NR)	0.64 (NR)	1.89	0.50	NR	●○○○ Very low [2]
EPDS-3A; GAD and MDD; 7	1 (240)	0.59 (NR)	0.67 (NR)	1.79	0.61	NR	●○○○ Very low [2]
Evidence statement: *It is uncertain if the EPDS-3A version has adequate sensitivity or specificity to detect anxiety disorder in pregnant or postpartum women (very low certainty evidence)*							
Footnotes: [1] Calculated from the pooled sensitivity and specificity using the following formulas: LR+ = sensitivity/(1-specificity); LR- = (1-sensitivity)/specificity [2] Single study of very low quality							

Abbreviations: AUC, area under the curve; CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; GAD, Generalised Anxiety Disorder; LR, likelihood ratio; MDD, major depressive disorder; NR, not reported

Table B6-15 Summary of Findings table for the GAD-7 for detection of anxiety in a mixed population of antenatal and postnatal women

Tool; Condition; Cut-off	No of studies[1] (participants)	Important outcomes		Critical outcomes			Overall certainty
		Sensitivity (95% CI)	Specificity (95% CI)	Positive Likelihood Ratio [1]	Negative Likelihood Ratio [1]	Pooled AUC	
GAD-7; GAD; 10	1 (240)	0.76 (NR)	0.52 (NR)	1.58	0.46	0.71 (NR)	●○○○ Very low [2]
GAD-7; GAD; 13	1 (240)	0.61 (NR)	0.73 (NR)	2.26	0.53	0.71 (NR)	●○○○ Very low [2]
GAD-7; GAD and MDD; 13	1 (240)	0.67 (NR)	0.69 (NR)	2.16	0.48	0.74 (NR)	●○○○ Very low [2]
Evidence statement: *It is uncertain if the GAD-7 has adequate sensitivity or specificity to detect anxiety disorder in pregnant or postpartum women (very low quality evidence).*							
Footnotes: [1] Calculated from the pooled sensitivity and specificity using the following formulas: LR+ = sensitivity/(1-specificity); LR- = (1-sensitivity)/specificity [2] Single study of very low quality							

Abbreviations: AUC, area under the curve; CI, confidence interval; GAD, Generalised Anxiety Disorder; LR, likelihood ratio; MDD, major depressive disorder; NR, not reported.

B6.4 NON-TECHNICAL CHARACTERISTICS OF RELEVANT INCLUDED TOOLS

The table below summaries the non-technical characteristics of the included anxiety screening tools for which there was evidence of technical performance (regardless of the certainty of that evidence). The complexity of scoring for each tool has been assessed as Simple, Moderate or High on the basis of information in the published literature and the experience of the EWG.

Table B6-16 Non-technical characteristics of the relevant included tools for anxiety screening

Tool	Number of items	Time to administer (mins)	Complexity of scoring	Available languages
EPDS	10	5-10 mins	Simple	Developed in English and translated into >50 languages Validated in English for anxiety screening
EPDS-3	3	<5 mins	Simple	English
GAD-7	7	5-10 mins	Simple	English
K-10	10	5-10 mins	Simple	English
STAI	20	<10 mins	Complex	English

Abbreviations: EPDS, Edinburgh Postnatal Depression Scale; GAD, Generalised Anxiety Disorder; K-10, Kessler 10; STAI, State Trait Anxiety Index

B6.5 CLINICAL USEFULNESS OF RELEVANT TOOLS

None of the individual studies identified in the supplementary search for anxiety screening tools (Grigoriadis 2011, Simpson 2014, Tran 2011) reported on acceptability, effectiveness or implementation of the specific tools. No SRs were identified that specifically related to anxiety screening, as this is often undertaken alongside depression screening. The reader is referred to the discussion of the clinical usefulness of depression screening (see above).

In addition to the information provided above, it should be noted that other studies (excluded because they did not provide evidence of technical performance) discuss the development of potentially useful perinatal anxiety screening tools.

A study from Australia describes the use of the Perinatal Anxiety Screening Scale (PASS; **Somerville 2015**), a 31-item self-report questionnaire with four sub-scales that measure general worry and specific fears, perfectionism, control and trauma, social anxiety, and acute anxiety and adjustment. This scale has been developed to assess severity of anxiety from minimal to severe anxiety, with the intention of monitoring fluctuating levels of anxiety during the pregnancy and after the infant is born. The authors argue that current screening processes that use a binary classification system (eg, 'at risk' versus 'not at risk' of anxiety) are likely to miss subtle but important changes in anxiety levels. The authors suggest that use of the PASS could supplement current perinatal mental health screening and aid decisions regarding appropriate services for referral and urgency of care.

A recent study from Canada describes the development of a short version of the STAI (**Bayrampour 2014**). This study was excluded from our assessment of technical performance as it did not meet our inclusion criteria, but it does describe initial work to validate the psychometric properties of three different 6-item forms of the STAI, that could potentially be more useful than the full STAI in a time-constrained clinical setting.

B6.6 OVERALL SUMMARY OF FINDINGS

An overall assessment of the technical performance, non-technical characteristics and clinical usefulness of anxiety screening tools is presented in **Table B6-17.**

Table B6-17 Overall Summary of Findings related to the use of perinatal anxiety screening tools

Tool(s)	Technical characteristics		Non-technical characteristics		Clinical usefulness		
	Performance[1]	**Certainty** [2]	**Ease of Administration**[3]	**Language availability**[4] **& cultural sensitivity**[5]	**Acceptability**[6]	**Effectiveness**[7]	**Implementability**[8]
EPDS	Perinatal: Uncertain	●●○○ Low	High	Multiple languages[9] Multiple populations[9]	High[9]	Unknown	High
EPDS-3	Perinatal: Uncertain	●○○○ Very low	High	English Western populations	Unknown But likely to be Good	Unknown	High
GAD-7	Perinatal: Uncertain	●○○○ Very low	High	English Western populations	Unknown But likely to be Good	Unknown	Moderate
K-10	Antenatal: Uncertain	●○○○ Very low	High	English Western populations	Unknown But likely to be Good	Unknown	High
STAI	Antenatal: Acceptable	●●●○ Moderate	Low	English Western populations	Unknown But likely to be Good	Unknown	Low

Footnotes

[1] Performance defined as sensitivity, specificity, positive likelihood ratio, negative likelihood ratio (defined as Acceptable, Limited, or Uncertain).

[2] Certainty assessed according to GRADE and QUADAS-2 criteria (defined as High, Moderate, Low or very Low).

[3] Ease of administration was based on judgment regarding the number of items, and the time and complexity of administering and scoring the tool (rated as High, Moderate, or Low)

[4] Language availability based on information from the included literature and the awareness of the EWG

[5] Cultural sensitivity was based on information from the included literature of any use in culturally and linguistically diverse populations

[6] Acceptability was based on the overall judgement of the EWG of the acceptability of each tool to women, health care professionals and/or the general public (rated as High, Moderate, Low or Unknown)

[7] Effectiveness was defined as positive impact on anxiety, services referred to or utilized, and impact on a woman's mental health (rated as High, Good, Limited, or Unknown)

[8] Implementability was based on the overall judgement of the EWG based on available information regarding the training requirements for use of the tool and implications for current models of care and staff and service availability

[9] Inferred from evidence of depression screening

Abbreviations: EPDS, Edinburgh Postnatal Depression Scale; EWG, Expert Working Group; GAD, Generalised Anxiety Disorder; K-10, Kessler 10; STAI, State Trait Anxiety Index.

B7 REFERENCES

Austin, M. P., D. Hadzi-Pavlovic, K. Saint and G. Parker (2005). "Antenatal screening for the prediction of postnatal depression: validation of a psychosocial Pregnancy Risk Questionnaire." Acta Psychiatr Scand 112(4): 310-317.

Austin, M. P., J. Colton, S. Priest, N. Reilly and D. Hadzi-Pavlovic (2013). "The antenatal risk questionnaire (ANRQ): acceptability and use for psychosocial risk assessment in the maternity setting." Women Birth 26(1): 17-25.

Bayrampour, H., S. McDonald, T. Fung and S. Tough (2014). "Reliability and validity of three shortened versions of the State Anxiety Inventory scale during the perinatal period." J Psychosom Obstet Gynaecol 35(3): 101-107.

Bernazzani, O., M. N. Marks, A. Bifulco, K. Siddle, P. Asten and S. Conroy (2005). "Assessing psychosocial risk in pregnant/postpartum women using the Contextual Assessment of Maternity Experience (CAME)--recent life adversity, social support and maternal feelings." Soc Psychiatry Psychiatr Epidemiol 40(6): 497-508.

Brealey, S. D., C. Hewitt, J. M. Green, J. Morrell and S. Gilbody (2010). "Screening for postnatal depression - is it acceptable to women and healthcare professionals? A systematic review and meta-synthesis." Journal of Reproductive and Infant Psychology 28(4): 328-344.

Carroll, J. C., A. J. Reid, A. Biringer, D. Midmer, R. H. Glazier, L. Wilson, J. A. Permaul, P. Pugh, B. Chalmers, F. Seddon and D. E. Stewart (2005). "Effectiveness of the Antenatal Psychosocial Health Assessment (ALPHA) form in detecting psychosocial concerns: a randomized controlled trial." Cmaj 173(3): 253-259.

El-Den, S., C. L. O'Reilly and T. F. Chen (2015). "A systematic review on the acceptability of perinatal depression screening." Journal of Affective Disorders 188: 284-303.

Gibson, J., K. McKenzie-Mcharg, J. Shakespeare, J. Price and R. Gray (2009). "A systematic review of studies validating the Edinburgh Postnatal Depression Scale in antepartum and postpartum women." Acta Psychiatrica Scandinavica 119(5): 350-364.

Goldin Evans, M., S. Phillippi and R. E. Gee (2015). "Examining the Screening Practices of Physicians for Postpartum Depression: Implications for Improving Health Outcomes." Women's Health Issues 25(6): 703-710.

Grant, K. A., C. McMahon and M. P. Austin (2008). "Maternal anxiety during the transition to parenthood: a prospective study." J Affect Disord 108(1-2): 101-111.

Grigoriadis, S., D. de Camps Meschino, E. Barrons, L. Bradley, A. Eady, A. Fishell, L. Mamisachvili, G. S. Cook, M. O'Keefe, S. Romans and L. E. Ross (2011). "Mood and anxiety disorders in a sample of Canadian perinatal women referred for psychiatric care." Arch Womens Ment Health 14(4): 325-333.

Hewitt CE, Gilbody SM, Brealey S, Paulden M, Palmer S, Mann R, et al. Methods to identify postnatal depression in primary care: an integrated evidence synthesis and value of information analysis. Health Technol Assess 2009;13(36).

Hewitt, C. E. and S. M. Gilbody (2009). "Is it clinically and cost effective to screen for postnatal depression: A systematic review of controlled clinical trials and economic evidence." BJOG: An International Journal of Obstetrics and Gynaecology 116(8): 1019-1027.

Hewitt, C. E., S. M. Gilbody, R. Mann and S. Brealey (2010). "Instruments to identify post-natal depression: Which methods have been the most extensively validated, in what setting and in which language?" International Journal of Psychiatry in Clinical Practice 14(1): 72-76.

Hewitt, C. E., S. M. Gilbody, S. Brealey, M. Paulden, S. Palmer, R. Mann, J. Green, J. Morrell, M. Barkham, K. Light and D. Richards (2009). "Methods to identify postnatal depression in primary care: An integrated evidence synthesis and value of information analysis." Health Technology Assessment 13(36): iii-206.

Johnson, M., V. Schmeid, S. J. Lupton, M. P. Austin, S. M. Matthey, L. Kemp, T. Meade and A. E. Yeo (2012). "Measuring perinatal mental health risk." Archives of Women's Mental Health 15(5): 375-386.

Kingston D, Austin M-P, Hegadoren K, McDonald S, Lasiuk G, McDonald S, Heaman M, Biringer A, Sword W, Giallo R, Patel T, Lane-Smith M, van Zanten SV (2014). Study protocol for a randomized, controlled, superiority trial comparing the clinical and cost-effectiveness of integrated online mental health assessment-referral-care in pregnancy to usual prenatal care on prenatal and postnatal mental health and infant health and development: the Integrated Maternal Psychosocial Assessment to Care Trial (IMPACT). Trials. 15:72.

Kingston D, Kingston D, McDonald S, Biringer A, Austin MP, McDonald SD, Giallo R, Ohinmaa A, Lasiuk G, MacQueen G, Van Zanten S (2015). Comparing the acceptability, clinical-, and cost-effectiveness of mental health E-screening to paper-based screening in pregnant women: a randomized, parallel-group, superiority trial. Eur Psychiatry. 30(Supplement 1):986.

Kohlhoff, J., R. Hickinbotham, C. Knox, V. Roach and B. Barnett Am (2016). "Antenatal psychosocial assessment and depression screening in a private hospital." Aust N Z J Obstet Gynaecol 56(2): 173-178.

Kozinszky, Z. and R. B. Dudas (2015). "Validation studies of the Edinburgh Postnatal Depression Scale for the antenatal period." Journal of Affective Disorders 176: 95-105.

Mann, R. and S. Gilbody (2011) "Validity of two case finding questions to detect postnatal depression: a review of diagnostic test accuracy (Structured abstract)." Journal of Affective Disorders 133, 388-397.

Matthey, S., J. Phillips, T. White, P. Glossop, U. Hopper, P. Panasetis, A. Petridis, M. Larkin and B. Barnett (2004). "Routine psychosocial assessment of women in the antenatal period: frequency of risk factors and implications for clinical services." Arch Womens Ment Health 7(4): 223-229.

Matthey, S., K. Souter, K. Mortimer, C. Stephens and A. Sheridan-Magro (2016). "Routine antenatal maternal screening for current mental health: evaluation of a change in the use of the Edinburgh Depression Scale in clinical practice." Arch Womens Ment Health 19(2): 367-372.

Meades, R. and S. Ayers (2011). "Anxiety measures validated in perinatal populations: A systematic review." Journal of Affective Disorders 133(1-2): 1-15.

Myers, E. R., N. Aubuchon-Endsley, L. A. Bastian, J. M. Gierisch, A. R. Kemper, G. K. Swamy, M. F. Wald, A. J. McBroom, K. R. Lallinger, R. N. Gray, C. Green and G. D. Sanders (2013) "Efficacy and safety of screening for postpartum depression (Provisional abstract)." Database of Abstracts of Reviews of Effects.

Nast, I., M. Bolten, G. Meinlschmidt and D. H. Hellhammer (2013). "How to measure prenatal stress? A systematic review of psychometric instruments to assess psychosocial stress during pregnancy." Paediatric and Perinatal Epidemiology 27(4): 313-322.

NICE (2015) National Collaborating Centre for Mental Health. Antenatal and Postnatal Mental Health: the NICE guideline on Clinical Management and Service Guidance. National Clinical Guideline Number 192: 1-922.

O'Connor, E., R. C. Rossom, M. Henninger, H. C. Groom, B. U. Burda, J. T. Henderson, K. D. Bigler and E. P. Whitlock (2016). Screening for depression in adults: an updated systematic evidence review for the U.S. Preventive Services Task Force. Evidence Synthesis No. 128. AHRQ Publication No. 14-05208-EF-1. Rockville, MD, Agency for Healthcare Research and Quality.

Paulden M, Palmer S, Hewitt C, Gilbody S (2009). Screening for postnatal depression in primary care: cost effectiveness analysis. BMJ. 340:b5203.

Quispel, C., T. A. Schneider, G. J. Bonsel and M. P. Lambregtse-van den Berg (2012). "An innovative screen-and-advice model for psychopathology and psychosocial problems among urban pregnant women: an exploratory study." J Psychosom Obstet Gynaecol 33(1): 7-14.

Reilly, N., C. Yin, L. Monterosso, S. Bradshaw, K. Neale, B. Harrison and M. P. Austin (2015). "Identifying psychosocial risk among mothers in an Australian private maternity setting: A pilot study." Aust N Z J Obstet Gynaecol 55(5): 453-458.

Reuland, D. S., A. Cherrington, G. S. Watkins, D. W. Bradford, R. A. Blanco and B. N. Gaynes (2009). "Diagnostic accuracy of Spanish language depression-screening instruments." Annals of Family Medicine 7(5): 455-462.

Rollans, M., V. Schmied, L. Kemp and T. Meade (2013). "Digging over that old ground: an Australian perspective of women's experience of psychosocial assessment and depression screening in pregnancy and following birth." BMC Womens Health 13: 18.

Shrestha, S. D., R. Pradhan, T. D. Tran, R. C. Gualano and J. R. W. Fisher (2016). "Reliability and validity of the Edinburgh Postnatal Depression Scale (EPDS) for detecting perinatal common mental disorders (PCMDs) among women in low-and lower-middle-income countries: A systematic review." BMC Pregnancy and Childbirth 16(1).

Simpson, W., M. Glazer, N. Michalski, M. Steiner and B. N. Frey (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." Can J Psychiatry 59(8): 434-440.

Somerville, S., S. L. Byrne, K. Dedman, R. Hagan, S. Coo, E. Oxnam, D. Doherty, N. Cunningham and A. C. Page (2015). "Detecting the severity of perinatal anxiety with the Perinatal Anxiety Screening Scale (PASS)." J Affect Disord 186: 18-25.

Spies, G., D. J. Stein, A. Roos, S. C. Faure, J. Mostert, S. Seedat and B. Vythilingum (2009). "Validity of the Kessler 10 (K-10) in detecting DSM-IV defined mood and anxiety disorders among pregnant women." Arch Womens Ment Health 12(2): 69-74.

Thombs, B. D., E. Arthurs, S. Coronado-Montoya, M. Roseman, V. C. Delisle, A. Leavens, B. Levis, L. Azoulay, C. Smith, L. Ciofani, J. C. Coyne, N. Feeley, S. Gilbody, J. Schinazi, D. E. Stewart and P. Zelkowitz (2014). "Depression screening and patient outcomes in pregnancy or postpartum: A systematic review." Journal of Psychosomatic Research 76(6): 433-446.

Tran, T. D., T. Tran, B. La, D. Lee, D. Rosenthal and J. Fisher (2011). "Screening for perinatal common mental disorders in women in the north of Vietnam: a comparison of three psychometric instruments." J Affect Disord 133(1-2): 281-293.

Walker, L. O., J. Gao and B. Xie (2015). "Postpartum Psychosocial and Behavioral Health: A Systematic Review of Self-Administered Scales Validated for Postpartum Women in the United States." Women's Health Issues 25(5): 586-600.

Wilkinson A, Anderson S, Wheeler SB (2017). Screening for and treating postpartum dpression and psychosis: A cost-effectiveness analysis. Matern Child Health J. 21;903-14.

B8 APPENDICES

B8.1 DETAILS OF SEARCHES FOR PSYCHOSOCIAL ASSESSMENT AND SCREENING TOOLS

B8.1.1 Search strings

B8.1.1.1 Systematic review search

Database/date	Search #	Search string	Results
Embase.com (MEDLINE, Embase) 01 Jun 2016	1	((pregnancy:ab,ti OR pregnant:ab,ti) OR (perinatal:ab,ti OR 'peri natal':ab,ti) OR (prenatal:ab,ti OR 'pre natal':ab,ti) OR (postnatal:ab,ti OR 'post natal':ab,ti) OR (postpartum:ab,ti OR 'post partum':ab,ti) OR (antenatal:ab,ti OR 'ante natal':ab,ti) OR puerper*:ab,ti OR maternal:ab,ti) AND ((depression:ab,ti OR depressive:ab,ti OR depressed:ab,ti) OR anxiety:ab,ti OR (psychosis:ab,ti OR psychotic:ab,ti) OR bipolar:ab,ti OR psychosocial:ab,ti) AND (('systematic review'/exp OR 'systematic review':ab,ti OR 'systematic literature review':ab,ti OR 'systematic literature search':ab,ti OR 'systematic search':ab,ti) OR ('meta analysis'/exp OR 'meta analysis':ab,ti OR metaanalysis:ab,ti) OR 'pooled analysis':ab,ti OR 'evidence synthesis':ab,ti) Limit 2009 to date	803
Cochrane Library (CDSR, DARE and HTA) 29 Jul 2016	1	(pregnancy OR pregnant) OR (perinatal OR 'peri natal') OR (prenatal OR 'pre natal') OR (postnatal OR 'post natal') OR (postpartum OR 'post partum') OR (antenatal OR 'ante natal') OR puerper* OR maternal in Title, Abstract, Keywords AND (depression OR depressive OR depressed) OR anxiety OR (psychosis OR psychotic) OR bipolar OR psychosocial OR (schizophrenia OR schizophrenic) OR "borderline personality disorder") Limit 2009 to date	153

Abbreviations: CDSR, Cochrane Database of Systematic Reviews; DARE, Database of Abstracts of Reviews of Effect; HTA, Health Technology Assessment database.

B8.1.1.2 Psychosocial assessment search

Database/date	Search #	Search string	Results
Embase, Medline, PsycInfo 15 Dec 2016	1	1 perinatal.ti,ab,kw,ot,sh,tw. (148697) 2 postnatal.ti,ab,kw,ot,sh,tw. (220691) 3 antenatal.ti,ab,kw,ot,sh,tw. (70585) 4 pregnan*.ti,ab,kw,ot,sh,tw. (1828303) 5 1 or 2 or 3 or 4 (2061750) 6 depression.ti,ab,kw,ot,sh,tw. (944636) 7 anxiety.ti,ab,kw,ot,sh,tw. (521244) 8 mental health.ti,ab,kw,ot,sh,tw. (408106) 9 mental disorder*.ti,ab,kw,ot,sh,tw. (272364) 10 6 or 7 or 8 or 9 (1724030) 11 assessment.ti,ab,kw,ot,sh,tw. (1962727) 12 psychological test*.ti,ab,kw,ot,sh,tw. (45796) 13 risk.ti,ab,kw,ot,sh,tw. (4325923) 14 screen*.ti,ab,kw,ot,sh,tw. (1495942) 15 questionnaire*.ti,ab,kw,ot,sh,tw. (1224545) 16 instrument*.ti,ab,kw,ot,sh,tw. (838057) 17 tool*.ti,ab,kw,ot,sh,tw. (1336184) 18 11 or 12 or 13 or 14 or 15 or 16 or 17 (9295418) 19 5 and 10 and 18 (31469) 20 'antenatal risk questionnaire'.af. (44) 21 'antenatal psychosocial health assessment'.af. (68) 22 'Australian routine psychosocial assessment'.af. (1) 23 ('Camberwell assessment of need' adj2 mothers').af. (18) 24 'Pregnancy risk questionnaire'.af. (64) 25 'Postnatal risk questionnaire'.af. (13) 26 'Contextual assessment of maternity experience'.af. (27) 27 20 or 21 or 22 or 23 or 24 or 25 or 26 (180) 28 'Risk factor assessment'.af. (1845) 29 ANRQ.ti,ab,kw,ot,sh,tw. (16) 30 ARPA.ti,ab,kw,ot,sh,tw. (271) 31 CAN-M.ti,ab,kw,ot,sh,tw. (20) 32 PRQ.ti,ab,kw,ot,sh,tw. (286) 33 PNRQ.ti,ab,kw,ot,sh,tw. (8) 34 RFA.ti,ab,kw,ot,sh,tw. (14370)	416

		35 28 or 29 or 30 or 31 or 32 or 33 or 34 (16806) 36 5 and 10 and 35 (45) 37 19 or 27 or 36 (31520) 38 limit 37 to yr="2011 -Current" (17190) 39 (conference abstract or conference poster).pt,sh,ti,tw. (2408391) 40 38 not 39 (14243) 41 limit 40 to english language (13585) 42 limit 41 to human (11955) 43 screening.ti,sh,tw. (1049514) 44 assessment.ti,sh,tw. (1934149) 45 questionnaire.ti,sh,tw. (1087130) 46 instrument.ti,sh,tw. (257146) 47 tool.ti,sh,tw. (899821) 48 psychological test.ti,sh,tw. (1871) 49 43 or 44 or 45 or 46 or 47 or 48 (4601157) 50 19 and 49 (15878) 51 *Depression/ (195372) 52 *Anxiety/ (115597) 53 *Mental Health/ (86859) 54 *Mental Disorders/ (177416) 55 51 or 52 or 53 or 54 (541717) 56 50 and 55 (4164) 57 exp Diagnosis/ (13785027) 58 56 and 57 (982) 59 *Pregnancy/ (207423) 60 58 and 59 (158) 61 27 or 36 or 60 (359) 62 remove duplicates from 61 (299) 63 limit 62 to yr="2011 -Current" (171) 64 *Anxiety Disorder/di (6156) 65 *Anxiety Disorders/di (4082) 66 *Depression/di (24413) 67 *Mental Disease/di (12856) 68 *Puerperal Depression/di (1039) 69 *Depression, Postpartum/di (1432) 70 *Mental Disorders/di (16396) 71 64 or 65 or 66 or 67 or 68 or 69 or 70 (56085) 72 5 and 71 (2499) 73 Screening Test.ti,sh. (69426) 74 Screening.ti,sh. (529219) 75 Assessment.ti,sh. (540086) 76 Functional Assessment.ti,sh. (60786) 77 Scoring System.ti,sh. (218209) 78 Rating Scale.ti,sh. (113085) 79 Severity of Illness Index.ti,sh. (225525) 80 Questionnaire.ti,sh. (607105) 81 Risk Assessment.ti,sh. (659086) 82 Surveys.af. or Questionnaires.sh. (416) 83 Psychometrics.sh. (107518) 84 Psychologic test.ti,sh. (38240) 85 Instrument.ti,sh. (36755) 86 73 or 74 or 75 or 76 or 77 or 78 or 79 or 80 or 81 or 82 or 83 or 84 or 85 (2811640) 87 72 and 86 (1344) 88 limit 87 to yr="2011 -Current" (505) 89 limit 88 to english language (489) 90 limit 89 to human (489) 91 27 or 36 or 90 (683) 92 limit 91 to yr="2011 -Current" (589) 93 limit 92 to english language (587) 94 limit 93 to human (585) 95 remove duplicates from 94 (439) 96 95 not 39 (430) 97 (note or letter or comment or news or editorial).pt,sh. (3979929) 98 96 not 97 (416) 99 from 98 keep 1- 416	

CINAHL 15 Dec 2016	2	S25 S24 Limiters - Published Date: 20110101-20161231; Exclude MEDLINE records; Limited to publication type Academic Journals 74 (keep records 1-74) S24 S6 OR S14 OR S23 990 S23 S1 AND S2 AND S22 22 S22 S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 1,676 S21 RFA 369 S20 PNRQ 2 S19 PRQ 64 S18 CAN-M 0 S17 ARPA 1 S16 ANRQ 1 S15 Risk factor assessment 1,239 S14 S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 99 S13 Contextual assessment of maternity experience 0 S12 Postnatal risk questionnaire 1 S11 Pregnancy risk questionnaire 20 S10 Camberwell assessment of need 74 S9 Australian routine psychosocial assessment 0 S8 Antenatal psychosocial health assessment 3 S7 Antenatal risk questionnaire 2 S6 S4 AND S5 877 S5 TI PREGNANCY 21,659 S4 S1 AND S2 AND S3 5,055 S3 assessment or psychological test* or risk or screen* or questionnaire* or instrument* or tool* 876,734 S2 depression or anxiety or mental health or mental disorder* 175,151 S1 perinatal or postnatal or antenatal or pregnan* 126,808	74

B8.1.1.3 Anxiety screening search

Database/date	Search #	Search string	Results
Embase, Medline, PyscInfo 21 Dec 2016	1	1 perinatal.ti,ab,kw,ot,sh,tw. (148832) 2 postnatal.ti,ab,kw,ot,sh,tw. (220860) 3 antenatal.ti,ab,kw,ot,sh,tw. (70668) 4 pregnan*.ti,ab,kw,ot,sh,tw. (1829440) 5 1 or 2 or 3 or 4 (2063077) 6 exp Anxiety/ (294702) 7 anxiety.ti,ab,kw,ot,sh,tw. (521915) 8 exp anxiety disorder/ (318501) 9 6 or 7 or 8 (713367) 10 'Edinburgh Postnatal Depression Scale'.ti,ab,kw,ot,sh,tw. (5399) 11 EPDS.ti,ab,kw,ot,sh,tw. (4208) 12 'Kessler-10'.ti,ab,kw,ot,sh,tw. (466) 13 K-10.ti,ab,kw,ot,sh,tw. (2054) 14 'Generali*ed Anxiety Disorder'.ti,ab,kw,ot,sh,tw. (21065) 15 GAD.ti,ab,kw,ot,sh,tw. (21588) 16 GAD-2.ti,ab,kw,ot,sh,tw. (161) 17 'General Health Questionnaire'.ti,ab,kw,ot,sh,tw. (11752) 18 GHQ.ti,ab,kw,ot,sh,tw. (8192) 19 'State-Trait Anxiety Inventory'.ti,ab,kw,ot,sh,tw. (12874) 20 STAI.ti,ab,kw,ot,sh,tw. (8564) 21 'Hospital Anxiety.mp. and Depression Scale*'.ti,ab,kw,ot,sh,tw. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fs, nm, kf, px, rx, an, eu, pm, ui, tc, id, tm] (19017) 22 'HADS-A'.ti,ab,kw,ot,sh,tw. (1115) 23 DASS-21.ti,ab,kw,ot,sh,tw. (858) 24 'Depression Anxiety Stress Scale* 21'.ti,ab,kw,ot,sh,tw. (272) 25 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 (91091) 26 5 and 9 and 25 (3334) 27 limit 26 to yr="2010 -Current" (2231) 28 *Anxiety/ (115735) 29 27 and 28 (774) 30 limit 29 to english language (750) 31 remove duplicates from 30 (519) 32 (comment or letter or note or short survey or editorial or conference abstract or conference poster).pt,sh. (6763939) 33 31 not 32 (401) 34 from 33 keep 1-401 (401)	401

CINAHL 21 Dec 2016	2	S23 S1 AND S6 AND S22 Limiters - Published Date: 20100101-20161231; Exclude MEDLINE records 131 S22 S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 262,299 S21 DASS-21 61 S20 'Depression Anxiety Stress Scales 21' 67 S19 'HADS-A' 256,550 S18 'Hospital Anxiety and Depression Scale* 1,587 S17 STAI 516 S16 'State-Trait Anxiety Inventory' 3,388 S15 GHQ 618 S14 'General Health Questionnaire' 1,075 S13 GAD-2 15 S12 GAD 611 S11 'Generali*ed Anxiety Disorder' 808 S10 K-10 73 S9 Kessler-10 44 S8 EPDS 369 S7 Edinburgh Postnatal Depression Scale 1,184 S6 S2 OR S3 OR S4 OR S5 127,128 S5 pregnan* 116,065 S4 antenatal* 5,239 S3 postnatal* 12,213 S2 perinatal* 12,213 S1 anxiety 39,298	131

B8.1.1.4 Economic search

A literature search was conducted to identify economic analyses of screening for perinatal depression using the EPDS. Hand searching of other EndNote libraries for this project was also conducted.

Database/date	Search #	Search string	Results
Ovid MEDLINE and Embase 9 May 2017	1	*economics/	29244
	2	exp "costs and cost analysis"/	513349
	3	(economic adj2 model*).mp.	15752
	4	(cost minimi* or cost-utilit* or health utilit* or economic evaluation* or economic review* or cost outcome or cost analys?s or economic analys?s or budget* impact analys?s).mp.	114439
	5	(cost-effective* or pharmacoeconomic* or pharmaco-economic* or cost-benefit or costs).ti,kf,kw.	144759
	6	(life year or life years or qaly* or cost-benefit analys?s or cost-effectiveness analys?s).ab,kf,kw.	57210
	7	(cost or economic*).ti,kf,kw.	284512
	8	(costs or cost-effectiveness or markov).ab.	454726
	9	7 and 8	112403
	10	1 or 2 or 3 or 4 or 5 or 6 or 9	660108
	11	("Edinburgh Postnatal Depression Scale" or EPDS).mp.	5293
	12	10 and 11	**56**
Ovid PsychINFO 9 May 2017	1	*economics/	12632
	2	exp "costs and cost analysis"/	22831
	3	(economic adj2 model*).mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fx, nm, kf, px, rx, an, eu, pm, ui, sy]	1261
	4	(cost minimi* or cost-utilit* or health utilit* or economic evaluation* or economic review* or cost outcome or cost analys?s or economic analys?s or budget* impact analys?s).mp.	17365
	5	(cost-effective* or pharmacoeconomic* or pharmaco-economic* or cost-benefit or costs).ti,kf,kw.	7071
	6	(life year or life years or qaly* or cost-benefit analys?s or cost-effectiveness analys?s).ab,kf,kw.	3169
	7	(cost or economic*).ti,kf,kw.	19069
	8	(costs or cost-effectiveness or markov).ab.	43654
	9	7 and 8	5309
	10	1 or 2 or 3 or 4 or 5 or 6 or 9	40744
	11	("Edinburgh Postnatal Depression Scale" or EPDS).mp.	2756
	12	10 and 11	**10**
Cochrane Library 18 May 2017	1	(perinatal or antenatal or ante natal or postnatal or post natal or (post and partum) or post partum or pregnancy or pregnan* or puerperal disorders or puerperal or post partum period or puerperium) in Title, Abstract, Keywords AND (depression or depressive or anxiety or "bipolar disorder" or schizophrenia or "mental health" or "mood disorder" or "mood disorders") in Title, Abstract, Keywords	2328
	2	("edinburgh postnatal depression scale" or EPDS) in All Text	378
	3	1 or 2	2369
	4	3 in Economic Evaluations	14

	5	3 in Technology Assessments	17
	6	4 and 5	**31**

B8.1.2 Exclusion of studies

B8.1.2.1 Systematic review search

	Status	No. citations excluded	No. citations included
Identified via literature search			*805*
Identified manually[1]			*5*
Duplicate citation		*92*	
TOTAL	**Included**		**718**
Title/abstract	Excluded	548	
TOTAL	**Included**		**170**
Full paper	Excluded – wrong population	14	
	Excluded – wrong indication	7	
	Excluded – wrong intervention	20	
	Excluded – wrong outcomes	8	
	Excluded – not in English	1	
	Excluded – duplicate data	3	
	Excluded – not a SR	30	
	Excluded – wrong study type	4	
	Excluded – superseded	1	
TOTAL	**Included**		**82**[2]
TOTAL	**Relevant to psychosocial assessment or screening**		**19**

B8.1.2.2 Psychosocial assessment search

	Status	No. citations excluded	No. citations included
Identified via literature searches			*490*
Duplicate citation		*0*	
Excluded – title/abstract		*444*	
TOTAL	**Full text retrieved**		**46**
Full paper	Excluded – wrong population	0	
	Excluded – wrong indication	3	
	Excluded – wrong intervention	4	
	Excluded – wrong outcomes	1	
	Excluded – wrong study type (incl SR)	32	
TOTAL	**Included**		**5**

B8.1.2.3 Anxiety screening search

	Status	No. citations excluded	No. citations included
Identified via literature searches			*532*
Duplicate citation		*43*	
TOTAL	**Included**		**489**
Full paper	Excluded – wrong population	5	
	Excluded – wrong indication	0	
	Excluded – wrong intervention	15	
	Excluded – wrong outcomes	4	
	Excluded – not in English	0	
	Excluded – wrong study type (incl SR)	462	
TOTAL	**Included**		**3**

B8.1.2.4 Economic search

	Status	No. citations excluded	No. citations included
Identified via literature search			73
Identified manually			0
Duplicate citation		29	
TOTAL	**Included**		**44**
Title/abstract	Excluded – wrong intervention	37	
	Excluded – conference abstract	1	
TOTAL	**Included**		**6**
Full paper	Excluded – not an economic analysis	2	
	Excluded – wrong intervention	4	

[1] Via the reference lists of included SRs.
[2] Includes 56 studies assessing screening, treatment or prevention only.

TOTAL	**Included**		**0**

B8.1.3 Excluded studies list

B8.1.3.1 Systematic review search

Full text citation	Reason for exclusion
Tsai, A. C., J. A. Scott, K. J. Hung, J. Q. Zhu, L. T. Matthews, C. Psaros and M. Tomlinson (2013). "Reliability and validity of instruments for assessing perinatal depression in African settings: Systematic review and meta-analysis." PLoS ONE 8(12).	Wrong population

B8.1.3.2 Psychosocial assessment search

Full list of excluded studies can be provided on request.

B8.1.3.3 Anxiety screening search

Full list of excluded studies can be provided on request.

B8.1.3.4 Economic search

Full text citation	Reason for exclusion
Hewitt, C., S. Gilbody, S. Brealey, M. Paulden, S. Palmer, R. Mann, J. Green, J. Morrell, M. Barkham, K. Light and D. Richards (2009). "Methods to identify postnatal depression in primary care: an integrated evidence synthesis and value of information analysis." Health Technol Assess 13(36).	Wrong intervention
Hewitt, C. E. and S. M. Gilbody (2009). "Is it clinically and cost effective to screen for postnatal depression: a systematic review of controlled clinical trials and economic evidence." BJOG: An International Journal of Obstetrics & Gynaecology 116(8): 1019-1027.	Not an economic analysis
NICE (2015) National Collaborating Centre for Mental Health. Antenatal and Postnatal Mental Health: the NICE guideline on Clinical Management and Service Guidance. National Clinical Guideline Number 192: 1-922.	Wrong intervention
Paulden, M., S. Palmer, C. Hewitt and S. Gilbody (2009). "Screening for postnatal depression in primary care: cost effectiveness analysis." BMJ 339: b5203.	Wrong intervention
Petrou, S., C. J. Morrell and M. Knapp (2015). An overview of health economic aspects of perinatal depression. Milgrom, Jeannette [Ed]; Gemmill, Alan W [Ed] (2015) Identifying perinatal depression and anxiety: Evidence-based practice in screening, psychosocial assessment, and management (pp 228-239) xvii, 274 pp Wiley-Blackwell.	Not an economic analysis
Wilkinson, A., S. Anderson and S. B. Wheeler (2017). "Screening for and Treating Postpartum Depression and Psychosis: A Cost-Effectiveness Analysis." Maternal & Child Health Journal 21(4): 903-914.	Wrong intervention

B8.2 COMPARISON OF INDIVIDUAL STUDIES WITHIN RELEVANT SRS OF SCREENING

B8.2.1 Depression screening SRs

Six SRs that included an assessment of the technical performance of tools for depression screening were compared for included studies as shown in the table below.

Table B 8-1 Individual included studies in published SRs – depression screening

		NICE 2015	O'Connor 2016[3]	Kozinsky 2015	Thombs 2014	Myers 2013[3]	Mann 2011	Gibson 2009
Literature search date		*Apr 2014*	*Jan 2015*	*Dec 2013*	*Apr 2013*	*Jul 2012*	*Apr-May 2010?*	*Jul 2008*
Tools evaluated		*EPDS; PHQ; Whooley qns; K-10*	*EPDS; PHQ*	*EPDS*	*EPDS; GHQ-12*	*[ANRQ**], BDI, BDI-11, EPDS, HRSD-17 HRSD-21, Leverton Q, PDSS, PHQ-9, 2Q screen*	*Whooley qns*	*EPDS*
Alvarado	2015		✓					
Toreki	2014		✓					
Chen	2013		✓					
Stewart	2013			✓				
Thiagayson	2013	✓						
Toreki	2013	✓	✓	✓				
Ekeroma	2012	✓				✓		
Mann	2012	✓	✓			✓		
Sidebottom	2012	✓						
Tandon	2012	✓	✓					
Yawn	2012		✓					
Bergink	2011	✓		✓				
Fernandes	2011	✓						
Flynn	2011	✓						
Hamdan	2011					✓		
Ji	2011					✓		
Leung	2011		✓		✓			
Rubertsson	2011	✓		✓				
Tran	2011	✓						
Austin	2010					✓		
Chaudron	2010	✓				✓		
Chibanda	2010	✓						
Edmonson	2010					✓		
Glavin	2010		✓					
Lau	2010	✓						
Pereira	2010					✓		
Smith	2010	✓	✓					
Bunevicius	2009a	✓	✓	✓				
Bunevicius	2009b		✓					
Csatordi	2009					✓		
Gjerdingen	2009	✓	✓			✓	✓	
Leonardou	2009	✓						
Morrell	2009		✓					
Phillips	2009	✓						
Spies	2009	✓						
Wang	2009			✓				

[3] Studies relating to benefits of treatment in those who screen positive for depression are not included in this table.

		NICE 2015	O'Connor 2016[3]	Kozinsky 2015	Thombs 2014	Myers 2013[3]	Mann 2011	Gibson 2009
Bass	2008							
Clarke	2008	✓	✓			✓		
Matthey	2008	✓						
Aguilar-Navarro	2007							✓
Baggaley	2007	✓						
Gausia	2007a	✓						
Mazhari	2007	✓						
Navarro	2007					✓		
Ortega Orcos	2007							✓
Pitanupong	2007	✓						
Santos	2007	✓						
Su	2007			✓				
Adewuya	2006	✓		✓				
Alvarado-Esquivel	2006	✓						
Felice	2006	✓	✓	✓		✓		
Jardri	2006					✓		
Werrett	2006	✓						
Adewuya	2005	✓						
Adouard	2005		✓	✓				
Agoub	2005	✓						
Beck	2005					✓		✓
Kadir	2005	✓						
Milgrom	2005	✓						
Teng	2005	✓	✓					
Wickberg	2005		✓					
Aydin	2004	✓						
Ascaso Terren	2003	✓						
Berle	2003	✓						
Garcia-Esteve	2003	✓	✓					✓
Mahmud	2003	✓						
Uwakwe	2003	✓						
Fernandez-San Martin	2002							✓
MacArthur	2002		✓					
Martinez de la Iglesia	2002							✓
Regmi	2002	✓						
Robison	2002							✓
Vega-Dienstmaier	2002							✓
Wulsin	2002							✓
Aragones Benaiges	2001							✓
Beck	2001a	✓	✓					
Eberhard-Gran	2001	✓						
Lee	2001		✓					
Yoshida	2001	✓						
Leverton	2000	✓	✓					
Muzik	2000	✓						

		NICE 2015	O'Connor 2016[3]	Kozinsky 2015	Thombs 2014	Myers 2013[3]	Mann 2011	Gibson 2009
Yamashita	2000		✓					
Baca	1999							✓
Barnett	1999	✓						
Benvenuti	1999	✓	✓					
Guedeney	1998	✓	✓					
Lee	1998	✓						
Carpiniello	1997	✓	✓					
Ghubash	1997	✓						
Cox	1996		✓					
Wickberg	1996	✓						
Jadresic	1995	✓						
Boyce	1993	✓						
Ring	1991							✓
Murray	1990a	✓	✓	✓				
Harris	1989	✓	✓					
Cox	1987	✓						

B8.3 LISTS OF INCLUDED STUDIES FROM SUPPLEMENTARY SEARCHES

B8.3.1 Anxiety screening

Table B8-2 List of individual studies included for anxiety screening

Study ID	Tool(s)	Nature of evidence included in current reviews			
		Technical performance	Acceptability	Effectiveness	Implementability
Pre-specified tools					
Grigoriadis 2011	EPDS; EPDS-3A	✓			
Simpson 2014	EPDS; GAD-7	✓			
Tran 2011	EPDS; GHQ-12	✓			

During the review of search hits a paper by Somerville 2015 was identified that describes the use of the Perinatal Anxiety Screening Scale (PASS). This tool has been developed and is in use in Australia. It was excluded from our review of technical performance as it did not report sensitivity and specificity with reference to a standard, but it is discussed in the narrative review of implementation of anxiety screening.

B8.3.2 Psychosocial assessment

Table B8-3 List of individual studies included for psychosocial assessment

Study ID	Tool(s)	Nature of evidence included in current reviews			
		Technical performance	Acceptability	Effectiveness	Implementability
Pre-specified tools					
Carroll 2005	ALPHA	✓	✓	✓	
Matthey 2004	ARPA	✓		✓	✓
Austin 2013	ANRQ (+ EPDS)	✓	✓		✓
Reilly 2015	ANRQ (+ EPDS)	✓	✓	✓	
Bernazzani 2005	CAME	✓			
Howard 2007	CAN-M	✓			
Austin 2005	PRQ	✓	✓		✓
EPDS plus structured psychosocial assessment					
Kohlhoff 2016	EPDS + PSA			✓	✓
Matthey 2016	EPDS + PSA			✓	
Quispel 2012	EPDS + PSA				✓

Abbreviations: ALPHA, Antenatal Psychosocial Health Assessment; ANRQ, Antenatal Risk Questionnaire; ARPA, Antenatal Routine Psychosocial Assessment; CAME, Contextual Assessment of Maternity Experience; CAN-M, Camberwell Assessment of Need – Mothers; EPDS, Edinburgh Postnatal Depression Scale; PRQ, Pregnancy Risk Questionnaire; PSA, structured psychosocial assessment not with a named tool.

B8.4 DETAILS OF QUADAS-2 ASSESSMENTS

B8.4.1 Questions for quality assessment of diagnostic studies

Patient selection - Risk of Bias:

- Was a consecutive or random sample of patients enrolled? Yes/No
- Was a case-control design avoided? Yes/No
- Did the study avoid inappropriate exclusions? Yes/No
- Could the selection of patients have introduced bias? Risk judged to be Low, High, or Unclear

Patient selection – Applicability concerns:

- Is there concern that the included patients do not match the review question? Concern judged to be Low, High, or Unclear

Index test(s) – Risk of Bias:

- Were the index test results interpreted without knowledge of the results of the reference standard? Yes/No
- If a threshold was used, was it pre-selected? Yes/No
- Could the conduct or interpretation of the index test have introduced bias? Risk judged to be Low, High, or Unclear

Index test(s) – Applicability concerns:

- Is there concern that the index test, its conduct, or interpretation differ from the review question? Concern judged to be Low, High, or Unclear

Reference standard – Risk of Bias:

- Is the reference standard likely to correctly classify the target condition? Yes/No
- Were the reference standard results interpreted without knowledge of the results of the index test? Yes/No
- Could the reference standard, its conduct, or its interpretation have introduced bias? Risk judged to be Low, High, or Unclear

Reference standard – Applicability concerns:

- Is there concern that the target condition defined by the reference standard does not match the review question? Concern judged to be Low, High, or Unclear

Flow and Timing – Risk of Bias:

- Was there an appropriate interval between index test(s) and reference standard? Yes/No
- Did all patients receive the reference standard? Yes/No
- Did patients receive the same reference standard? Yes/No
- Were all patients included in the analysis? Yes/No
- Could the patient flow have introduced bias? Risk judged to be Low, High, or Unclear

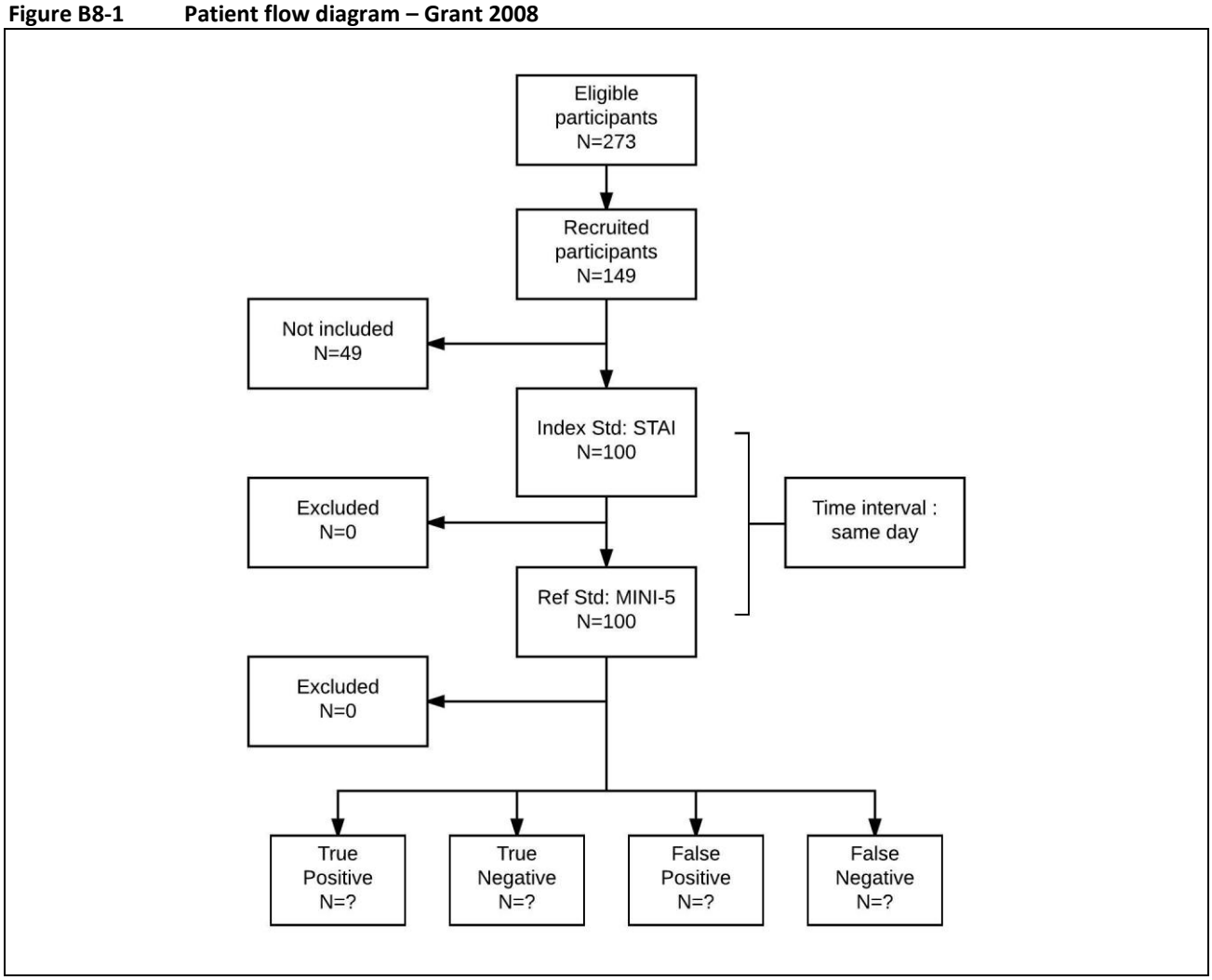
B8.4.2 Checklists for additional screening studies

Table B8-4 Key aspects of the review question

The review question:	
Patients (presentation, setting, intended use of index test, prior testing)	• Pregnant women or women in the first 12 month postpartum • Primary care setting: General Practice; non-psychiatric, out-patient antenatal or postnatal clinics; midwife clinics; parenting groups; mother-and-baby units • Test used as screening tool to identify individuals for further mental health assessment • No prior mental health testing
Index test(s)	• DASS-21; EPDS (full or 3A version); GAD-2; GHQ; HADS; HADS-A; K-10; STAI
Reference standard and target condition (definition of a 'case')	• Structured diagnostic interview such as the MINI-5, DSM IV • Anxiety: anxiety or anxiety disorder (generalized or specific e.g. social phobia, panic disorder, PTSD)

Abbreviations: DASS21, Depression Anxiety Stress Scales; DSM, Diagnostic and Statistical Manual; EPDS, Edinburgh Postnatal Depression Scale; GAD-2, Generalised Anxiety Disorder scale– 2-item scale; GHQ, General Health Questionnaire; HADS, Hospital Anxiety and Depression Scale; HADS-A, Hospital Anxiety and Depression Scale for anxiety; K-10, Kessler 10 item questionnaire; PTSD, post-traumatic stress disorder; STAI, State-Trait Anxiety Inventory; MINI, Mini-international neuropsychiatric interview.

Figure B8-1 Patient flow diagram – Grant 2008



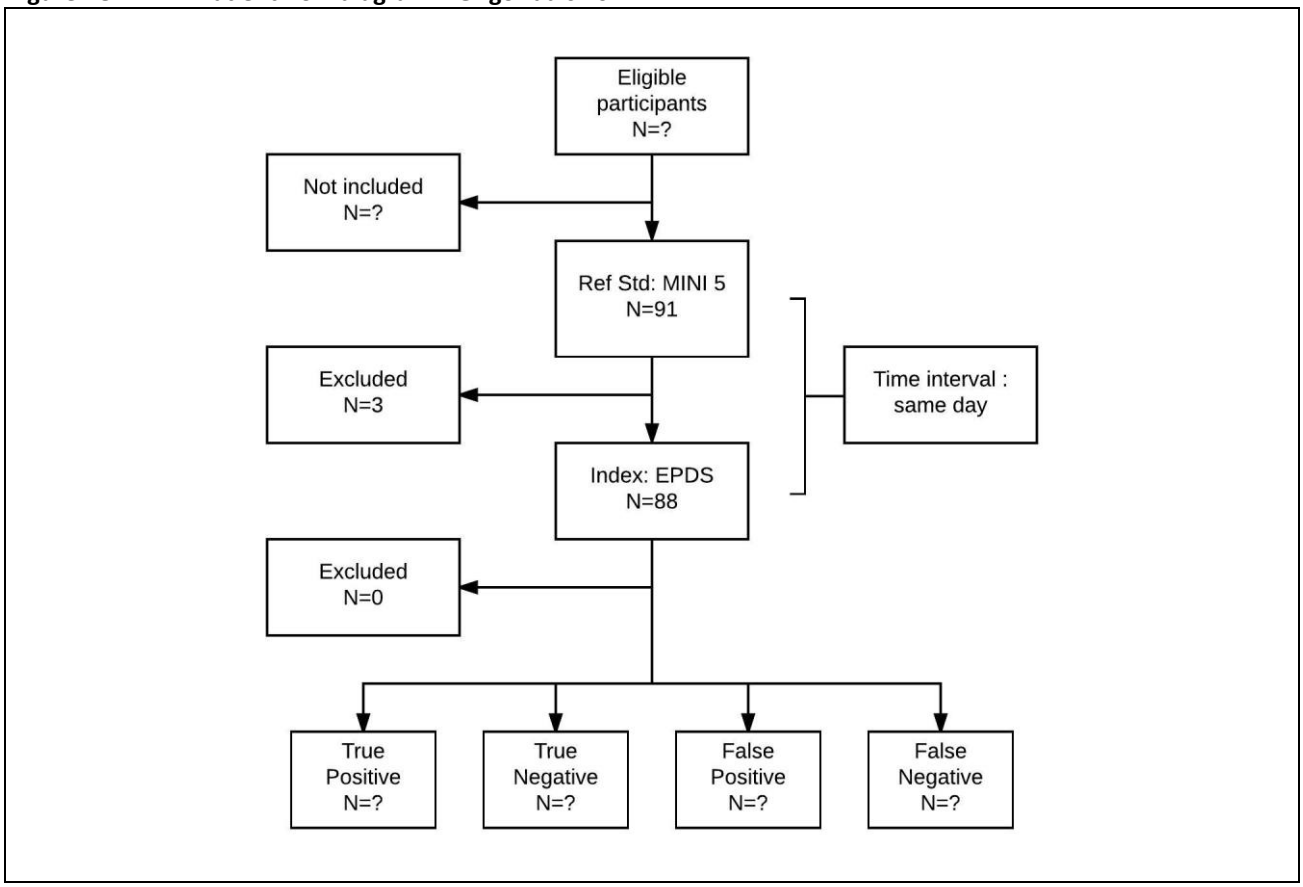
Abbreviations: MINI, Mini-international neuropsychiatric interview; STAI, State-Trait Anxiety Inventory; Std, standard.

Table B8-5 Risk of bias and applicability judgments – Grant 2008

Risk-of-bias assessment, Grant 2008	
Domain 1: Patient selection	
Risk of Bias	
Description of patient selection methods:	All pregnant women presenting to a single Australian obstetric clinic were invited to participate.
Was a consecutive or random sample of patients enrolled?	Consecutive
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Risk: Low
Concerns regarding applicability	
Description of included patients:	Women who were pregnant and booking in for their first antenatal appointment (average of 15 weeks' gestation, range 7-32 weeks)
Is there concern that the included patients do not match the review question?	Concern: Low
Domain 2: Index test	
Risk of Bias	
Description of the index test and how it was conducted and interpreted	State Trait Anxiety Scale was self-completed by participants and interpreted according to published methods..
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	Yes
Could the conduct or interpretation of the index test have introduced bias?	Risk: Low
Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differ from the review question?	Concern: Low

Risk-of-bias assessment, Grant 2008	
Domain 3: Reference Standard	
Risk of bias	
Description of the reference standard and how it was conducted and interpreted	Mini-Plus International Neuropsychiatric Interview version 5.0 was administered but no details on who administered it or how it was interpreted.
Is the reference standard likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index test	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Risk: Low
Concerns regarding applicability	
Is there concern that the target condition as defined by the refence standard does not match the review question?	Concern: Low
Domain 4: Flow and Timing	
Risk of bias	
Description of patients who did not receive the index text(s) and/or reference standard or who were excluded from the 2x2 table	All participants received index test and reference standard but 2x2 tables not reported.
Description of the time interval and any interventions between the index test(s) and reference standard.	Index test and reference standard conducted on the same day.
Was there an appropriate interval between index test(s) and reference standard?	Yes
Did all patients receive the reference standard?	Yes
Did patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	Risk: Low

Figure B8-2 Patient flow diagram – Grigoriadis 2011



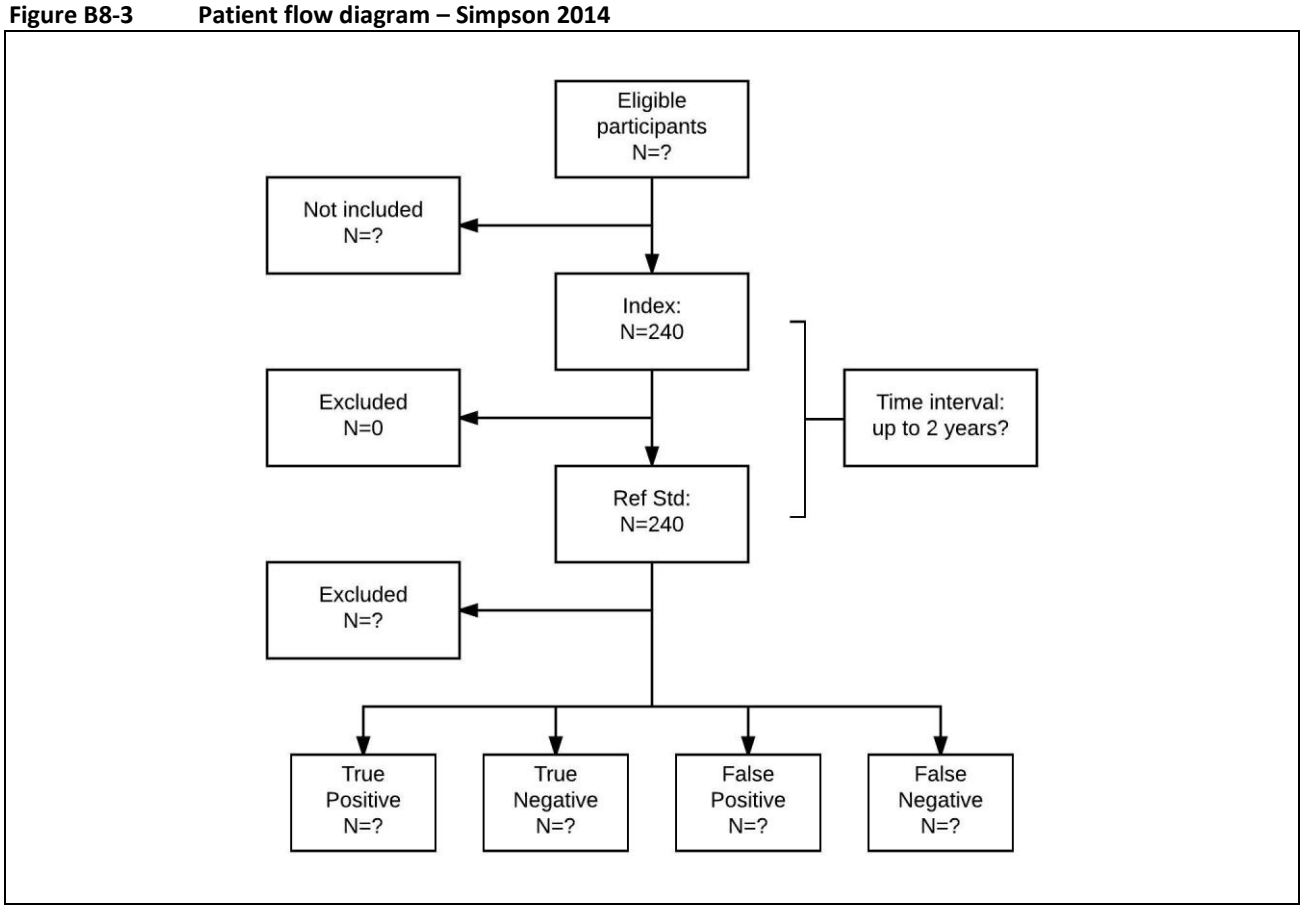
Abbreviations: EPDS, Edinburgh Postnatal Depression Scale; MINI, Mini-international neuropsychiatric interview; Std, standard.

Table B8-6 Risk of bias and applicability judgments – Grigoriadis 2011

Risk-of-bias assessment, Grigoriadis 2011	
Domain 1: Patient selection	
Risk-of-bias assessment	
Description of patient selection methods:	All women presenting to a single Canadian perinatal mental health clinic (except those with a diagnosis of substance use or psychotic disorder) were eligible and were invited to participate.
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Risk: Unclear
Concerns regarding applicability	
Description of included patients:	Women who were pregnant (any trimester) or <12 months postpartum and accessing psychiatric services.
Is there concern that the included patients do not match the review question?	Concern: High
Domain 2: Index test	
Risk of Bias	
Description of the index test and how it was conducted and interpreted	The full version of the EPDS was self-completed by participants; standard scoring was used.
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	Yes – a number of pre-specified thresholds were examined
Could the conduct or interpretation of the index test have introduced bias?	Risk: Unclear
Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differ from the review question?	Concern: Low
Domain 3: Reference Standard	
Risk of bias	
Description of the reference standard and how it was conducted and interpreted	The Mini International Neuropsychiatric Interview version 5 (MINI-5) was conducted by the same trained research assistant; rater reliability was confirmed prior to study commencement. Case defined as: generalized anxiety disorder
Is the reference standard likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index test	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Risk: Unclear
Concerns regarding applicability	
Is there concern that the target condition as defined by the refence standard does not match the review question?	Concern: Low
Domain 4: Flow and Timing	
Risk of bias	
Description of patients who did not receive the index text(s) and/or reference standard or who were excluded from the 2x2 table	91 subjects received the reference standard and 88 received the index test; no information is provided on the 3 subjects who did not complete the EPDS; a 2x2 table is not provided by the authors.
Description of the time interval and any interventions between the index test(s) and reference standard.	The index test and reference standard were conducted on the same day.
Was there an appropriate interval between index test(s) and reference standard?	Yes
Did all patients receive the reference standard?	Yes
Did patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Unclear
Could the patient flow have introduced bias?	Risk: Unclear

Abbreviations: EPDS, Edinburgh Postnatal Depression Scale; MINI, Mini-international neuropsychiatric interview.

Figure B8-3 Patient flow diagram – Simpson 2014



Abbreviations: Std, standard.

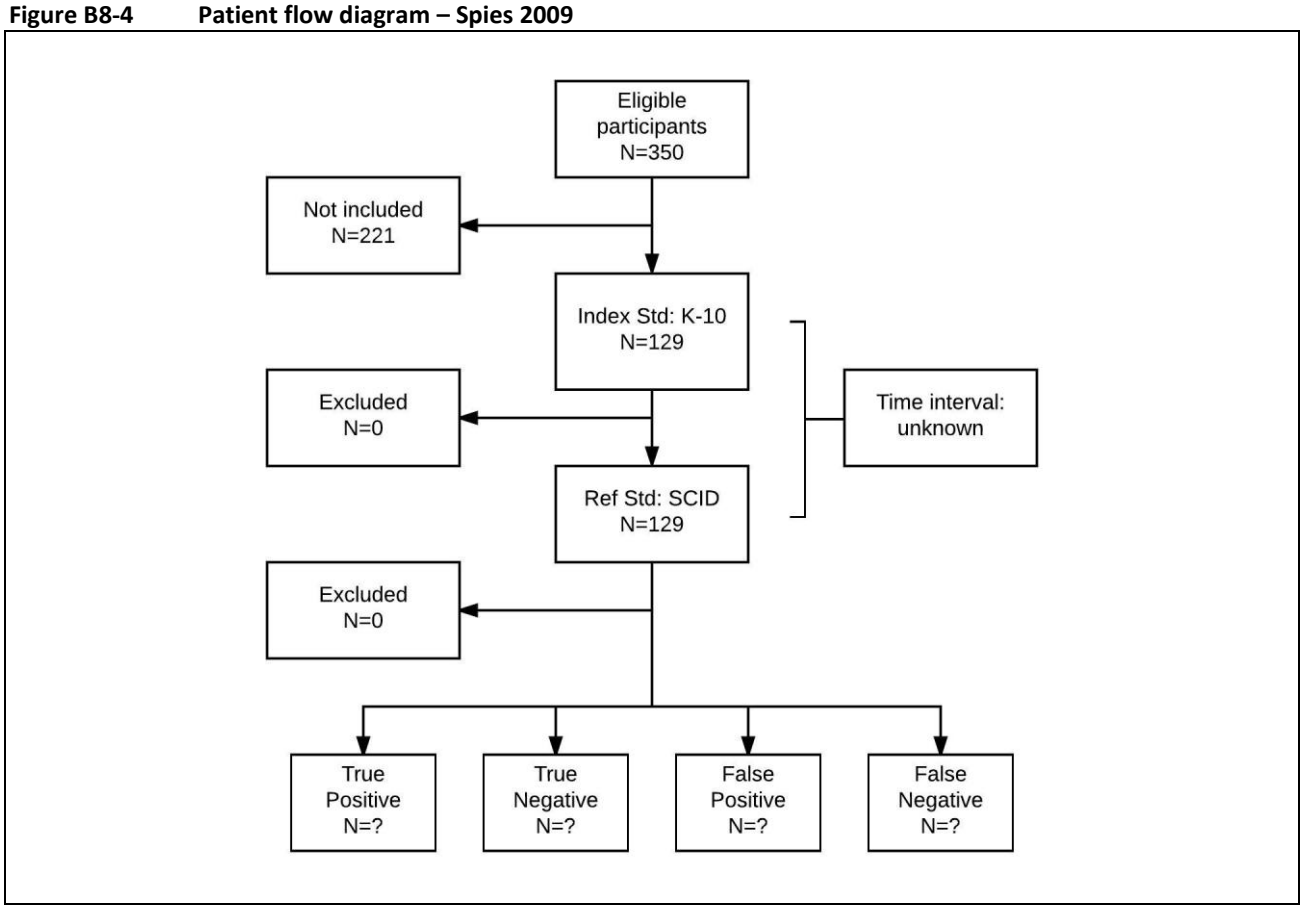
aTable B8-7 Risk of bias and applicability judgments – Simpson 2014

Risk-of-bias assessment, Simpson 2014	
Domain 1: Patient selection	
Risk of Bias	
Description of patient selection methods:	Individuals were selected from referrals to one health centre.
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Risk: High
Concerns regarding applicability	
Description of included patients:	Individuals were pregnant (any trimester) or postpartum (time since birth not specified) women who had been referred for psychiatric consultation.
Is there concern that the included patients do not match the review question?	Concern: High
Domain 2a: Index test – GAD-7	
Risk of Bias	
Description of the index test and how it was conducted and interpreted	Was administered on 'initial assessment day' but not stated when this was, or who administered the test; appears to have used the original methods of Spitzer 2006 to score the results, but not stated explicitly.
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	Yes
Could the conduct or interpretation of the index test have introduced bias?	Risk: High
Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differ from the review question?	Concern: Unclear
Domain 2b: Index test – EPDS (full and 3A)	
Risk of Bias	
Description of the index test and how it was conducted and interpreted	Was administered on 'initial assessment day' but not stated when this was, or who administered the test; appears to have used standard methods to score each version of the EPDS, but not stated explicitly
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear

Risk-of-bias assessment, Simpson 2014	
If a threshold was used, was it pre-specified?	Yes
Could the conduct or interpretation of the index test have introduced bias?	Risk: High
Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differ from the review question?	Concern: Unclear
Domain 3: Reference Standard	
Risk of bias	
Description of the reference standard and how it was conducted and interpreted	'Clinical diagnosis by a psychiatrist' but not otherwise defined except that it was *not* a structured interview; appears that clinical diagnosis was extracted retrospectively via chart review.
Is the reference standard likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index test	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Risk: High
Concerns regarding applicability	
Is there concern that the target condition as defined by the refence standard does not match the review question?	Concern: Low
Domain 4: Flow and Timing	
Risk of bias	
Description of patients who did not receive the index text(s) and/or reference standard or who were excluded from the 2x2 table	Stated that all patients received both index tests and the reference standard, but no 2x2 table was presented
Description of the time interval and any interventions between the index test(s) and reference standard.	Not clear if the reference standard was conducted before or after the index texts; also not clear what the time interval was between the index tests and reference standard and could potentially be up to 2 years
Was there an appropriate interval between index test(s) and reference standard?	No
Did all patients receive the reference standard?	Yes – but not verifiable
Did patients receive the same reference standard?	Unclear – no information provided on inter-rater reliability of diagnosis amongst psychiatrists at the centre
Were all patients included in the analysis?	Yes - but not verifiable
Could the patient flow have introduced bias?	Risk: High

Abbreviations: EPDS, Edinburgh Postnatal Depression Scale.

Figure B8-4 Patient flow diagram – Spies 2009



Abbreviations: SCID, Structural Clinical Interview for DSM Disorders; Std, standard.

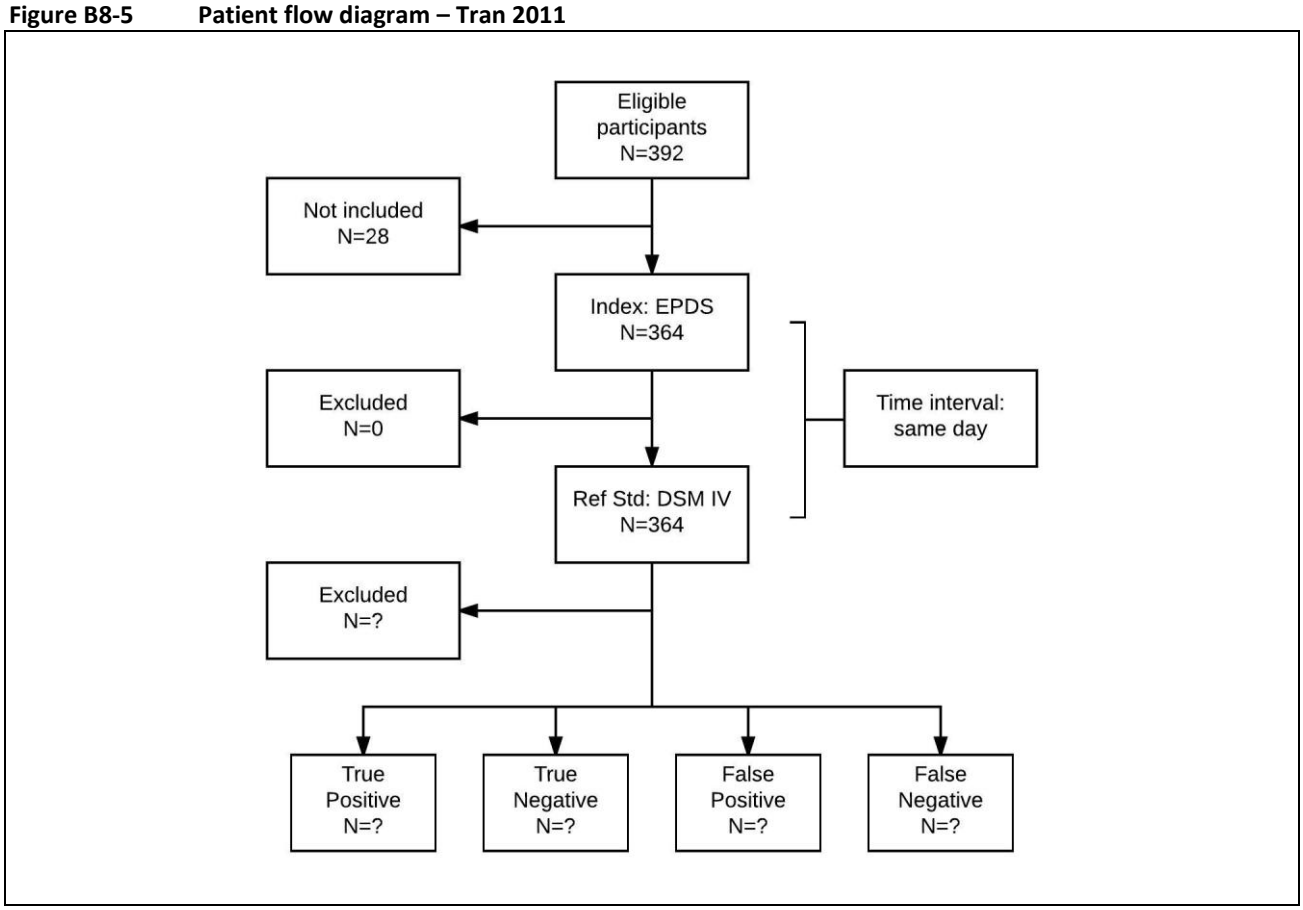
Table B8-8 Risk of bias and applicability judgments – Spies 2009

Risk-of-bias assessment, Spies 2009	
Domain 1: Patient selection	
Risk of Bias	
Description of patient selection methods:	All pregnant women (>18 years of age) with gestation of <20 weeks were invited to participate, representing subset of larger prospective study.
Was a consecutive or random sample of patients enrolled?	Consecutive
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Risk: Low
Concerns regarding applicability	
Description of included patients:	Pregnant women with a mean age of 25 years presenting for their first antenatal visit in South Africa.
Is there concern that the included patients do not match the review question?	Concern: Low
Domain 2: Index test	
Risk of Bias	
Description of the index test and how it was conducted and interpreted	The Kessler-10 was translated into Afrikaans and self-completed by participants with reading assistance from the researchers; findings interpreted according to published methods.
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	No threshold was pre-specified, but a range of values were examined.
Could the conduct or interpretation of the index test have introduced bias?	Risk: Unclear
Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differ from the review question?	Concern: Unclear
Domain 3: Reference Standard	
Risk of bias	
Description of the reference standard and how it was conducted and interpreted	The Structured Clinical Interview for DSM (SCID) was administered by the same researcher for all participants and interpreted according to published methods.
Is the reference standard likely to correctly classify the target condition?	Yes

Risk-of-bias assessment, Spies 2009	
Were the reference standard results interpreted without knowledge of the results of the index test	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Risk: Low
Concerns regarding applicability	
Is there concern that the target condition as defined by the refence standard does not match the review question?	Concern: Low
Domain 4: Flow and Timing	
Risk of bias	
Description of patients who did not receive the index text(s) and/or reference standard or who were excluded from the 2x2 table	Stated that all patients received both index tests and the reference standard, but no 2x2 table was presented.
Description of the time interval and any interventions between the index test(s) and reference standard.	Not reported
Was there an appropriate interval between index test(s) and reference standard?	Unknown
Did all patients receive the reference standard?	Yes
Did patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	Risk: High

Abbreviations: DSM, Diagnostic and Statistical Manual; SCID, Structural Clinical Interview for DSM Disorders.

Figure B8-5 Patient flow diagram – Tran 2011



Abbreviations: DSM, Diagnostic and Statistical Manual; EPDS, Edinburgh Postnatal Depression Scale; Std, standard.

Table B8-9 Risk of bias and applicability judgments – Tran 2011

Risk-of-bias assessment	**Tran 2011**
Domain 1: Patient selection	
Risk of Bias	
Description of patient selection methods:	Health centres were randomly selected across North Vietnam, and all eligible women at each centre were invited to participate in the study.
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear

Risk-of-bias assessment	**Tran 2011**
Could the selection of patients have introduced bias?	Risk: Unclear
Concerns regarding applicability	
Description of included patients:	Women who were at least 28 weeks' gestation or were 4-6 weeks postpartum and attending health centres as part of their standard perinatal care.
Is there concern that the included patients do not match the review question?	Concern: Low
Domain 2a: Index test - EPDS	
Risk of Bias	
Description of the index test and how it was conducted and interpreted	The full version of the EPDS was administered via individual interviews, conducted by the researchers; results were scored using the original methods of Cox 1987.
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes – a number of pre-specified thresholds were examined.
Could the conduct or interpretation of the index test have introduced bias?	Risk: Low
Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differ from the review question?	Concern: Low
Domain 2b: Index test – GHQ-12	
Risk of Bias	
Description of the index test and how it was conducted and interpreted	The 12-item version of the GHQ was administered via individual interviews, conducted by the researchers; results were scored using the methods of Goldberg 1988.
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes – a number of pre-specified thresholds were examined.
Could the conduct or interpretation of the index test have introduced bias?	Risk: Low
Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differ from the review question?	Concern: Low
Domain 3: Reference Standard	
Risk of bias	
Description of the reference standard and how it was conducted and interpreted	Individual structured clinical interviews for DSM IV Axis 1 Diagnoses (for depression, generalized anxiety, and panic disorder) were conducted and interpreted by a psychiatrist. Case defined as: depression or generalized anxiety or panic disorder.
Is the reference standard likely to correctly classify the target condition?	Yes – but is a composite outcome, not anxiety alone.
Were the reference standard results interpreted without knowledge of the results of the index test	Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	Risk: Low
Concerns regarding applicability	
Is there concern that the target condition as defined by the refence standard does not match the review question?	Concern: Low
Domain 4: Flow and Timing	
Risk of bias	
Description of patients who did not receive the index text(s) and/or reference standard or who were excluded from the 2x2 table	All participants received both index tests and the reference standard; no 2x2 table is provided and it is not known if all findings were included in the analyses
Description of the time interval and any interventions between the index test(s) and reference standard.	Both of the index tests and the reference standard were conducted on the same day
Was there an appropriate interval between index test(s) and reference standard?	Yes
Did all patients receive the reference standard?	Yes
Did patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Unclear
Could the patient flow have introduced bias?	Risk: Unclear

Abbreviations: DSM, Diagnostic and Statistical Manual; EPDS, Edinburgh Postnatal Depression Scale; GHQ, General Health Questionnaire.