Mental health care in the perinatal period: Australian clinical practice guideline

2023 Update

Technical Report Part B:

Psychosocial assessment and screening for depression or anxiety in the perinatal period

Prepared by Kate Sawkins, Sue Campbell, Emma Donoghue, Allison Wyndham, Samara Lewis, Adrian Fann & Agnes Wilson

ISBN: 978-0-6485095-4-7

clinical evidence for the real world

13 February 2023

CONTENTS

TABLES

Abbreviations

AD	Anxiety and related disorders
ALPHA	Antenatal Psychosocial Health Assessment
ANRQ	Antenatal Risk Questionnaire
ANRQ-R	Antenatal Risk Questionnaire - Revised
ANRQ-2A	2 'anxiety' items from the Antenatal Risk Questionnaire
ARPA	Antenatal Routine Psychosocial Assessment
AUC	area under the curve
AUROC	area under the receiver-operating characteristics curve
CALD	culturally and linguistically diverse
CAME	Contextual Assessment of Maternity Experience
CAN-M	Camberwell Assessment of Need – Mothers
CBR	Consensus-based recommendation
CBT	cognitive behavioural therapy
CI	confidence interval
CIDI	Composite International Diagnostic Interview
CIS	Clinical Interview Schedule
CIS-R	Clinical Interview Schedule - Revised
COPE	Centre of Perinatal Excellence
DASS-21	Depression Anxiety Stress Scales
DIGS	Diagnostic Interview for Genetic Studies
DSM	Diagnostic and Statistical Manual of Mental Disorders (DSM-IV, 4th edition; DSM-V, 5th edition)
EBR	Evidence-based recommendation
EPDS	Edinburgh Postnatal Depression Scale
EPDA-3A	Edinburgh Postnatal Depression Scale – Anxiety subscale
EWG	Expert Working Group
GAD	Generalised Anxiety Disorder
GAD-2	Generalized Anxiety Disorder 2-item scale
GAD-7	Generalized Anxiety Disorder 7-item scale
GHQ	General Health Questionnaire
GRADE	Grading of Recommendations, Assessment, Development and Evaluation
HADS	Hospital Anxiety and Depression Scale
HADS-A	Hospital Anxiety and Depression Scale – Anxiety subscale
HAM-D	Hamilton Depression Rating Scale
HCP	health care professional
HTA	health technology assessment
ICD	International Statistical Classification of Diseases and Related Health Problems (ICD-10, 10th revision)
KMMS	Kimberley Mum's Mood Scale
K10	Kessler Psychological Distress Scale (10 item)
LGBTQI+	lesbian, gay, bisexual, transgender, queer/questioning, intersex
LMIC	low middle income countries
LR	likelihood ratio
LR-	negative likelihood ratio

LR+	positive likelihood ratio
MA	meta-analysis
MDD	major depressive disorder
MDE	major depressive episode
MINI	Mini-International Neuropsychiatric Interview
NA	not applicable
NICE	National Institute of Health and Care Excellence
NPV	negative predictive value
NR	not reported
OCD	obsessive compulsive disorder
OR	odds ratio
PHQ	Patient Health Questionnaire
PHQ-2	first 2 items of the PHQ-9
PHQ-9	Patient Health Questionnaire-9
PICO	population-intervention-comparator-outcome
PND	postnatal depression
PNMH	perinatal mental health
PNRQ	Postnatal Risk Questionnaire
PNRQ-R	Postnatal Risk Questionnaire - Revised
PP	Practice point
PPD	postpartum depression
PPV	positive predictive value
PRQ	Pregnancy Risk Questionnaire
PTSD	post-traumatic stress disorder
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
RCT	randomised controlled trial
ROC	receiver operating characteristics
SAGE-SR	Series of Assessments for Guiding Evaluation – Self-Report
SCID	Structured Clinical Interview for DSM Disorders
SIGH-ADS	Structured Interview Guide for the Hamilton Depression Rating Scale with Atypical Depression Supplement
SR	systematic review
SSRI	selective serotonin reuptake inhibitor
STAI	State-Trait Anxiety Inventory
UK	United Kingdom
UMIC	upper middle income countries
US	United States

B1. Introduction

In October 2017, the Centre of Perinatal Excellence (COPE) published a national clinical practice guideline on *Effective Mental Health Care in the Perinatal Period* (hereafter referred to as the **2017 Australian Guideline**).

The aim of this Evidence Review Update is to assess the body of evidence – including the 'new' evidence – relating to:

- psychosocial assessment of birthing parents at risk of mental health problems in the perinatal period
- screening birthing parents for depression and anxiety in the perinatal period.

Consistent with the 2017 Australian Guideline, 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 quantitative evidence (e.g. screening test performance appraised using the Quality Assessment of Diagnostic Accuracy Studies [QUADAS]-2 tool) and narrative reviews of the acceptability, effectiveness and implementation issues associated with perinatal mental health assessment (psychosocial assessment as well as depression/anxiety screening). The overall summary of findings also considers non-technical characteristics of the tools, such as time to administer and complexity of scoring.

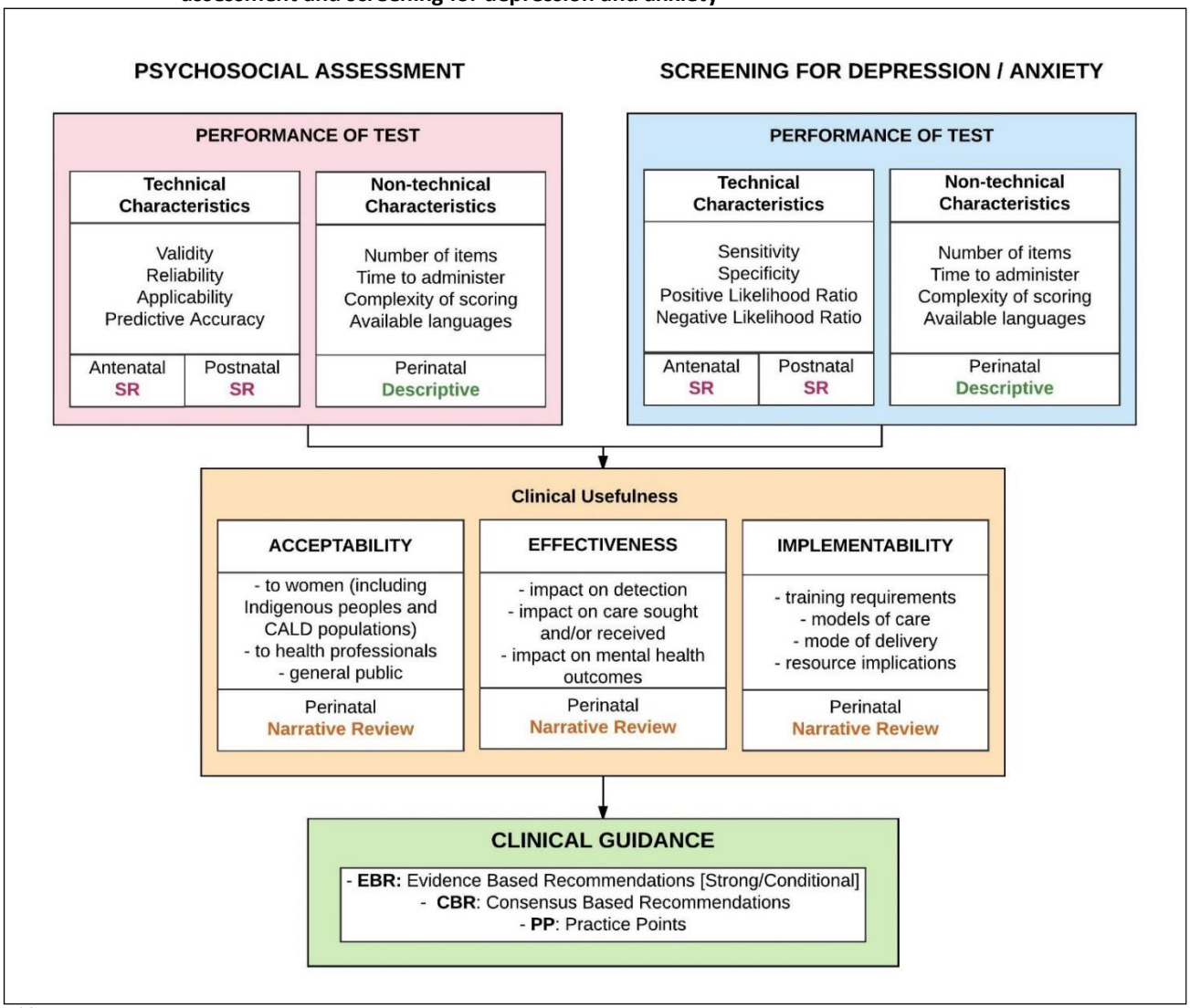
Given the sensitive and personal nature of the questions asked during perinatal mental health assessment, acceptability to women and non-mental health professionals (such as midwives, child and family health nurses, general practitioners and obstetricians) is of paramount importance. Particular attention has been given to evidence of acceptability to Aboriginal and Torres Strait Islander women, and to women from culturally and linguistically diverse backgrounds in Australia.

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 general approach to assessment of the evidence for psychosocial assessment and mental health screening in the perinatal period follows the Evidence to Decision framework developed for the 2017 Australian Guideline (see Figure 1). The sections within this Technical Report follow the format of this Evidence to Decision framework.

This Technical Report includes an overview of the methods used to identify and appraise the evidence, and the key findings (presented as Summary of Findings Tables, where appropriate). Details of the literature search strategies, included/excluded studies, characteristics of included studies, and risk of bias assessments are included in the Appendices.

Figure 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. Methodology

B2.1 Clinical questions

The Research Protocol for this update of the evidence review outlined three main research questions relating to the most appropriate methods for psychosocial assessment and screening of birthing parents in the perinatal period. These questions were addressed via systematic review.

The research questions are similar to those investigated for the 2017 Australian Guideline.

Table 1 Research questions relating to psychosocial assessment and mental health screening

Research questions and sub-questions
Q1. What are the most appropriate methods for psychosocial assessment of the birthing parent at risk of mental health problems in the perinatal period?
Q1a. What is the performance (defined as reliability, validity, and accuracy) of validated multidimensional tools for perinatal psychosocial assessment? *[addressed via SR – quantitative analysis]*
Q1b. What are the non-technical characteristics (defined as number of items, time to administer, perinatal/postnatal timing, complexity of scoring, training requirements, and available languages) of validated multidimensional tools for perinatal psychosocial assessment? *[addressed via descriptive review]*
Q1c. What is the acceptability to the birthing parent, health professionals, and the general public of validated multidimensional tools for perinatal psychosocial assessment? *[addressed via narrative review]*
Q1d. 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]*
Q1e. What are the implications (for resourcing, workforce, and models of care) of implementing perinatal psychosocial assessment (via different modes of delivery) with a validated multidimensional tool? *[addressed via narrative review]*
Q2. What are the most appropriate methods for screening the birthing parent for depression in the perinatal period?
Q2a. 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 SR – quantitative analysis]*
Q2b. What are the non-technical characteristics (defined as number of items, time to administer, perinatal/postnatal timing, complexity of scoring, training requirements, and available languages) of validated tools for perinatal depression screening? *[addressed via descriptive review]*
Q2c. What is the acceptability to the birthing parent, health professionals, and the general public of screening for perinatal depression? *[addressed via narrative review]*
Q2d. 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]*
Q2e. What are the implications (for resourcing, workforce, and models of care) of implementing perinatal depression screening (via different modes of delivery) with a validated tool? *[addressed via narrative review]*
Q3. What are the most appropriate methods for screening the birthing parent for anxiety in the perinatal period?
Q3a. 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 SR – quantitative analysis]*
Q3b. What are the non-technical characteristics (defined as number of items, time to administer, perinatal/postnatal timing, complexity of scoring, training requirements, and available languages) of validated tools for perinatal anxiety screening? *[addressed via descriptive review]*
Q3c.What is the acceptability to the birthing parent, health professionals, and the general public of screening for perinatal anxiety? *[addressed via narrative review]*
Q3d.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]*
Q3e. What are the implications (for resourcing, workforce, and models of care) of implementing perinatal anxiety screening (via different modes of delivery) with a validated tool? *[addressed via narrative review]*

B2.2 Criteria for determining study eligibility

B2.2.1 Study eligibility criteria for psychosocial assessment

For psychosocial assessment, the focus of the Evidence Review Update is the same as that in the 2017 Australian Guideline; that is, validated tools developed to identify a range of factors in a birthing parent's current situation or past that might place them 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 is the identification of one or more factors known to influence perinatal mental health. Instruments that are designed to identify current mental health problems are not included within our definition of psychosocial assessment but are instead included under mental health screening (see B2.2.2 and B2.2.3).

For this Evidence Review Update, 'new' evidence for psychosocial assessment was included if it met the PICO criteria in Table 2 and was published after the literature search date for the 2017 Australian Guideline (June 2016). The EWG added three tools that were not included in the PICO for the 2017 Guideline: Antenatal Risk Questionnaire-Revised (ANRQ-R), Postnatal Risk Questionnaire-Revised (PNRQ-R) and the Kimberly Mum's Mood Scale (KMMS).

Table 2 Detailed PICO criteria for Q1: Psychosocial assessment

Question 1	**What are the most appropriate methods for psychosocial assessment of birthing parents at risk of mental health problems in the perinatal period?**
Population	Pregnant or postnatal women (birthing parent) *Subgroups of interest:* • Aboriginal and Torres Strait Islander pregnant or postnatal women • Refugee and asylum seeker pregnant or postnatal women • Pregnant or postnatal women from migrant or CALD background • LGBTQI+ birthing parents and non-birthing partners with or without a previous history of abuse
Intervention	• Validated psychosocial assessment tools to identify people at risk of mental health problems in the perinatal period ○ Limited to tools investigated in the 2017 Australian Guideline (ALPHA, ANRQ, ARPA, CAME[a], CAN-M[b], PNRQ, PRQ) and the revised versions of the ANRQ and PNRQ (ANRQ-R and PNRQ-R), and the KMMS
Comparator	• Subsequent manifestation of mental health issues or any standard clinical/diagnostic interview as a reference standard
Outcomes	**Tool performance** *Critical outcomes* • Validity • Reliability • Predictive accuracy (OR odds of identifying a factor of concern) *Important outcomes* • Sensitivity • Specificity **Clinical usefulness** *Critical outcomes* • Acceptability to pregnant or postnatal women, to healthcare providers, to the general public

Abbreviations: ALPHA, Antenatal Psychosocial Health Assessment; ANRQ, Antenatal Risk Questionnaire; ANRQ-R, Antenatal Risk Questionnaire – Revised; ARPA, Antenatal Routine Psychosocial Assessment; CALD, culturally and linguistically diverse; CAME, Contextual Assessment of Maternity Experience; CAN-M, Camberwell Assessment of Need-Mothers; KMMS, Kimberly Mum's Mood Scale; LGBTQI+, lesbian, gay, bisexual, transgender, queer/questioning, intersex; PNRQ, Postnatal Risk Questionnaire; PNRQ-R, Postnatal Risk Questionnaire – Revised; PRQ, Pregnancy Risk Questionnaire.

a 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. Women with a history of MDD and women living in poverty comprise a subset of the target population.

b The **CAN-M** has been designed for use in pregnant women and mothers with current severe mental illness who are already receiving mental health care, which is very different to the target population for the current Guideline (women under routine antenatal care with unknown past or current mental health status).

B2.2.2 Study eligibility criteria for depression screening

For depression screening, the focus of the Evidence Review Update is the same as that in the 2017 Australian Guideline, which is on validated tools that have been developed or assessed in antenatal and/or postnatal women to examine current mental health. It should be noted that some tools used to screen for depression are also used to screen for anxiety. Consequently, as for the 2017 Australian Guideline, there will likely be some overlap in the evidence included for depression screening and for anxiety screening.

The primary goal of a tool for screening for perinatal depression in the Australian setting 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.

For this Evidence Review Update, 'new' evidence for depression screening was included if it met the PICO criteria in Table 3 and was published after the literature search date (April 2014) for the 2015 National Institute for Health and Care Excellence (NICE) guideline on *Antenatal and postnatal mental health*, which was the basis for the depression screening evidence in the 2017 Australian Guideline. The EWG added one additional tool that was not included in the PICO for the 2017 Australian Guideline: Hospital Anxiety and Depression Scale (HADS).

Studies of technical performance (including studies comparing the performance of two or more validated tools) were only eligible if a standardised diagnostic interview was used as the reference standard.

Table 3 Detailed PICO criteria for Q2: Screening for depression

Question 2	**What are the most appropriate methods for screening the birthing parent for depression in the perinatal period?**
Population	Pregnant or postnatal women (birthing parent) *Subgroups of interest:* • Aboriginal and Torres Strait Islander pregnant or postnatal women • Refugee and asylum seeker pregnant or postnatal women • Pregnant or postnatal women from migrant or CALD background • LGBTQI+ birthing parents and non-birthing partners with or without a previous history of abuse
Intervention	• Validated screening tools to identify people with depression in the perinatal period o Limited to tools investigated in the Australian Guideline (EPDS, PHQ [PHQ-2 or PHQ-9], K10, Whooley questions) and the HADS
Comparator	• Any type of standardised diagnostic interview, defined as a structured interview (such as the SCID, CIDI or MINI) delivered by trained staff, or an ICD mental health diagnosis by a psychiatrist or clinical psychologist • A different screening tool (from the list above)
Outcomes	**Tool performance** *Critical outcomes* • Positive Likelihood Ratio (LR+) • Negative Likelihood Ratio (LR-) • AUROC *Important outcomes* • Sensitivity • Specificity • Youden's index **Clinical usefulness** *Critical outcomes* • Acceptability to women, to healthcare providers, to the general public • Mental health outcomes *Important outcomes* • Impact on help-seeking behaviour (services sought or utilised) • Impact of detection (e.g., referral rates if screen positive)

Abbreviations: AUROC, area under the receiver-operating characteristics curve; CIDI, Composite International Diagnostic Interview; DASS-21, Depression Anxiety Stress Scales; DSM, Diagnostic and Statistical Manual of Mental Disorders; EPDS, Edinburgh Postnatal Depression Scale; HADS, Hospital Anxiety and Depression Scale; ICD, International Statistical Classification of Diseases and Related Health Problems; K10, Kessler Psychological Distress Scale (10 item); MINI, Mini-International Neuropsychiatric Interview; PHQ-2, first 2 items of the PHQ-9; PHQ-9, Patient Health Questionnaire-9; SCID, Structured Clinical Interview for DSM Disorders.

The EWG agreed that although discrimination properties of screening instruments are important, the most useful test performance measures are those that predict the probability of the condition in an individual. 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), because LR+ and LR- are independent of prevalence, whereas PPV and NPV are not. The EWG recognise the importance of sensitivity and specificity as test measures, and how comparing these measures at different cut-off thresholds is important for clinical interpretation of results. Area under the curve (AUC) of the receiver-operating characteristics (ROC) curve is a valuable global measure of test performance.

B2.2.3 Study eligibility criteria for anxiety screening

The focus in this Evidence Review Update is the same as that in the 2017 Australian Guideline, which is on validated tools for screening for anxiety that have been developed or assessed in birthing parents.

'New' evidence was included if it met the PICO criteria in Table 4 and was published after the literature search date for the 2017 Australian Guideline (June 2016). The EWG added one tool that was not included in the PICO for the 2017 Australian Guideline, namely the two anxiety items in the ANRQ (ANRQ-2A).

Studies of technical performance (including studies comparing the performance of two or more validated tools) were only eligible if a standardised diagnostic interview was used as the reference standard.

Table 4 Detailed PICO criteria for Q3: Screening for anxiety

Question 3	**What are the most appropriate methods for screening the birthing parent for anxiety in the perinatal period?**
Population	Pregnant or postnatal women (birthing parent) *Subgroups of interest:* • Aboriginal and Torres Strait Islander pregnant or postnatal women • Refugee and asylum seeker pregnant or postnatal women • Pregnant or postnatal women from migrant or CALD background • LGBTQI+ birthing parents and non-birthing partners with or without a previous history of abuse
Intervention	• Validated screening tools to identify people with anxiety in the perinatal period o Limited to tools investigated in the 2017 Australian Guideline (EPDS, DASS-21, GAD-2/GAD-7, GHQ, HADS, HADS-A, K10, STAI) or the ANRQ-2A
Comparator	• Any type of standardised diagnostic interview, defined as a structured interview (such as the SCID, CIDI or MINI) delivered by trained staff, or an ICD mental health diagnosis by a psychiatrist or clinical psychologist • A different screening tool (from the list above)
Outcomes	**Tool performance** *Critical outcomes* • Positive Likelihood Ratio (LR+) • Negative Likelihood Ratio (LR-) • AUROC *Important outcomes* • Sensitivity • Specificity **Clinical usefulness** *Critical outcomes* • Acceptability to women, to healthcare providers, to the general public • Mental health outcomes *Important outcomes* • Impact on help-seeking behaviour (services sought or utilised) • Impact of detection (e.g. referral rates if screen positive)

Abbreviations: ANRQ-2A, 2 'anxiety' items from the Antenatal Risk Questionnaire; AUROC, area under the receiver-operating characteristics curve; CIDI, Composite International Diagnostic Interview; DASS-21, Depression Anxiety Stress Scales; DSM, Diagnostic and Statistical Manual of Mental Disorders; EPDS, Edinburgh Postnatal Depression Scale; GAD-2, Generalized Anxiety Disorder 2-item scale; GAD-7, Generalized Anxiety Disorder 7-item scale; GHQ, General Health Questionnaire; HADS, Hospital Anxiety and Depression Scale; HADS-A, Hospital Anxiety and Depression Scale – Anxiety subscale; ICD, International Statistical Classification of Diseases and Related Health Problems; K10, Kessler Psychological Distress Scale (10 item); MINI, Mini-International Neuropsychiatric Interview; STAI, State-Trait Anxiety Inventory; SCID, Structured Clinical Interview for DSM Disorders.

As for depression screening, the EWG agreed that the LR+ and LR-, and the AUROC (for different cut-offs) are the critical outcomes for assessing test performance, and sensitivity and specificity are important outcomes.

B2.3 Literature search

As this is a guideline update, the search strings used for the 2017 Australian Guideline were updated to reflect changes in search terminology and additional psychosocial assessment and mental health screening tools since the original search was undertaken. Search strings for identification of evidence relating to psychosocial assessment and mental health screening are shown in **Appendix 1**.

Consistent with the 2017 Australian Guideline, terms related to screening for bipolar disorder, borderline personality disorder, schizophrenia and postpartum psychosis were not included in the evidence review search strings; although these lower prevalence conditions are in scope, they are not typically screened for in primary practice. It is expected that these conditions would be identified through the mental health history item/s that are included as part of a particular psychosocial assessment measure.

Searches were restricted to English-language, full text articles. As per the Research Protocol, primary studies were eligible; conference abstracts, letters, editorials, narrative reviews and dissertations were excluded. The literature searches were conducted on 22 February 2022 in CINAHL and PsychINFO, and on 07 March 2022 in Medline and Embase (using EMBASE.com). The start date was 01 January 2014 for depression screening (the literature searches for NICE included RCTs from the late 1990s to 07 April 2014), and June 2016 for psychosocial assessment and anxiety screening (in line with literature search dates for the 2017 Australian Guideline).

After deduplicating records in EndNote, unique records were uploaded into systematic review software, DistillerSR, for determination of study eligibility.

In addition to the formal literature search, EWG members were provided with a full list of potentially included studies and were asked to forward any additional studies that were missing from the list. Articles identified by the EWG were considered for inclusion if they met the pre-specified eligibility criteria.

The searches did not specifically aim to identify or limit retrieval of articles to studies that addressed socioeconomic, Aboriginal or Torres Strait Islander populations. However, the reviewers were required to document any papers addressing these populations for specific consideration by the EWG. Implications for rural and remote areas, and the Indigenous population, have been considered and documented in the clinical guidance.

B2.4 Study eligibility

For this Evidence Review Update, a two-step eligibility process was undertaken. **Step 1** involved standard inclusion of potentially relevant studies on the basis of the broad PICO criteria outlined in the Research Protocol. **Step 2** required judgement to determine whether potentially included studies from Step 1 met the threshold to inform recommendations for the Australian setting.

B2.4.1 Step 1 – PICO-based eligibility

Study eligibility in **Step 1** was informed by the PICO criteria outlined in Table 2 (psychosocial assessment), Table 3 (screening for depression) and Table 4 (screening for anxiety). All evidence selection criteria were applied in two stages: first to the titles/abstracts and then to the full publications/reports of potentially included studies. Records were excluded for the following reasons:

- Wrong **publication type** – not a full-text report (excludes protocols, conference abstracts, editorials, letters)
- Wrong **population** – study was not conducted in pregnant or postnatal women/birthing parents
- Wrong **intervention** – study did not examine at least one of the psychosocial assessment or mental health screening tools listed in the PICO
- Wrong **comparator** – applies to studies of technical performance without an eligible reference standard (standardised diagnostic interview)
- Wrong **outcome** – study did not examine at least one of the outcomes listed in the PICO
- Not in **English** – full text article not published in English language.

The application of the eligibility criteria above is summarised in **Appendix 1.2**. Note that validation studies of translated versions of mental health screening tools in other countries were excluded as were studies using non-English versions of screening tools in other countries. Studies of tools not specifically mentioned in the PICO (e.g. Perinatal Depression Inventory [PDI-14]) were also excluded.

Appendix 2.1 provides the citation details and reason for exclusion of studies excluded at full text.

B2.4.2 Step 2 – Applying a threshold for appraisal

As this is a guideline update, **Step 2** involved a process to identify studies included in Step 1 that could potentially inform recommendations in the Australian setting. Step 2 was instigated as a pragmatic solution so that evidence appraisal could focus on studies that provide sufficient data on technical performance or clinical usefulness and are applicable to the Australian context. Particular attention was paid to studies of Aboriginal and Torres Strait Islander women, or women from culturally and linguistically diverse backgrounds in Australia.

In total, 19 studies were excluded at Step 2 and are listed in **Appendix 2.2** with reasons for exclusion.

Many of the excluded studies lacked applicability to the Australian context where screening for mental health is already administered online. Studies reporting effectiveness outcomes (e.g. referral rates) were excluded in Step 2 if the data were not comparative (e.g. before implementation of screening compared with post implementation). Qualitative studies of particularly small size (<10 participants) were also excluded, as were systematic reviews with search dates that predated the search dates for the 2017 Australian Guideline.

Appendix 3 provides the citation details for all studies that met the two-step eligibility criteria and were taken through the evidence appraisal process.

B2.5 Assessment of the evidence

B2.5.1 Psychosocial assessment

Quality assessment of studies reporting on the technical performance of psychosocial assessment tools was 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) 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 follows the 'GRADE-style' assessment, which was developed for the 2017 Australian Guideline. 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 aspects of technical performance, non-technical characteristics and clinical usefulness.

To make a judgement 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.

The non-technical aspects of psychosocial assessment (number of items, time to administer, complexity of scoring, and available languages) were addressed in the 2017 Australian Technical Report Part B and are reproduced in this report, with the addition of 'new' tools with evidence of technical performance (namely KMMS and ANRQ-R).

Each of the clinical usefulness aspects of psychosocial assessment has been addressed as a narrative review, based on the findings from the systematic searches. Assessment of the technical performance of the included tools was completed first, and a judgement made by the EWG regarding the strength of the evidence for each tool.

In Australian 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.

B2.5.2 Screening for depression or anxiety

To enable the development of GRADE-style recommendations on mental health screening, an **overall quality** for each screening study has been determined using the following framework (taken from the 2017 Australian Guideline):

- **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.

Complete QUADAS-2 assessments are provided in **Appendix 5** for each included study reporting technical performance outcomes.

Once the results across studies are pooled by type of tool, cut-off threshold and type of mental health issue (depression or anxiety), 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 (●○○○).

Given the place of depression screening in perinatal care, and recognition that under-reporting is more likely than over-reporting, the EWG agreed it is most important to minimise false negatives, even if that is associated with an over-representation of false positives.

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).

Each of the clinical usefulness aspects of psychosocial assessment has been addressed as a narrative review, based on the findings from the systematic searches. Assessment of the technical performance of the included tools was completed first, and a judgement made by the EWG regarding the strength of the evidence for each tool.

B2.6 Evidence to recommendation process

A structured evidence-to-decision framework was used to assist the EWG to develop new recommendations or amend existing recommendations from the 2017 Australian Guidelines.

EWG members were provided with a summary of the recommendations from the 2017 Australian Guideline together with the new evidence identified in the Evidence Review Update (this Technical Report). The EWG met on the 12 September 2022 and agreed on edits to existing recommendations, and the addition of a new practice point. Evidence to decision tables describing the deliberations of the EWG at the 12 September 2022 meeting are provided in an Appendix to the Guideline.

B2.7 Cost-effectiveness

To address potential resourcing implications of perinatal psychosocial assessment and mental health screening, a separate search was undertaken to identify economic evaluations/cost-effectiveness analyses from Australia. Details of the literature search are included in **Appendix 1.1.2**, **Appendix 1.2.2** and **Appendix 3.4**.

The literature search identified one relevant study of integrated psychosocial assessment using tools specified in the PICO (ANRQ-R and EPDS). A summary of this study (Chambers 2022) is provided in **Appendix 5.1.2**.

B3. Psychosocial assessment

What are the most appropriate methods for psychosocial assessment of birthing parents at risk of mental health problems in the perinatal period?

B3.1 Studies identified in the Evidence Review Update

The Evidence Review Update identified three studies investigating methods used for psychosocial assessment of birthing parents at risk of mental health problems in the perinatal period (see Table 5). One study investigated the ANRQ (Kalra 2018), and one study investigated the KMMS (Marley 2017). The other study (Kingston 2017) utilised ALPHA and EPDS, but investigated the acceptability of web-based electronic mental health screening compared with paper-based screening in pregnant women. It did not examine the technical performance or clinical usefulness of ALPHA as a psychosocial assessment tool. One of the studies reported on technical performance (Marley 2017), three reported on acceptability (Kalra 2018, Kingston 2017, Marley 2017) and two reported on implementability (Kingston 2017, Marley 2017).

One study evaluating the ANRQ-R (Reilly 2022) was identified in the Evidence Review Update but was excluded by the EWG because the reference standard (Series of Assessments for Guiding Evaluation – Self-Report [SAGE-SR]) did not meet the eligibility criteria outlined in Section B2.2.1. The ANRQ-R has been identified as an emerging tool by the EWG, and further information about the study by Reilly (2022) is included in Appendix 5.1.

Table 5 List of individual studies included for psychosocial assessment

Study ID	Study type	Tool(s)	Nature of evidence included in identified studies		
			Technical performance	Acceptability	Implementability
Pre-specified tools					
Kalra 2018	Primary	ANRQ (+ EPDS)		✓	
Kingston 2017	Primary	Wed-based screening (ALPHA + EPDS)		✓	✓
Marley 2017	Primary	KMMS	✓	✓	✓

Abbreviations: ALPHA, Antenatal Psychosocial Health Assessment; ANRQ, Antenatal Risk Questionnaire; EPDS, Edinburgh Postnatal Depression Scale; KMMS, Kimberley Mum's Mood Scale.

A summary of the findings for technical performance, clinical usefulness, and an overall summary of the findings is provided below. Where the evidence is not related to a specific screening tool/s, the findings for clinical usefulness are presented as general evidence in section B3.6.

B3.2 Technical performance of relevant tools

Key characteristics of the studies included in the Evidence Review Update are presented in Table 6.

Of the tools specified in the PICO for psychosocial assessment, only one paper has been published in the search period that reported on technical performance (Marley 2017 for KMMS). No new evidence for technical performance was identified/included for the ALPHA, ANRQ/ANRQ-R, ARPA, CAN-M, CAME or PRQ tools.

Marley 2017 conducted a cross-sectional study to determine whether the KMMS is a reliable and valid tool to identify Kimberley Aboriginal perinatal women at risk of anxiety and/or depression compared to a semi-structured clinical interview by a general practitioner (GP). The semi-structured clinical interview was developed by a female Kimberley Aboriginal GP, experienced in mental health assessment and with

extensive connections to the local community. Diagnosis was according to the DSM-IV for anxiety and depressive disorders. Ninety-seven Kimberley Aboriginal perinatal women completed the KMMS, and 91 of these women completed the reference standard GP assessment (97% of reference standard assessments were completed within 24 hours of KMMS completion). Overall, the authors concluded that the KMMS is an effective tool for identifying those at risk of anxiety and depressive disorders amongst Kimberley Aboriginal perinatal women. Results of the study are summarised in Table 6 and Table 7. The authors acknowledged the relatively small sample size (and therefore wide confidence intervals) and the use of a convenience sample as study limitations. They also noted that while GP assessment as the reference standard may be seen as a study limitation, the use of GPs with an understanding of the local context was important for understanding and responding to participants, and preferable to the use of fly-in-fly-out mental health professionals. The study also examined the acceptability of the KMMS to participants and healthcare providers (see Section B3.5).

B3.3 Evidence profile tables

Evidence regarding the predictive capacity of the KMMS is presented in Table 7.

Table 6 Design and psychometric properties of individual studies since 2017

Study ID	Study characteristics (N)	Content validity	Reliability	Applicability	Quality Notes
KMMS					
Marley 2017	*Study design:* Cross-sectional in-community study with reference standard to ascertain validity *Study population:* Pregnant (>6wks) or postpartum (>7d) Aboriginal women, N=97 Blinded DSM-IV assessment (target within 24hr, max 7d) by GP, N=91	*Included domains:* Part 1 covering same areas and scoring as EPDS, moderated by Part 2 Part 2 psychosocial tool based on SAFESTART Guidelines. Domains are key issues identified by Kimberley Aboriginal women: • supports • major stressors • self-esteem, anxiety levels • relationships, • childhood experiences • social emotional wellbeing including substance use *Method of development:* Part 1 adapted from EPDS using 'Kimberley' English, locally developed graphics and a visual Likert scale focusing on feelings not numbers.	*Internal consistency*: Part 1 reliability Cronbach's α= 0.89	*Country:* Australia *Setting:* Indigenous community primary care (15 sites in Kimberley, WA) *Normative data:* Yes; describes sociodemographic factors and psychosocial profile relative to KMMS scores	**High** (●●●●) Based on development of methods, pre-specified population, robust reference standard, validity measures. Directly applicable population and setting.

Abbreviations: DSM-IV, Diagnostic and Statistical Manual of Mental Disorders 4th Edition; EPDS, Edinburgh Postnatal Depression Scale; KMMS, Kimberley Mum's Mood Scale; SAGE-SR, SCID, Structured Clinical Interview for DSM Disorders.

Table 7 Evidence Profile Table: 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	
KMMS				
1 (97)	Marley 2017	Correctly classified: 85.7% Risk equivalence was determined as moderate.	Sensitivity: 83% Specificity: 87% PPV: 68% NPV: 94% AUROC (95% CI) 0.90 (0.83–0.97)	Acceptable **High** (●●●●)
Evidence Statements: *The KMMS is effective at identifying cases of anxiety and depression in Kimberley Aboriginal women (high quality evidence).*				

Abbreviations: AUROC, area under the receiver-operating characteristics curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value

B3.4 Non-technical characteristics of relevant tools

The table below summarises the non-technical characteristics of four psychosocial assessment tools. The ALPHA, ANRQ and PRQ were assessed during the 2017 Australian Guideline development, and all had high or moderate quality evidence of technical performance. The KMMS was added to the research protocol for the Evidence Review Update and was assessed to have high 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 8 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[a]	Simple *Three-point scoring for each question*	English
ANRQ	9 standard items, 3 extra questions if yes answered to certain questions	5-10 minutes	Moderate *Combination of categorical and continuous data (requires skip logic)*	Translated and available digitally in >25 languages (including English)
PRQ	21[b]	10-20 minutes	Moderate *Five-point Likert scale for each question*	English
KMMS	Part 1 – 10 Part 2 – 6 domains	30-60 minutes	Moderate *Four-point scoring for each question in Part 1* *Information from Part 2 is used to contextualise the score from Part 1* *Overall risk of no/low/moderate/high assigned*	English (initially developed for Aboriginal women in the Kimberley region in Western Australia).

Abbreviations: ALPHA, Antenatal Psychosocial Health Assessment; ANRQ, Antenatal Risk Questionnaire; KMMS, Kimberley Mum's Mood Scale; PRQ, Pregnancy Risk Questionnaire.
a assumed based on number of items and comparison with PRQ
b originally 23 items. Latest version has 21 items comprised of 18 antenatal items and 3 early postnatal items.

B3.5 Clinical usefulness of relevant tools

The Kimberley Mum's Mood Scale (KMMS) is a two-part scale that was developed to improve perinatal screening for anxiety and depression in Kimberley Aboriginal women (Marley 2017). Part 1 is adapted from the EPDS and Part 2 is a psychosocial tool used to contextualise the scores from Part 1. **Marley 2017** provided evidence of the acceptability of the KMMS. The tool was administered to 97 Kimberley Aboriginal women who were pregnant or had given birth in the prior 12 months. The acceptability of the KMMS was determined using a qualitative approach. Findings from the study with respect to acceptability and implementability of the KMMS are provided in Table 9 and Table 10 respectively.

Table 9 Evidence regarding the acceptability of the KMMS

Study ID	Acceptability		
	To women (pregnant/postpartum)	To healthcare providers	To general public
Marley 2017	• Median scores for Q1 (understood the questions), Q2 (felt comfortable) and Q5 (talking about childhood/home life) were 10, 10 and 8 (out of 10), respectively (N=81) • Participants found the process easy and useful • The KMMS was accepted by nearly all participants	• 89% reported that doing the KMMS was considerably or extremely useful and superior to the EPDS (N=9) • One study personnel felt that Part 2 of the KMMS was not in their scope of practice • The KMMS was accepted by all but one study personnel	• Not assessed

Abbreviations: KMMS, Kimberley Mum's Mood Scale; EPDS, Edinburgh Postnatal Depression Scale.

Table 10 Evidence regarding implementability of the KMMS

Study ID	Implementability		
	Training requirements	Resource implications	Other
Marley 2017	• Culturally safe training and support	• At least 30 minutes should be allowed by healthcare providers and patients for administering the KMMS	• Further evaluation in a larger Kimberley population is recommended during real world implementation • Testing of applicability in other remote regions required prior to a recommendation for wider use

Abbreviations: KMMS, Kimberley Mum's Mood Scale.

B3.6 General evidence of clinical usefulness

Kingston (2017) reported on the acceptability and feasibility of web-based electronic mental health screening (using the ALPHA and EPDS) compared with paper-based screening in pregnant women. The authors reported that women in both groups felt that they would be able to disclose any mental health concerns, and that women in the e- screening group were more likely to report a preference for e-screening than women in the paper-based screening group. Overall, the authors concluded that e-screening is feasible and acceptable to pregnant women.

Table 11 Evidence regarding the acceptability of web-based screening (using ALPHA and EPDS)

Study ID	Acceptability		
	To women (pregnant)	To healthcare providers	To general public
Kingston 2017	• 57.9% (175/302) women in the e-screening group strongly or somewhat agreed that they would like to use or did like using a tablet to answer questions on emotional health (vs 37.2% (121/325) in the paper-based screening group). • 46.0% (139/302) women in the e-screening group would or did prefer using a tablet to paper (vs 29.2% (95/325) in the paper-based screening group).	• Not assessed	• Not assessed

Abbreviations: ALPHA, Antenatal Psychosocial Health Assessment; EPDS, Edinburgh Postnatal Depression Scale.

Table 12 Evidence related to the implementability of web-based screening (using ALPHA and EPDS)

Study ID	Implementability	
	Training requirements	Feasibility
Kingston 2017	• Minimal staff training	• Findings support the feasibility of e-screening • Women in both the e-screening and paper screening groups reported being able to disclose concerns regarding their mental health

Abbreviations: ALPHA, Antenatal Psychosocial Health Assessment; EPDS, Edinburgh Postnatal Depression Scale.

Kalra 2018 included evidence of the acceptability of the ANRQ and EPDS administered as part of routine psychosocial assessment in a private maternity setting in Australia. The study involved a retrospective audit of medical records of 455 women who received routine screening (ANRQ and EPDS administered at the first antenatal booking visit), as well as a prospective survey of 101 women who underwent routine psychosocial assessment at their first booking visit.

The authors reported that the acceptability of routine psychosocial assessment among the sample of 101 women was high, with the women's experience of the assessment being overwhelmingly positive. Acceptability was assessed using a 12-question online survey developed by the researchers. The survey included statements such as "I felt comfortable being asked questions about my emotional health", and "I

think enquiry about emotional health is an important part of antenatal care", with women asked to indicate their level of agreement with each statement. Acceptability to healthcare providers and the general public was not assessed.

Table 13 Evidence regarding the acceptability of psychosocial assessment (using ANRQ and EPDS)

Study ID	Acceptability		
	To women (pregnant) (N)	To healthcare providers	To general public
Kalra 2018	• Acceptability of depression screening and psychosocial risk assessment was high (N=101)	• Not assessed	• Not assessed

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

B3.7 Overall summary of findings

Table 14 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, the tool that scores highest across all domains of interest is the ANRQ. 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.

Table 14 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]	Implementability[7]
ALPHA	Limited	Moderate (●●●○)	Moderate	English; Cultural sensitivity unknown	Moderate	Limited
ANRQ	Acceptable	Moderate (●●●○)	High	Translated and available digitally in >25 languages (including English) Cultural sensitivity unknown	High	High
PRQ	Acceptable	Moderate (●●●○)	Moderate	English; Cultural sensitivity unknown	Unknown	Limited
KMMS	Acceptable	High (●●●●)	Low	English only; initially developed for Aboriginal women in the Kimberley region in Western Australia. Developed to be culturally specific to Kimberley Indigenous women. Being trialled in other Aboriginal and Torres Strait Islander population groups.	High	Context specific

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 judgement 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] 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; KMMS, Kimberley Mum's Mood Scale; PRQ, Pregnancy Risk Questionnaire; EWG, Expert Working Group.

B4. Screening for depression

What are the most appropriate methods for screening the birthing parent for depression in the perinatal period?

B4.1 Studies identified in the Evidence Review Update

The Evidence Review Update identified 14 studies investigating methods used for depression screening in the perinatal period (see Table 15). One additional study (Kingston 2017) investigated the acceptability of web-based electronic mental health screening compared with paper-based screening in pregnant women (using EPDS and ALPHA) but is not included in this section because it did not examine the technical performance or clinical usefulness of EPDS as a depression screening tool.

Of the 14 studies, three addressed both anxiety and depression screening (Chan 2022, Blackmore 2022, Nithianandan 2016). The screening tools evaluated across the 14 studies included EPDS, PHQ-9, PHQ-2, Whooley questions, and EPDS via the iCOPE digital platform translated to Dari. One study (Nithianandan 2016) investigated barriers and enablers to perinatal mental health screening, and implementation of screening and referral programs in women of refugee background. Five of the studies reported on technical performance (Wang 2021, Ezirim 2021, Blackmore 2022, Levis 2020, Kotz 2021), nine reported on acceptability of depression screening (Chan 2021, Kotz 2021, Chan 2022, Vik 2021, Lawson 2019, Yapp 2019, Logsdon 2018, BenDavid 2016, Nithianandan 2016), two reported comparative effectiveness (Avalos 2016, BenDavid 2016), and one reported on implementability (Chan 2021).

Table 15 List of individual studies included for depression screening

Study ID	Study type	Tool(s)	Nature of evidence included in identified studies			
			Technical performance	Acceptability	Comparative effectiveness	Implementability
Blackmore 2022	Primary	EPDS (iCOPE digital platform) translated to Dari	✓			
Chan 2022	Primary	EPDS		✓		
Chan 2021	SR	EPDS		✓		✓
Kotz 2021	SR	EPDS	✓	✓		
Wang 2021	SR	PHQ-9	✓			
Ezirim 2021	Primary	EPDS	✓			
Vik 2021	Primary	EPDS		✓		
Levis 2020	SR	EPDS	✓			
Lawson 2019	Primary	PHQ-2 via text		✓		
Yapp 2019	Primary	Whooley		✓		
Logsdon 2018	Primary	EPDS		✓		
Avalos 2016	Primary	PHQ-9			✓	
BenDavid 2016	Primary	EPDS		✓	✓	
Nithianandan 2016	Primary	EPDS		✓		

Abbreviations: EPDS, Edinburgh Postnatal Depression Scale; iCOPE, Centre of Perinatal Excellence digital screening platform; PHQ-2, first 2 items of the PHQ-9; PHQ-9, Patient Health Questionnaire-9.

A summary of the findings for technical performance, clinical usefulness, and an overall summary of the findings is provided below. Where the evidence is not related to a specific screening tool/s, the findings for clinical usefulness are presented as general evidence in Section B4.5 (e.g. electronic versus paper

screening). The systematic reviews that reported on technical performance of the PHQ-9 (Wang 2021) and EPDS (Levis 2020) have been summarised narratively below (detailed data from the publications are not reproduced in this report).

B4.2 Technical performance of relevant tools

B4.2.1 Evidence from systematic reviews

B4.2.1.1 EPDS for screening to detect major depression

Levis (2020) conducted a study on behalf of the DEPRESsion Screening Data (DEPRESSD) EPDS group, an international collaborative project that aims to synthesize the global depression screening data in order to develop and apply rigorous methods to assessing depression screening tool accuracy (https://www.depressd.ca/home). This study was a systematic review and meta-analysis of individual patient datasets from included studies to examine the accuracy of the EPDS for screening to detect major depression among pregnant and postpartum women. Eligible datasets included both EPDS scores and major depression classification based on validated diagnostic interviews. EPDS sensitivity and specificity were estimated compared with fully structured, semi structured and MINI diagnostic interviews separately. Eligible individual patient data were able to be obtained for 58 studies (15,557 participants, 2069 with major depression).

The authors principal findings were based on the studies using semi-structured interviews, which are designed to closely replicate clinical diagnoses by mental health professionals. Among these studies (9066 participants, 1330 with major depression), the sensitivity and specificity with 95% confidence intervals were as follows:

- cut-off score of 10 or higher, sensitivity 0.85 (0.79 to 0.90) and specificity 0.84 (0.79 to 0.88)
- cut-off score of 11 or higher, sensitivity 0.81 (0.75 to 0.87) and specificity 0.88 (0.85 to 0.91)
- cut-off score of 13 or higher, sensitivity 0.66 (0.58 to 0.74) and specificity 0.95 (0.92 to 0.96).

Accuracy was similar across reference standards and subgroups, including for pregnant and postpartum women.

The authors concluded that an EPDS cut-off value of 11 or higher maximised combined sensitivity and specificity; a cut-off value of 13 or higher was less sensitive but more specific. To identify pregnant and postpartum women with higher symptom levels, a cut-off of 13 or higher could be used. *Lower cut-off values could be used if the intention is to avoid false negatives and identify most patients who meet diagnostic criteria.*

The authors also noted that their recommendation of a major depression cut-off score of 11 or higher was at odds with the most commonly used EPDS cut-off score of 13 or higher, and the recommendation of 12 or higher from a previous meta-analysis.

B4.2.1.2 PHQ-9 versus EPDS for screening for perinatal depression

Wang (2021) conducted a systematic review and meta-analysis to examine the validity of the PHQ-9 for screening for perinatal depression. This review included 35 studies (19,760 participants); 10 studies (5,235 participants) looked at the criterion validity of the PHQ-9 compared with standard psychiatric interview, and 25 studies examined the convergent validity of the PHQ-9 compared with the EPDS or other validated depression measures in a perinatal population. Of the 19 studies that administered both the PHQ-9 and EPDS to participants, 15 provided one or more psychometric comparisons between the two scales.

Meta-analysis of the 7 criterion validity studies[1] using the standard PHQ-9 cut-off point ≥10 showed pooled sensitivity 0.84 (95%CI 0.75 to 0.90), specificity 0.81 (95%CI 0.74 to 0.86) and AUC of 0.89. Operating characteristics of the PHQ-9 and EPDS were nearly identical in head-to-head comparison studies. The median correlation between the PHQ-9 and EPDS was 0.59 (based on six studies), and categorical agreement was moderate. The authors concluded that the PHQ-9 appears to be a viable option for perinatal depression screening with operating characteristics similar to the EPDS.

B4.2.1.3 KMMS and EPDS in Aboriginal and Torres Strait Islander women

A systematic review by **Kotz (2021)** aimed to investigate the cultural safety of the EPDS in the Australian Indigenous context. This review included four studies but found no studies of psychometric or qualitative validation of the standard EPDS for Australian Indigenous women. They concluded that there is currently no evidence to support its use as a screening tool in this population.

Among the four included studies, one was a small validation study that compared the KMMS with a blinded mental health assessment by an Indigenous general practitioner as the gold standard (Marley 2017). This study found the KMMS to have 87% sensitivity and 75% specificity with a positive predictive value of 54.1% and a negative predictive value of 94.4%. The authors noted that with the added inclusion of psychosocial inquiry, these results appeared promising. However, due to the small sized, convenience samples, the results did not adequately meet evaluation criteria that would support generalised use.

B4.2.2 Evidence from primary studies

B4.2.2.1 Characteristics of included studies

The key characteristics of the two primary studies identified in the Evidence Review Update that reported on relevant technical performance outcomes for depression screening are presented in Table 16. The table summarises the quality of each study, as assessed using QUADAS-2, together with the overall quality rating.

The Australian study by **Blackmore 2022** was undertaken to validate the use of a Dari translation of the EPDS among women of refugee background. Fifty-two Dari-speaking women were administered the EPDS Dari version via the iCOPE digital platform at a public pregnancy clinic in Melbourne. A structured clinical interview for DMS-5 Research Version (SCID-5-RV) was conducted as the reference standard, blind to EPDS screening results. The results of the study are presented in Table 17 and Table 18. Overall, the authors concluded that study results support the use of the EPDS Dari version to screen for depression and anxiety symptoms during pregnancy. The results also supported the use of a lowered cut-off score of ≥9 for depression screening (consistent with the recommendation for culturally and linguistically diverse groups), and a cut-off of ≥5 for anxiety screening. The study was judged to be of high quality based on QUADAS-2 methods (see Table 16).

The study by **Ezirim 2021** aimed to determine the efficacy of the EPDS immediately postpartum (i.e. 3-24 hours postpartum). The authors hypothesised that if EPDS scores immediately postpartum were predictive of EPDS scores at the 6-week postpartum visit, psychiatric services may be easier to access through the inpatient setting. EPDS data from 848 participants collected immediately postpartum and again at 6-weeks postpartum were analysed as part of this longitudinal observational study. The study found that EPDS scores immediately postpartum are not a reliable predictor of elevated EPDS scores at 6-weeks postpartum. Findings from the study are presented in Table 17 and Table 18. The quality of the study was assessed as 'very low' using QUADAS-2 methods (see Table 16) due to a high risk of selection bias, risk of

[1] One of the 7 studies were assessed as having low risk of bias across all four QUADAS domains, three studies had low risk of bias in three domains, two studies had low risk of bias in two domains and one study had low risk of bias in only one domain.

bias associated with the reference standard (the reference standard was only undertaken on participants with an elevated EPDS) and risks associated with the timing of assessments (see Table App. 17).

B4.2.2.2Evidence summaries

Evidence summaries of the two primary studies for technical performance of depression screening are presented in Table 17.

B4.2.2.3 Summary of findings regarding technical performance

Table 18 presents the Summary of Findings for the critical and important outcomes, as determined by the EWG. Unpooled sensitivity and specificity results are presented. Where LR+ and LR- values were not reported by study authors, these have been calculated based on the corresponding 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 presented according to the population in the studies: antenatal women only, or postnatal women only.

Table 16 Key characteristics of primary studies using screening tools to identify perinatal depression

Study ID	Tool(s)	Country; Setting (Population sample)	Case Definition	Patient selection		Index test(s)		Ref. Standard		Flow & Timing	Study Quality[a]
				Risk of bias	Applicability concerns	Risk of bias	Applicability concerns	Risk of bias	Applicability concerns	Risk of bias	
Antenatal											
Blackmore 2022	EPDS	Australia; Antenatal clinic (Community – refugees)	Any current depressive or anxiety disorder	Low	Low	Low	Low	Low	Low	Low	**High**
Postnatal											
Ezirim 2021	EPDS	US; Hospital, maternity ward (Community)	Depression	High	Low	Unclear	Low	High	Low	High	**Very Low**

Abbreviations: EPDS, Edinburgh Postnatal Depression Scale; US, Unites States.
a Determined by the EWG based on QUADAS-2 methods (see **Appendix 5**).

Table 17 Evidence Summary: EPDS for detection of depression in perinatal women

Citation	Population	N	Cut-off	Reference Standard	LR+ (95% CI)	LR– (95% CI)	Consistency Cronbach's α	Sensitivity (95% CI)	Specificity (95% CI)	AUROC (95% CI)	Youden index	Study Quality
Antenatal												
Blackmore 2022	Antenatal, Australia (CALD group, women of refugee background with Dari as their preferred language)	52	>9	SCID-5-RV	NR [PPV = 0.80 (0.56–0.94)]	NR [NPV = 1.00 (0.88–1.00)]	0.79	1.00 (0.79–1.00)	0.88 (0.73–0.97	0.90 (0.82–0.99	NR	High
Postnatal												
Ezirim 2021	Postpartum (3-24 hours postpartum)	848	>9	DSM-V	4.21 (2.9–6.2)	0.76 (0.7–0.9)]	NR	28.79 % (21.2–37.3)	93.16 % (91.1–94.9)	0.76 (0.73–0.79)	0.424	Very Low

Abbreviations: AUROC, area under the receiver-operating characteristics curve; CALD, culturally and linguistically diverse; CI, Confidence Interval; EPDS, Edinburgh Postnatal Depression Scale; LR+, positive likelihood ratio; LR-, negative likelihood ratio, NPV, Negative Predictive Value; NR, Not Reported; PPV, positive predictive value.

Table 18 Summary of Findings: EPDS for detection of depression in perinatal women

Tool; Condition; Cut-off	No. of studies (participants)	Critical outcomes			Important outcomes			Overall certainty
		LR+ (95% CI)	LR– (95% CI)	AUROC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Youden index	
Antenatal								
EPDS; any current depressive or anxiety disorder; >9	1 (52)[1]	8.33[2] [CI ranges: NR]	0.01[2] [CI ranges: NR]	0.90 (0.82–0.99)	1.00 (0.79–1.00)	0.88 (0.73–0.97)	NR	●●○○ Low [3]
Postnatal								
EPDS; depression (not specified); >9	1 (848)[4]	4.21 (2.9–6.2)	0.76 (0.7–0.9)	0.76 (0.73–0.79)	28.79 % (21.2–37.3)	93.16 % (91.1–94.9)	0.424	●○○○ Very low[5]
Evidence statements: *It is uncertain if the Dari version of the EPDS has adequate sensitivity or specificity to detect any depressive or anxiety disorder in Dari-speaking pregnant women of refugee background (low certainty evidence).* *It is uncertain if the EPDS has adequate sensitivity or specificity to detect depression in women at 3-24 hours postpartum (very low certainty evidence).*								

Footnotes:

[1] Blackmore 2022
[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] One study of high quality.
[4] Ezirim 2021
[5] One study of very low quality.

Abbreviations: AUROC, area under the receiver-operating characteristics curve; CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; LR+, positive likelihood ratio; LR-, negative likelihood ratio, NPV, negative predictive value; NR, not reported; PPV, positive predictive value.

B4.3 Non-technical characteristics of relevant tools

The table below summaries the non-technical characteristics of the EPDS. The complexity of scoring was assessed as Simple, Moderate or High on the basis of information in the published literature and the experience of the EWG.

Table 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 Available digitally in >25 languages

Abbreviations: EPDS, Edinburgh Postnatal Depression Scale.

B4.4 Clinical usefulness of relevant tools

B4.4.1 Acceptability of depression screening

B4.4.1.1 Depression screening of Aboriginal and Torres Strait Islander women

Chan (2022) examined the acceptability of the EPDS via semi-structured interviews/yarns with 13 pregnant Aboriginal women and 10 non-Aboriginal midwives based in Perth. A grounded theory approach using thematic analysis was used to analyse the interview transcripts. The authors reported that, overall, the pregnant women found the EPDS easy and straightforward to use. In contrast, the midwives perceived that the responses of the pregnant women may be influenced by what they felt was the desirable answer, and they perceived that the women felt the EPDS was a waste of time and not meaningful. “Midwives strongly advised pairing the EPDS with conversation to understand psychosocial and contextual factors, with a focus on relationship building and prioritisation of engagement” (p7). The authors advised that in the absence of a validated perinatal mental health screening tool for Aboriginal women, using the EPDS seems preferable to not screening.

Chan (2021) conducted a systematic review of the cultural suitability of the EPDS to Indigenous mothers globally. For the purposes of this report, findings from the Australian studies are summarised. In terms of health care providers views of the EPDS, the results were mixed. Some considered it acceptable, whilst others found it culturally inappropriate, were not confident in the results due to concerns such as language barriers or felt it did not accurately identify issues. One Australian study (Freeman 2017) included in the review found the rate of EPDS completion in Indigenous women (43%) to be lower than the rate for mothers in the general population (98%). The review authors acknowledged that a lack of discussion of the potential reasons for EPDS rejection was a design limitation of the original study. Overall, the review concluded that the acceptability of the EPDS in an indigenous context remains unclear.

In a systematic review by **Kotz (2021)**, the cultural safety of the EPDS in the Australian Indigenous context was investigated. The review found that the EPDS had not been validated for use with Australian Indigenous women, and as such concluded that there is currently no evidence to support its use as a screening tool in this population. Whilst the acceptability of the EPDS to Australian Indigenous women was not specifically addressed, one study included in the review identified that some Indigenous mothers and healthcare professionals noted difficulties in understanding some language in the EPDS (Campbell 2008), and another noted biases in women’s responses (Kotz 2016).

B4.4.1.2 Depression screening of women of refugee background in Australia

Nithianandan (2016) investigated the factors affecting the implementation of perinatal mental health screening in women of refugee background in Australia. They conducted semi-structured interviews with 28 health professionals and 9 women of refugee background and analysed the data via thematic analysis to identify barriers and enablers to screening. They identified behavioural change techniques and provided recommendations for implementation of screening and referral programs in women of refugee background. The authors reported that "almost all participants perceived the need for, and understood the rationale behind, routine antenatal mental health screening" (p4), and that participants generally felt the EPDS was appropriate and easy to understand when they had an accurate interpreter.

In terms of health professional acceptance of the EPDS, the authors reported inconsistent feedback from health professionals, with some considering the EPDS the best available screening tool, and others citing concerns regarding its cultural appropriateness. In terms of implementation, staff training, inter-disciplinary roles to support referral, and clear and robust referral pathways were identified by health professionals as factors affecting implementation of mental health screening in women of refugee background.

Women of refugee background identified continuity of care, female interpreters and health professionals, social support and useful follow-up care as priorities. In-person interpreters, translated EPDS versions, and sufficient time and capacity for mental health screening were identified as important environmental factors. The authors recommended providing culturally appropriate information about mental health screening at earlier appointments to improve acceptance of subsequent screening. Recommendations were also made to normalise screening, manage referral expectations, reinforce the professionalism of interpreters and the importance of follow-up care in order to improve women's engagement with the health services. The authors recommended that further research be undertaken to examine cross-cultural understandings of each EPDS item across key refugee communities.

B4.4.1.3 Depression screening in international populations

Vik (2021) examined the experiences of Norwegian health visitors and midwives (all registered nurses) with their use of the EPDS. Data was collected via two focus group interviews (N=6 participants per group) and analysed using thematic network analysis. The authors reported that, overall, the EPDS was a well-accepted screening tool by the participants in the study. It provided them with information about mental health problems in mothers, aided in the initiation of conversations about problems of early motherhood, and promoted multidisciplinary collaborations.

BenDavid (2016) conducted a pilot project of a telephone-based screening program for PPD in an urban primary care setting serving a vulnerable population in the United States. Prior to the project, PPD screening was not routinely performed at the health centre involved. Participants were screened using the EPDS via telephone between two to three weeks postpartum. The authors reported that 100% of participants (n=27) found the PPD screening acceptable, however no detail was provided regarding how acceptability was measured, and therefore it is not clear whether acceptability related to the EPDS as a tool, the mode of screening (i.e. telephone), or PPD screening more generally.

Yapp (2019) examined the acceptability of depression screening (using the Whooley questions) among women attending an antenatal booking appointment at an inner-city service in the United Kingdom. Face to face interviews were conducted with 545 women (287 Whooley positive, 258 Whooley negative) and analysed qualitatively using thematic and framework approaches. The women were asked "What was it like for you answering the questions about your mood?", "Were there any questions you found upsetting, distressing or confronting?" and "Did the midwife give you some feedback about your answers?" (p128). The study reported that the majority of women found the questions acceptable to ask, however some women experienced difficulty due to concerns about the consequences of a disclosure, because of their emotional responses to the questions, or because they felt the midwife did not validate their responses.

Some women reported that the screening felt rushed due to time constraints of the booking appointment, or like a tick-box exercise. In terms of study implications, the authors recommended that midwives should normalise mental health disclosures, provide reassurance that mental health concerns can be supported, and explore mental health concerns in order to establish appropriate support. Furthermore, women should be provided with an opportunity to disclose mental health concerns without others present and be reassured that mental health concerns are very common and can be accommodated. The importance of appropriate training for midwives was highlighted by the authors.

B4.4.2 Effectiveness of depression screening

Avalos (2016) conducted a retrospective cohort study (n=97,678) to assess the effectiveness of implementing a perinatal depression screening program at a large group practice in Northern California, United States. Following implementation of the program, women were screened using the PHQ-9 at their first pre-natal visit, at 26-28 weeks gestation, and at 3-8 weeks postpartum. Routine perinatal mental health screening was not in place prior to the implementation of the program. Identification of new perinatal depression diagnosis increased from 8.2% pre-implementation to 11.7% post-implementation. Contrary to expectations, the observed percentage of women with a new depression diagnosis who received treatment decreased from pre-implementation (60.9%) to post-implementation (47.1%). The authors hypothesised that this was due to underdiagnosis of depression in the pre-implementation phase and hence a smaller number of women identified as needing treatment. After adjusting for this, the percentage of women receiving treatment of those expected to have a new depression diagnosis increased (42.6% pre-implementation vs. 47.1% post-implementation). Depression symptoms improved significantly at up to 6 months post-diagnosis following implementation of the program; however, there was insufficient data available in the pre-implementation phase to make comparisons.

In addition to assessing acceptability (see B4.4.1), **BenDavid (2016)** also reported rates of referral acceptance by women who screened positive on the EPDS in an urban primary care setting in the United States. Acceptance rates in the intervention group were compared with historical data reported in the literature. Referrals to primary care providers for further mental health assessment and possible treatment were accepted by 64.3% of the women, and referrals to support services were also accepted by 64.3% of women. This exceeded the 60% benchmark established based on historical data. Rates of referral were not compared to a control group of no-screening or an alternative screening tool.

B4.4.3 Implementation of depression screening

In Australia, the iCOPE app is currently freely available to all public hospitals and maternal child health clinics as part of the Commonwealth Government-funded National Perinatal Mental Health Check initiative. The digital platform screens for symptoms of depression and anxiety using the EPDS and assesses psychosocial risk factors using the ANRQ. iCOPE is also being rolled out to the private healthcare sector.

The systematic review by **Chan (2021)** briefly touched on implementation of the EPDS in Indigenous populations. The authors acknowledged that adaptive approaches to implementing the EPDS as reported by some health professionals, such as employing a conversational style approach rather than structured question and answer, may assist with engagement, but may also have implications for both validity and interpretation of EPDS results clinically and in research.

B4.5 General evidence of clinical usefulness

Lawson (2019) examined the feasibility of using text messages for mental health screening in the immediate postpartum period (using the PHQ-2). A survey of participants' satisfaction with the text messages was conducted at 12-13 weeks postpartum, with a response rate of 63% (N=589). Of the

respondents, 78% recommended that all women be screened for postpartum depression via text messaging. Eighty-seven percent (N=511) reported a preference for screening via text messages to other forms of screening such as face-to-face or via telephone. Based on the results of the survey the authors concluded that screening for postpartum depression via text messaging was feasible and well accepted. Of note, Lawson 2019 also reported technical performance of the PHQ-2 using the EPDS as a reference standard.

Logsdon (2018) evaluated the acceptability to new mothers of postpartum depression screening using the EPDS on the night before hospital discharge. Acceptability was assessed via a telephone interview conducted by research staff two to four weeks after hospital discharge. Participants were asked ''Was it acceptable for a hospital-based perinatal nurse to ask you about symptoms of PPD before hospital discharge?''(p326). The majority of participants (81%) responded positively to this question, leading the authors to report that the majority of new mothers found depression screening by hospital-based perinatal nurses acceptable.

B4.6 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 20. This table is reproduced from the 2017 Technical Report. The 'new' evidence does not alter the conclusions from 2017.

With regard to the timing of EPDS administration, findings by Ezirim (2021) do not support administration of the EPDS in the immediate postpartum period (i.e. within 3-24 hours postpartum).

Whilst new evidence was available on the acceptability of the EPDS to Aboriginal and Torres Strait Islander women, the results of these studies were unclear. One new study was available regarding the acceptability of the EPDS to women of refugee background in Australia (Nithianandan 2016). In general, acceptability of the EPDS was high amongst women of refugee background; however, acceptability amongst health professionals was unclear. The authors provided detailed implementation recommendations to support the implementation of perinatal screening programs in women of refugee background.

Table 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 Postnatal: Acceptable	●●●● High ●●●● High	High	>50 languages (including English) Available digitally in >25 languages Multiple populations	High	Good	High
PHQ-9	Antenatal: Uncertain Postnatal: Uncertain	●●○○ Low ●●○○ Low	High	>50 languages (including English)[2] Western populations	Unknown but likely to be Good	Limited	High
Whooley questions	Antenatal: Uncertain Postnatal: Uncertain	●●○○ Low ●○○○ Very low	High	English Western populations	High	Limited	High
K10	Antenatal: Uncertain Postnatal: Uncertain	●●○○ Low ●○○○ Very low	High	At least 35 languages (including English)[2] Western populations	Unknown but likely to be Good	Unknown	High

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 judgement 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 utilised, 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-9, Patient Health Questionnaire-9; K10, Kessler Psychological Distress Scale (10 item)

[2] https://www.dhi.health.nsw.gov.au/transcultural-mental-health-centre-tmhc/health-professionals/cross-cultural-mental-health-care-a-resource-kit-for-gps-and-health-professionals/multilingual-assessment-tools

B5. Screening for anxiety

What are the most appropriate methods for screening the birthing parent for anxiety in the perinatal period?

B5.1 Studies identified in the Evidence Review Update

The Evidence Review Update identified five studies investigating methods used for anxiety screening in the perinatal period (see Table 21). One additional study (Kingston 2017) investigated the acceptability of web-based electronic mental health screening compared with paper-based screening in pregnant women (using EPDS and ALPHA) but is not included in this section because it did not examine the technical performance or clinical usefulness of EPDS as an anxiety screening tool.

Of the five studies, three addressed both anxiety and depression screening (Chan 2022, Blackmore 2022, Nithianandan 2016). The screening tools evaluated across the six studies included GAD-2, GAD-7, EPDS, EPDS-3A, and EPDS via the iCOPE digital platform translated to Dari. One study (Nithianandan 2016) investigated barriers and enablers to perinatal mental health screening, and implementation of screening and referral programs in women of refugee background. Two of the studies reported on technical performance (Blackmore 2022, Fairbrother 2019), two reported on acceptability (Chan 2022, Nithianandan 2016), and one reported on comparative effectiveness (Lieb 2020).

One additional study from Australia (Austin 2022) was identified in the Evidence Review Update but was excluded by the EWG as the reference standard (SAGE-SR) did not meet the eligibility criteria outlined in Section B2.2.3.

Table 21 List of individual studies included for anxiety screening

Study ID	Study type	Tool(s)	Nature of evidence included in identified studies			
			Technical performance	Acceptability	Comparative Effectiveness	Implementability
Blackmore 2022	Primary	EPDS (iCOPE digital platform) translated to Dari	✓			
Chan 2022	Primary	EPDS		✓		
Lieb 2020	Primary	GAD-2 (+PHQ-2)			✓	
Fairbrother 2019	Primary	EPDS; EPDS-3A; GAD-7; GAD-2	✓			
Nithianandan 2016	Primary	EPDS		✓		

Abbreviations: EPDS, Edinburgh Postnatal Depression Scale; EPDS-3A, Edinburgh Postnatal Depression Scale – Anxiety subscale; GAD-2, Generalized Anxiety Disorder 2-item scale; GAD-7, Generalized Anxiety Disorder 7-item scale; iCOPE, Centre of Perinatal Excellence digital screening platform; PHQ-2, first 2 items of the PHQ-9.

A summary of the findings for technical performance, clinical usefulness, and an overall summary of the findings is provided below.

B5.2 Technical performance of relevant tools

B5.2.1 Characteristics of included studies

The key characteristics of the two primary studies identified in the Evidence Review Update that reported on relevant technical performance outcomes for anxiety screening are presented in Table 22. The table summarises the quality of each study, as assessed using QUADAS-2, together with the overall quality rating.

B5.2.2 Evidence summaries

Evidence summaries of the two primary studies for technical performance of anxiety screening are presented in Table 23.

The Australian study by **Blackmore 2022** is relevant to depression and anxiety screening and was judged to be of high quality. A brief summary of the study is provided in Section B4.2.2.

Fairbrother 2019 presented an assessment of tools for screening for anxiety and related disorders in 310 Canadian women at approximately 3 months postpartum. The authors evaluated five tools: the EPDS and its anxiety subscale (EPDS-3A), the GAD-7 and GAD-2 scales and a clinically derived alternative 'Anxiety Disorder – 13' scale (AD-13; not in use in Australia). The tools were used to screen for core anxiety disorders (panic, agoraphobia, generalised anxiety disorder, social anxiety disorder, and specific phobia) and a broader set of core plus two additional related disorders (obsessive compulsive disorder [OCD] and post-traumatic stress disorder [PTSD]). The authors note there is limited evidence to support use of these four tools in a perinatal population for anxiety screening.

The authors pre-specified four performance criteria they considered suitable for a tool intended for widespread clinical application: AUC ≥ 0.8; Youden's index ≥ 0.5; NPV ≥ 0.8, and LR+ ≥ 4.0. All five tools met the NPV criterion, but none of the four tools in use in Australia (EPDS, EPDS-3A, GAD-7, GAD-2) met the other criteria (only the AD-13 met all four criteria). Based on AUC values for these four tools, there was a general correlation between better tool performance and greater number of specific anxiety items in the tool. For example, the GAD-7 (a seven-item scale) did best, followed by GAD-2 (two items) and EPDS-3A (three items) performing similarly and the EPDS tool having the lowest AUC (three anxiety items but with the depression-focused domains included).

The study concluded that since these tools did not meet the performance criteria *"Neither the EPDS/EPDS 3-A, nor the GAD-7/GAD-2 can be recommended for widespread use as a perinatal [anxiety disorder] screening tool"*. However, this study was judged to be of very low quality (Table 22) because the reference standard (SCID-IV) was conducted up to 7 weeks after the index tests and was only undertaken in women with elevated screening results (partial verification bias).

B5.2.3 Summary of findings regarding technical performance

The tables below present the Summary of Findings for each tool identified in the Evidence Review Update: EPDS (Table 24), EPDS-3A (Table 25), GAD-7 (Table 26) and GAD-2 (Table 27).

Results are grouped together 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.

Table 22 Key characteristics of primary studies using screening tools to identify perinatal anxiety

Study ID	Tool(s)	Country; Setting (Population sample)	Case Definition	Patient selection		Index test(s)		Ref. Standard		Flow & Timing	Study Quality[a]
				Risk of bias	Applicability concerns	Risk of bias	Applicability concerns	Risk of bias	Applicability concerns	Risk of bias	
Antenatal											
Blackmore 2022	EPDS	Australia; Antenatal clinic (Community – refugees)	Any current depressive or anxiety disorder	Low	Low	Low	Low	Low	Low	Low	**High**
Perinatal											
Fairbrother 2019	EPDS, EPDS-3A, GAD-7, GAD-2	Canada; Perinatal(Community)	Full composite of AD; 5 core anxiety disorders[c]	Unclear	Low	Low	Low	High	Low	High	**Very Low**

Abbreviations: AD, anxiety and related disorders; EPDS, Edinburgh Postnatal Depression Scale; EPDS-3A, Edinburgh Postnatal Depression Scale – Anxiety subscale; GAD-2, Generalized Anxiety Disorder 2-item scale; GAD-7, Generalized Anxiety Disorder 7-item scale.

a Determined by the EWG based on QUADAS-2 methods (see **Appendix 5**).

b Generalised anxiety disorder, panic disorder, agoraphobia, social anxiety disorder, obsessive compulsive disorder.

c Five core anxiety disorders: panic disorder, agoraphobia, generalised anxiety disorder, social anxiety disorder, and specific phobia.

Table 23 Evidence Summary: tools for detection of anxiety in perinatal women

Citation	Population	N	Cut-off	Reference Standard	LR+ (95% CI)	LR– (95% CI)	Consistency Cronbach's α	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	Youden index	Study Quality
EPDS												
Blackmore 2022	Antenatal (women of refugee background, preferred language Dari)	52	≥5	SCID-5-RV	NR [PPV = 0.58 (0.33–0.80)]	NR [NPV = 1.00 (0.89–1.00)]	0.79	1.00 (0.72–1.00)	0.80 (0.65–0.91)	0.94 (0.88–1.00)	NR	High
Fairbrother 2019	Perinatal (35wk; 3mo postpartum)	310 [n=115]	6	SCID-IV	1.91 [CI ranges: NR]	0.498 [CI ranges: NR]	NR	0.680 [CI ranges: NR]	0.643 [CI ranges: NR]	0.744 (0.663–0.824)	<0.5	Very Low
EPDS-3A												
Fairbrother 2019	Perinatal (35wk; 3mo postpartum)	310 [n=115]	4	SCID-IV	2.87 [CI ranges: NR]	0.463 [CI ranges: NR]	NR	0.640 [CI ranges: NR]	0.777 [CI ranges: NR]	0.757 (0.678–0.836)	<0.5	Very Low
GAD-7												
Fairbrother 2019	Perinatal (35wk; 3mo postpartum)	310 [n=115]	6	SCID-IV	3.78 [CI ranges: NR]	0.516 [CI ranges: NR]	NR	0.560 [CI ranges: NR]	0.852 [CI ranges: NR]	0.780 (0.704–0.856)	<0.5	Very Low
GAD-2												
Fairbrother 2019	Perinatal (35wk; 3mo postpartum)	310 [n=115]	2	SCID-IV	3.03 [CI ranges: NR]	0.540 [CI ranges: NR]	NR	0.560 [CI ranges: NR]	0.815 [CI ranges: NR]	0.752 (0.675–0.829)	<0.5	Very Low

Abbreviations: AUC, area under the curve; CALD, culturally and linguistically diverse; CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; EPDS-3A, Edinburgh Postnatal Depression Scale – Anxiety subscale; GAD-2, Generalized Anxiety Disorder 2-item scale; GAD-7, Generalized Anxiety Disorder 7-item scale; LR+, positive likelihood ratio; LR-, negative likelihood ratio; mo, months; NPV, negative predictive value; NR, not reported; PPV, positive predictive value; SCID-5-RV, Structured Clinical Interview for DSM Disorders, 5th edition Research Version; wk, weeks.

Table 24 Summary of Findings: EPDS for detection of anxiety in perinatal women

Tool; Condition; Cut-off	No. of studies (participants)	Critical outcomes			Important outcomes			Overall certainty
		LR+ (95% CI)	LR– (95% CI)	AUROC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Youden index	
Antenatal								
EPDS; any current depressive or anxiety disorder; ≥5	1 (52)[1]	5.00[2] [CI ranges: NR]	0.013[2] [CI ranges: NR]	0.94 (0.88–1.00)	1.00 (0.72–1.00)	0.80 (0.65–0.91)	NR	●●○○ Low[3]
Perinatal								
EPDS; full composite of anxiety and related disorders; 6	1 (115)[4]	1.91 [CI ranges: NR]	0.498 [CI ranges: NR]	0.744 (0.663–0.824)	0.680	0.643	<0.5	●○○○ Very low[5]
Evidence statement: *It is uncertain if the Dari version of the EPDS has adequate sensitivity or specificity to detect any depressive or anxiety disorder in Dari-speaking pregnant women of refugee background (low certainty evidence).* *It is uncertain if the EPDS has adequate sensitivity or specificity to detect the full composite of anxiety and related disorders in perinatal women (very low quality evidence).*								
Footnotes: [1] Blackmore (2022) [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] One study of high quality. [4] Fairbrother (2019) [5] One study of very low quality.								

Abbreviations: AUROC, area under the receiver-operating characteristics curve; CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; LR+, positive likelihood ratio; LR-, negative likelihood ratio; NR, not reported.

Table 25 Summary of Findings: EPDS-3A for detection of anxiety in perinatal women

Tool; Condition; Cut-off	No. of studies (participants)	Critical outcomes			Important outcomes			Overall certainty
		LR+ (95% CI)	LR– (95% CI)	AUROC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Youden index	
Perinatal								
EPDS-3A; full composite of anxiety and related disorders; 4	1 (115)[1]	2.87 [CI ranges: NR]	0.463 [CI ranges: NR]	0.757 (0.678–0.836)	0.640 [CI ranges: NR]	0.777 [CI ranges: NR]	<0.5	●○○○ Very low[2]
Evidence statement: *It is uncertain if the EPDS-3A has adequate sensitivity or specificity to detect anxiety disorders in perinatal women (very low quality evidence).*								
Footnotes: [1] Fairbrother (2019) [2] One study of very low quality.								

Abbreviations: AUROC, area under the receiver-operating characteristics curve; CI, confidence interval; EPDS-3A, Edinburgh Postnatal Depression Scale – Anxiety subscale; GAD, generalised anxiety disorder; LR+, positive likelihood ratio; LR-, negative likelihood ratio; NR, not reported.

Table 26 Summary of Findings: GAD-7 for detection of anxiety in perinatal women

Tool; Condition; Cut-off	No. of studies (participants)	Critical outcomes			Important outcomes			Overall certainty
		LR+ (95% CI)	LR– (95% CI)	AUROC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Youden index	
Perinatal								
GAD-7; full composite of anxiety and related disorders; 6	1 (115)[1]	3.78 [CI ranges: NR]	0.516 [CI ranges: NR]	0.780 (0.704–0.856)	0.560 [CI ranges: NR]	0.852 [CI ranges: NR]	<0.5	●○○○ Very low[2]
Evidence statement: *It is uncertain if the GAD-7 has adequate sensitivity or specificity to detect anxiety disorders in perinatal women (very low quality evidence).*								
Footnotes: [1] Fairbrother (2019) [2] One study of very low quality.								

Abbreviations: AUROC, area under the receiver-operating characteristics curve; CI, confidence interval; GAD, generalised anxiety disorder; GAD-7, Generalized Anxiety Disorder 7-item scale; LR+, positive likelihood ratio; LR-, negative likelihood ratio; NR, not reported.

Table 27 Summary of Findings: GAD-2 for detection of anxiety in perinatal women

Tool; Condition; Cut-off	No. of studies (participants)	Critical outcomes			Important outcomes			Overall certainty
		LR+ (95% CI)	LR– (95% CI)	AUROC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Youden index	
Perinatal								
GAD-2; full composite of anxiety and related disorders; 2	1 (115)[1]	3.03 [CI ranges: NR]	0.540 [CI ranges: NR]	0.752 (0.675–0.829)	0.560 [CI ranges: NR]	0.815 [CI ranges: NR]	<0.5	●○○○ Very low[2]
Evidence statement: *It is uncertain if the GAD-2 has adequate sensitivity or specificity to detect anxiety disorders in perinatal women (very low quality evidence).*								
Footnotes: [1] Fairbrother (2019) [2] One study of very low quality.								

Abbreviations: AUROC, area under the receiver-operating characteristics curve; CI, confidence interval; GAD, generalised anxiety disorder; GAD-2, Generalized Anxiety Disorder 2-item scale; LR+, positive likelihood ratio; LR-, negative likelihood ratio; NR, not reported.

B5.3 Non-technical characteristics of relevant tools

The table below summaries the non-technical characteristics of the included anxiety screening tools for which there was evidence of technical performance in the 2017 Australian Guideline or the Evidence Review Update. 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 28 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 Available digitally in >25 languages
EPDS-3A	3	<5 mins	Simple	Developed in English and translated into >50 languages Available digitally in >25 languages
GAD-7	7	5-10 mins	Simple	Developed in English Translated into >50 languages [3]
GAD-2	2	<5 mins[a]	Simple[a]	English
K10	10	5-10 mins	Simple	At least 35 languages (including English)[4]
STAI	40	>10 mins	Complex	English

Abbreviations: EPDS, Edinburgh Postnatal Depression Scale; EPDS-3A, Edinburgh Postnatal Depression Scale – Anxiety subscale; GAD-2, Generalized Anxiety Disorder 2-item scale; GAD-7, Generalized Anxiety Disorder 7-item scale; K10, Kessler Psychological Distress Scale (10 item); STAI, State-Trait Anxiety Inventory

a assumed based on number of items and comparison with GAD-7

B5.4 Clinical usefulness of relevant tools

Chan (2022) and **Nithianandan (2016)** both reported on the clinical usefulness of the EPDS for mental health screening. The findings from these studies have been summarised in Section B4 on screening for depression.

Lieb (2020) investigated the effectiveness of adding anxiety screening (using GAD-2) to existing depression screening (using PHQ-2) in pregnant women in the United States. The study was not powered to distinguish the primary outcomes between the two groups due to higher-than-expected exclusion rates. Whilst the addition of anxiety screening did detect women who would have been missed by depression screening alone, the difference in positive screening rates between the groups was not statistically significant (OR 3.24, 95% CI 0.39-26.88). Overall, the rate of referral during pregnancy was not significantly different between groups (OR 1.95, 95% CI, 0.76-4.97), however, the authors noted an increase in referrals in the depression plus anxiety screening group for women with a history of mental health diagnosis or substance abuse, leading them to conclude that anxiety screening may be particularly useful in these populations.

[3] https://www.phqscreeners.com/select-screener

[4] https://www.dhi.health.nsw.gov.au/transcultural-mental-health-centre-tmhc/health-professionals/cross-cultural-mental-health-care-a-resource-kit-for-gps-and-health-professionals/multilingual-assessment-tools

B5.5 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 29. This table is reproduced from the 2017 Technical Report, with the addition of the GAD-2. The 'new' evidence does not alter the conclusions from 2017.

Table 29 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	Antenatal: Uncertain Perinatal: Uncertain	●●○○ Low	High	Developed in English and translated into >50 languages Available digitally in >25 languages Multiple populations[9]	High[9]	Unknown	High
EPDS-3A	Antenatal: Uncertain Perinatal: Uncertain	●○○○ Very low	High	Developed in English and translated into >50 languages Available digitally in >25 languages Western populations	Unknown But likely to be Good	Unknown	High
GAD-7	Antenatal: Uncertain Perinatal: Uncertain	●○○○ Very low	High	English Western populations	Unknown But likely to be Good	Unknown	Moderate
GAD-2	Antenatal: Uncertain Perinatal: Uncertain	●○○○ Very low	High	English	Unknown	Unknown	High
K10	Antenatal: Uncertain	●○○○ Very low	High	At least 35 languages (including English)[5] 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 judgement 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 utilised, 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 on 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; EPDS-3A, Edinburgh Postnatal Depression Scale – Anxiety subscale; EWG, Expert Working Group; GAD-2, Generalized Anxiety Disorder 2-item scale; GAD-7, Generalized Anxiety Disorder 7-item scale; K10, Kessler Psychological Distress Scale (10 item); STAI, State-Trait Anxiety Inventory.

[5] https://www.dhi.health.nsw.gov.au/transcultural-mental-health-centre-tmhc/health-professionals/cross-cultural-mental-health-care-a-resource-kit-for-gps-and-health-professionals/multilingual-assessment-tools

Appendix 1 Literature search

1.1 Search strings

The updated literature search covered the period from **01 January 2014**[6] to **07 March 2022**. The search was undertaken to identify studies of any type relating to psychosocial assessment or mental health screening (depression or anxiety) in the perinatal period using the tools specified in the PICO.

1.1.1 Technical performance and clinical usefulness

Table App. 1 EMBASE.com search string – Psychosocial assessment and mental health screening

Search set		EMBASE.com search string (concurrently searches Embase and Medline)	Records
Perinatal period	#1	pregnancy:ti,ab,kw OR pregnant:ti,ab,kw OR perinatal:ti,ab,kw OR 'peri natal':ti,ab,kw OR peripartum:ti,ab,kw OR 'peri partum':ti,ab,kw OR prenatal:ti,ab,kw OR 'pre natal':ti,ab,kw OR postnatal:ti,ab,kw OR 'post natal':ti,ab,kw OR postpartum:ti,ab,kw OR 'post partum':ti,ab,kw OR antenatal:ti,ab,kw OR 'ante natal':ti,ab,kw OR antepartum:ti,ab,kw OR 'ante partum':ti,ab,kw OR parturition:ti,ab,kw OR puerper*:ti,ab,kw OR maternal:ti,ab,kw OR 'after birth':ti,ab	1,195,034
Mental health problems	#2	'depression'/exp OR 'postnatal depression'/exp OR depress*:ti,ab,kw OR 'anxiety disorder'/exp OR anxiety:ti,ab,kw OR 'mental health'/exp OR 'mental disease'/exp OR 'mental health':ti,ab,kw OR 'mental disorder':ti,ab,kw OR 'mood disorder'/exp OR 'mood disorder':ti,ab,kw	3,030,094
Psychosocial assessment tools	#3	'antenatal psychosocial health assessment':ti,ab,kw OR 'ante natal psychosocial health assessment':ti,ab,kw OR 'antenatal risk questionnaire':ti,ab,kw OR 'ante natal risk questionnaire':ti,ab,kw OR anrq:ti,ab,kw OR 'antenatal routine psychosocial assessment':ti,ab,kw OR arpa:ti,ab,kw OR 'camberwell assessment of need*':ti,ab,kw OR 'can m':ti,ab,kw OR 'contextual assessment of maternity experience':ti,ab,kw OR 'kimberly mum*s mood scale':ti,ab,kw OR 'pregnancy risk questionnaire':ti,ab,kw OR 'postnatal risk questionnaire':ti,ab,kw OR 'post natal risk questionnaire':ti,ab,kw OR 'perinatal risk questionnaire':ti,ab,kw OR 'peri natal risk questionnaire':ti,ab,kw OR prq:ti,ab,kw OR pnrq:ti,ab,kw	798
	#4	#1 AND #2 AND #3	50
Depression screening tools	#5	'depression anxiety stress scale':ti,ab,kw OR 'depression, anxiety and stress scale':ti,ab,kw OR dass*:ti,ab,kw OR 'edinburgh postnatal depression scale':ti,ab,kw OR 'edinburgh post natal depression scale':ti,ab,kw OR epds:ti,ab,kw OR 'hospital anxiety and depression scale':ti,ab,kw OR 'hads-a':ti,ab,kw OR hads:ti,ab,kw OR 'kessler psychological distress scale':ti,ab,kw OR 'k 10':ti,ab,kw OR k10:ti,ab,kw OR 'k-10':ti,ab,kw OR 'patient health questionnaire':ti,ab,kw OR 'phq':ti,ab,kw OR 'phq-2':ti,ab,kw OR 'phq-9':ti,ab,kw OR 'whooley questions':ti,ab,kw	49,227
	#6	#1 AND #2 AND #5	5,663
Anxiety screening tools	#7	'general anxiety disorder-7':ti,ab,kw OR 'generalised anxiety disorder-7':ti,ab,kw OR 'generalized anxiety disorder-7':ti,ab,kw OR 'general anxiety disorder-2':ti,ab,kw OR 'generalised anxiety disorder-2':ti,ab,kw OR 'generalized anxiety disorder-2':ti,ab,kw OR 'gad 7':ti,ab,kw OR 'gad-7':ti,ab,kw OR 'gad 2':ti,ab,kw OR 'gad-2':ti,ab,kw OR 'general health questionnaire':ti,ab,kw OR ghq:ti,ab,kw OR 'state trait anxiety inventory':ti,ab,kw OR 'state-trait anxiety inventory':ti,ab,kw OR stai:ti,ab,kw	22,803
	#8	#1 AND #2 AND #7	1,703
Year limits	#9	#4 AND [2016-2022]/py	24
	#10	#6 AND [2014-2022]/py	3,678
	#11	#8 AND [2016-2022]/py	757
Combined	#12	#9 OR #10 OR #11	4,151
Other limits	#13	#12 NOT ([conference abstract]/lim OR [conference review]/lim OR [letter]/lim OR [editorial]/lim)	3,103
	#14	#13 NOT [animals]/lim	3,077
	#15	#14 AND [english]/lim	3,000
SR set	#16	'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 OR 'technology assessment':ab,ti OR hta:ab,ti OR 'cochrane':ab,ti	591,719
	#17	#15 AND #16	**82**
Primary study set	#18	#15 NOT #17	**2,918**

[6] In the 2017 Australian Guideline, the NICE 2015 SR (with a literature search in 2014) was used as the basis for the clinical evidence relating to depression screening. The search for evidence relating to anxiety screening and psychosocial assessment was conducted in 2016.

Table App. 2 CINAHL search string – Psychosocial assessment and mental health screening

Search set		CINAHL search string	Limiters/Expanders	Records
Perinatal period	S1	(MH "Postnatal Period") OR (MH "Pregnancy") OR (MH "Prenatal Care") OR (MH "Perinatal Care")	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	
	S2	AB ((pregnancy or pregnant or perinatal or peri natal or peripartum or peri partum or prenatal or pre natal or postnatal or post natal or postpartum or post partum or antenatal or ante natal or antepartum or ante partum or parturition or puerper* or maternal or after birth)) OR TI ((pregnancy or pregnant or perinatal or peri natal or peripartum or peri partum or prenatal or pre natal or postnatal or post natal or postpartum or post partum or antenatal or ante natal or antepartum or ante partum or parturition or puerper* or maternal or after birth))	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	
	S3	S1 OR S2	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	
Mental health problems	S4	(MH "Depression, Postpartum") OR (MH "Mental Health") OR (MH "Mental Disorders") OR (MH "Affective Disorders") OR (MH "Anxiety Disorders")	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	
	S5	TI ((depress* or anxiety or mental health or mental disorder* or mood disorder*)) OR AB ((depress* or anxiety or mental health or mental disorder* or mood disorder*))	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	
	S6	S4 OR S5	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	
Psychosocial assessment tools	S7	(MH "Diagnosis, Psychosocial") OR (MH "Psychological Tests+")	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	
	S8	TI ((antenatal psychosocial health assessment or ante natal psychosocial health assessment or alpha or antenatal risk questionnaire* or ante natal risk questionnaire* or anrq or 'Antenatal routine psychosocial assessment or ARPA or Camberwell assessment of need or Camberwell assessment of needs or CAN-M or contextual assessment of maternity experience or kimberly mum* mood scale or pregnancy risk questionnaire or postnatal risk questionnaire or post natal risk questionnaire or perinatal risk questionnaire or peri natal risk questionnaire or prq or PNRQ or risk factor assessment)) OR AB ((antenatal psychosocial health assessment or ante natal psychosocial health assessment or alpha or antenatal risk questionnaire* or ante natal risk questionnaire* or anrq or 'Antenatal routine psychosocial assessment or ARPA or Camberwell assessment of need or Camberwell assessment of needs or CAN-M or contextual assessment of maternity experience or kimberly mum* mood scale or pregnancy risk questionnaire or postnatal risk questionnaire or post natal risk questionnaire or perinatal risk questionnaire or peri natal risk questionnaire or prq or PNRQ or risk factor assessment))	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	
	S9	S7 OR S8	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	
Depression screening tools	S10	TI ((depression anxiety stress scale or dass* or edinburgh postnatal depression scale or edinburgh post natal depression scale or epds or hospital anxiety depression scale or HADS-A or HADS or kessler psychological distress scale or k 10 or k10 or k-10 or patient health questionnaire or phq or phq-2 or phq-9 or Whooley question*)) OR AB ((depression anxiety stress scale or dass* or edinburgh postnatal depression scale or edinburgh post natal depression scale or epds or hospital anxiety depression scale or HADS-A or HADS or kessler psychological distress scale or k 10 or k10 or k-10 or patient health questionnaire or phq or phq-2 or phq-9 or Whooley question*))	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	

Search set		CINAHL search string	Limiters/Expanders	Records
Anxiety screening tools	S11	TI ((general anxiety disorder-7 or generalised anxiety disorder-7 or generalized anxiety disorder-7 or general anxiety disorder-2 or generalised anxiety disorder-2 or generalized anxiety disorder-2 or gad 7 or gad-7 or gad 2 or gad-2 or general health questionnaire or GHQ or state trait anxiety inventory or state-trait anxiety inventory or stai)) OR AB ((general anxiety disorder-7 or generalised anxiety disorder-7 or generalized anxiety disorder-7 or general anxiety disorder-2 or generalised anxiety disorder-2 or generalized anxiety disorder-2 or gad 7 or gad-7 or gad 2 or gad-2 or general health questionnaire or GHQ or state trait anxiety inventory or state-trait anxiety inventory or stai))	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	
Limits	S12	S9 OR S11	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	
	S13	S3 AND S6 AND S10	Limiters - English Language; Published Date: 20140101-20221231 Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	
	S14	S3 AND S6 AND S12	Limiters - English Language; Published Date: 20160101-20221231 Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	
Combined	S15	S13 OR S14	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	
SR set	S16	(TI (systematic* n3 review*)) or (AB (systematic* n3 review*)) or (TI (systematic* n3 bibliographic*)) or (AB (systematic* n3 bibliographic*)) or (TI (systematic* n3 literature)) or (AB (systematic* n3 literature)) or (TI (comprehensive* n3 literature)) or (AB (comprehensive* n3 literature)) or (TI (comprehensive* n3 bibliographic*)) or (AB (comprehensive* n3 bibliographic*)) or (TI (integrative n3 review)) or (AB (integrative n3 review)) or (JN "Cochrane Database of Systematic Reviews") or (TI (information n2 synthesis)) or (TI (data n2 synthesis)) or (AB (information n2 synthesis)) or (AB (data n2 synthesis)) or (TI (data n2 extract*)) or (AB (data n2 extract*)) or (TI (medline or pubmed or psyclit or cinahl or (psycinfo not "psycinfo database") or "web of science" or scopus or embase)) or (AB (medline or pubmed or psyclit or cinahl or (psycinfo not "psycinfo database") or "web of science" or scopus or embase)) or (MH "Systematic Review") or (MH "Meta Analysis") or (TI (meta-analy* or metaanaly*)) or (AB (meta-analy* or metaanaly*))	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	
	S17	S15 AND S16	Limiters - English Language Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	**232**
Deduplicated				**231**
Primary study set	S18	S15 NOT S17	Limiters - English Language Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	**4,624**
Deduplicated				**4,583**

Table App. 3 PsychINFO search string – Psychosocial assessment and mental health screening

Search set		PsychINFO search string	Records
Perinatal period	1	(pregnancy or pregnant or perinatal or peri natal or peripartum or peri partum or prenatal or pre natal or postnatal or post natal or postpartum or post partum or antenatal or ante natal or antepartum or ante partum or parturition or puerper* or maternal or after birth).ti,ab.	129,372
	2	Perinatal Period/ or Pregnancy/ or Prenatal Care/ or Postnatal Period/	32,772
	3	or/1-2	131,516
Mental health problems	4	depression/ or postnatal depression/ or anxiety disorder/ or mental health/ or mental disease/ or mood disorder/ or postpartum depression/ (139196)	139,196
	5	(depress* or anxiety or mental health or mental disorder* or mood disorder*).ti,ab.	636,289

Search set		PsychINFO search string	Records
	6	or/4-5	654,727
Psychosocial assessment tools	7	(antenatal psychosocial health assessment or ante natal psychosocial health assessment or alpha or antenatal risk questionnaire* or ante natal risk questionnaire* or anrq or Antenatal routine psychosocial assessment or ARPA or Camberwell assessment of need or Camberwell assessment of needs or CAN-M or contextual assessment of maternity experience or kimberly mum* mood scale or pregnancy risk questionnaire or postnatal risk questionnaire or post natal risk questionnaire or perinatal risk questionnaire or peri natal risk questionnaire or prq or PNRQ or risk factor assessment).ti,ab.	57,746
	8	exp Psychosocial Assessment/	304
	9	or/7-8	58,036
Depression screening tools	10	(depression anxiety stress scale or dass* or edinburgh postnatal depression scale or edinburgh post natal depression scale or epds or hospital anxiety depression scale or HADS-A or HADS or kessler psychological distress scale or k 10 or k10 or k-10 or patient health questionnaire or phq or phq-2 or phq-9 or Whooley question*).ti,ab.	14,108
Anxiety screening tools	11	(general anxiety disorder-7 or generalised anxiety disorder-7 or generalized anxiety disorder-7 or general anxiety disorder-2 or generalised anxiety disorder-2 or generalized anxiety disorder-2 or gad 7 or gad-7 or gad 2 or gad-2 or general health questionnaire or GHQ or state trait anxiety inventory or state-trait anxiety inventory or stai).ti,ab.	12,906
Limits	12	and/3,6,10	2,436
	13	limit 12 to yr="2014 -Current"	1,401
	14	or/9,11	70,489
	15	and/3,6,14	1,193
	16	limit 15 to yr="2016 -Current"	437
	17	or/13,16	1,699
	18	limit 17 to (human and english language)	**1,401**
Deduplicated			**1,398**

1.1.2 Economic analyses

A literature search was conducted on **03 November 2022** to update the search undertaken in May 2017 for economic analyses of perinatal psychosocial assessment and mental health screening. The literature search was restricted to Australian economic analyses of any tools for psychosocial assessment or mental health screening, including any of the tools specified in the PICO.

Table App. 4 EMBASE.com search string – Economic analyses of psychosocial assessment and mental health screening

Search set		EMBASE.com search string (concurrently searches Embase and Medline)	Records
Economic/cost terms	#1	'economic evaluation'/exp OR 'health care cost'/de OR 'economic model'/exp OR 'health utility'/de OR 'economics'/de	689,546
	#2	(((cost* OR economic OR markov) NEAR/3 (model OR analysis OR analyses)):ti,ab,kw) OR 'cost impact$':ti,ab,kw OR 'economic impact$':ti,ab,kw OR 'cost outcome$':ti,ab,kw OR 'budget impact$':ti,ab,kw	123,036
	#3	'life year$':ti,ab,kw OR qaly$:ti,ab,kw	41,001
Psychosocial assessment tools	#4	'antenatal psychosocial health assessment$':ti,ab,kw OR 'ante natal psychosocial health assessment$':ti,ab,kw OR 'antenatal risk questionnaire':ti,ab,kw OR 'ante natal risk questionnaire':ti,ab,kw OR anrq:ti,ab,kw OR 'antenatal routine psychosocial assessment$':ti,ab,kw OR arpa:ti,ab,kw OR 'camberwell assessment of need*':ti,ab,kw OR 'can m':ti,ab,kw OR 'contextual assessment of maternity experience$':ti,ab,kw OR 'kimberly mum*s mood scale':ti,ab,kw OR 'pregnancy risk questionnaire':ti,ab,kw OR 'postnatal risk questionnaire':ti,ab,kw OR 'post natal risk questionnaire':ti,ab,kw OR 'perinatal risk questionnaire':ti,ab,kw OR 'peri natal risk questionnaire':ti,ab,kw OR prq:ti,ab,kw OR pnrq:ti,ab,kw	842
Depression screening tools	#5	'depression, anxiety and stress scale':ti,ab,kw OR dass*:ti,ab,kw OR 'edinburgh postnatal depression scale':ti,ab,kw OR 'edinburgh post natal depression scale':ti,ab,kw OR epds:ti,ab,kw OR 'hospital anxiety and depression scale':ti,ab,kw OR 'hads-a':ti,ab,kw OR hads:ti,ab,kw OR 'kessler psychological distress scale':ti,ab,kw OR 'k 10':ti,ab,kw OR k10:ti,ab,kw OR 'k-10':ti,ab,kw OR 'patient health questionnaire':ti,ab,kw OR 'phq':ti,ab,kw OR 'phq-2':ti,ab,kw OR 'phq-9':ti,ab,kw OR 'whooley questions':ti,ab,kw	53,732

Search set		EMBASE.com search string (concurrently searches Embase and Medline)	Records
Anxiety screening tools	#6	'general anxiety disorder-7':ti,ab,kw OR 'generalised anxiety disorder-7':ti,ab,kw OR 'generalized anxiety disorder-7':ti,ab,kw OR 'general anxiety disorder-2':ti,ab,kw OR 'generalised anxiety disorder-2':ti,ab,kw OR 'generalized anxiety disorder-2':ti,ab,kw OR 'gad 7':ti,ab,kw OR 'gad-7':ti,ab,kw OR 'gad 2':ti,ab,kw OR 'gad-2':ti,ab,kw OR 'general health questionnaire':ti,ab,kw OR ghq:ti,ab,kw OR 'state trait anxiety inventory':ti,ab,kw OR 'state-trait anxiety inventory':ti,ab,kw OR stai:ti,ab,kw	24,588
General psychosocial assessment, depression, and anxiety screening terms	#7	((psychosocial NEAR/3 assessment):ti,ab,kw) OR 'psychosocial assessment'/de OR 'psychosocial assessment tool'/de OR ((('mental health' OR depression OR anxiety) NEAR/3 (screening OR screen)):ti,ab,kw)	15,322
Australia terms	#8	australian:ti,ab,kw OR australia:ti,ab,kw OR nsw:ti,ab,kw OR 'new south wales':ti,ab,kw OR victoria:ti,ab,kw OR qld:ti,ab,kw OR queensland:ti,ab,kw OR tasmania:ti,ab,kw OR 'northern territory':ti,ab,kw OR 'western australia':ti,ab,kw OR 'australian capital territory':ti,ab,kw	242,949
Economic evaluation	#9	#1 OR #2 OR #3	739,022
Psychosocial assessment and screening tools	#10	#4 OR #5 OR #6 OR #7	85,084
Australian studies	#11	#8 AND #9 AND #10	65
Year limit	#12	#11 AND [2017-2022]/py	34

Table App. 5 Cochrane Library search string – Economic analyses of psychosocial assessment and mental health screening

Search set		Cochrane Library search string	Records
Economic/cost terms	#1	MeSH descriptor: [Costs and Cost Analysis] explode all trees	11,580
	#2	(((cost OR economic OR markov) NEAR/3 (model OR analysis OR analyses)):ti,ab,kw) OR "cost impact*":ti,ab,kw OR "economic impact*":ti,ab,kw OR "cost outcome*":ti,ab,kw OR "budget impact*":ti,ab,kw	25,607
	#3	("life year*" OR QALY*):ti,ab,kw	6,109
Psychosocial assessment tools	#4	"antenatal psychosocial health assessment":ti,ab,kw OR "ante natal psychosocial health assessment":ti,ab,kw OR "antenatal risk questionnaire":ti,ab,kw OR "ante natal risk questionnaire":ti,ab,kw OR anrq:ti,ab,kw OR "antenatal routine psychosocial assessment*":ti,ab,kw OR arpa:ti,ab,kw OR "camberwell assessment of need*":ti,ab,kw OR "can m":ti,ab,kw OR "contextual assessment of maternity experience*":ti,ab,kw OR "kimberly mum*s mood scale":ti,ab,kw OR "pregnancy risk questionnaire":ti,ab,kw OR "postnatal risk questionnaire":ti,ab,kw OR "post natal risk questionnaire":ti,ab,kw OR "perinatal risk questionnaire":ti,ab,kw OR "peri natal risk questionnaire":ti,ab,kw OR prq:ti,ab,kw OR pnrq:ti,ab,kw	66
Depression screening tools	#5	"depression, anxiety and stress scale":ti,ab,kw OR dass*:ti,ab,kw OR "edinburgh postnatal depression scale":ti,ab,kw OR "edinburgh post natal depression scale":ti,ab,kw OR epds:ti,ab,kw OR "hospital anxiety and depression scale":ti,ab,kw OR "hads-a":ti,ab,kw OR hads:ti,ab,kw OR "kessler psychological distress scale":ti,ab,kw OR "k 10":ti,ab,kw OR k10:ti,ab,kw OR "k-10":ti,ab,kw OR "patient health questionnaire":ti,ab,kw OR "phq":ti,ab,kw OR "phq-2":ti,ab,kw OR "phq-9":ti,ab,kw OR "whooley questions":ti,ab,kw	16,556
Anxiety screening tools	#6	"general anxiety disorder-7":ti,ab,kw OR "generalised anxiety disorder-7":ti,ab,kw OR "generalized anxiety disorder-7":ti,ab,kw OR "general anxiety disorder-2":ti,ab,kw OR "generalised anxiety disorder-2":ti,ab,kw OR "generalized anxiety disorder-2":ti,ab,kw OR "gad 7":ti,ab,kw OR "gad-7":ti,ab,kw OR "gad 2":ti,ab,kw OR "gad-2":ti,ab,kw OR "general health questionnaire":ti,ab,kw OR ghq:ti,ab,kw OR "state trait anxiety inventory":ti,ab,kw OR "state-trait anxiety inventory":ti,ab,kw OR stai:ti,ab,kw	7,082
General psychosocial assessment, depression, and anxiety screening terms	#7	(psychosocial NEAR/3 assessment):ti,ab,kw OR (('mental health' OR depression OR anxiety) near/3 (screening OR screen)):ti,ab,kw	3,244
Australia terms	#8	(Australian or Australia or NSW or "New South Wales" or Victoria or QLD or Queensland or Tasmania or "Northern Territory" OR "Western Australia" OR "Australian Capital Territory"):ti,ab,kw with Cochrane Library publication date Between Jan 2017 and Sep 2022	21,731

Search set		Cochrane Library search string	Records
Economic evaluation	#9	#1 OR #2 OR #3	28,769
Psychosocial assessment and screening tools in Australia	#10	(#4 OR #5 OR #6 OR #7) AND #8	1,117
Economic evaluation of psychosocial assessment and screening tools	#11	#9 AND #10	86
Year limit	#12	#11 with Cochrane Library publication date Between Jan 2017 and Nov 2022	73

1.2 Study inclusion/exclusion

1.2.1 Technical performance and clinical usefulness

Table App. 6 Inclusion/exclusion – Studies for psychosocial assessment and screening

	No. records
Records identified via literature search of Embase and Medline on 07 March 2022	3,000
Records identified via literature searches of CINAHL and PsychINFO on 22 February 2022	6,212
Records identified manually or via EWG	0
Total records identified, excluding duplicates	6,598
Records included after title/abstract screen	95
Records excluded after full text screen	55
Excluded – wrong publication type	2
Excluded – wrong population	2
Excluded – wrong intervention	7
Excluded – no comparator	3
Excluded – wrong comparator	12
Excluded – wrong outcomes	16
Excluded - duplicate	4
Excluded – already included in 2017 guideline	9
Records included after full text screen	40
*Records not taken through evidence appraisal process (see **Appendix 2.2** for reasons)*	*21*
Total records included for psychosocial assessment (**Section B3**)	3
Total records included for screening for mental health problems	16
Screening for depression (**Section B4**)	14[a]
Screening for anxiety (**Section B5**)	5[a]

a Three studies included for depression screening were also relevant to anxiety screening.

1.2.2 Economic analyses

Table App. 7 Inclusion/exclusion – Studies reporting economic analyses

	No. records
Records identified via literature search of Embase.com on 03 November 2022	34
Records identified via literature searches of Cochrane Library on 03 November 2022	73
Records identified manually or via EWG	0
Total records identified, excluding duplicates	94
Records excluded after title/abstract screen	93

	No. records
Excluded – wrong population	58
Excluded – wrong intervention	32
Excluded – wrong outcome	2
Excluded – wrong study type	1
Records included after title/abstract screen	1
Records included after full text screen	1

Appendix 2 Excluded studies list

This appendix contains citation details for studies that were excluded at full text review (Section 2.1) and studies that were not taken through the evidence appraisal process (Section 2.2). Refer to Section B2.4 for a description of the two-step eligibility process.

2.1 Studies excluded at full text review, with reason for exclusion

Alzahrani, A. D. (2019). Risk Factors for Postnatal Depression among Primipara Mothers. The Spanish journal of psychology, 22(e35), 1-8
Reason for exclusion: wrong outcome
Ref ID: 406

Anderson, M. L., Wolf Craig, K. S., Hostovsky, S., Bligh, M., Bramande, E., Walker, K., Biebel, K., Byatt, N. (2021). Creating the Capacity to Screen Deaf Women for Perinatal Depression: A Pilot Study. Midwifery, 92(2021), 102867
Reason for exclusion: wrong outcome
Ref ID: 418

Arefadib, N., Cooklin, A., Nicholson, J., Shafiei, T. (2021). Postnatal depression and anxiety screening and management by maternal and child health nurses in community settings: A scoping review. Midwifery, 100(2021), 103039
Reason for exclusion: wrong outcome
Ref ID: 3

Badiya, P. K., Siddabattuni, S., Dey, D., Hiremath, A. C., Nalam, R. L., Srinivasan, V., Vaitheswaran, S., Ganesh, A., Prabhakar, Y., Ramamurthy, S. S. (2021). Effect of mode of administration on Edinburgh Postnatal Depression Scale in the South Indian population: A comparative study on self-administered and interviewer-administered scores. Asian Journal of Psychiatry, 66(2021), 102890
Reason for exclusion: wrong outcome
Ref ID: 477

Berger, E., Wu, A., Smulian, E. A., Quiñones, J. N., Curet, S., Marraccini, R. L., Smulian, J. C. (2015). Universal versus risk factor-targeted early inpatient postpartum depression screening. Journal of Maternal-Fetal and Neonatal Medicine, 28(7), 739-744
Reason for exclusion: wrong outcome
Ref ID: 557

Bhat, A., Nanda, A., Murphy, L., Ball, A. L., Fortney, J., Katon, J. (2022). A systematic review of screening for perinatal depression and anxiety in community-based settings. Archives of Women's Mental Health, 25(1), 33-49
Reason for exclusion: wrong outcome
Ref ID: 5

Bränn, E., Fransson, E., Wikman, A., Kollia, N., Nguyen, D., Lilliecreutz, C., Skalkidou, A. (2021). Who do we miss when screening for postpartum depression? A population-based study in a Swedish region. Journal of Affective Disorders, 287(2021), 165-173
Reason for exclusion: wrong outcome
Ref ID: 633

Brodey, B. B., Goodman, S. H., Baldasaro, R. E., Brooks-DeWeese, A., Wilson, M. E., Brodey, I. S. B., Doyle, N. M. (2016). Development of the Perinatal Depression Inventory (PDI)-14 using item response theory: a comparison of the BDI-II, EPDS, PDI, and PHQ-9. Archives of Women's Mental Health, 19(2), 307-316
Reason for exclusion: wrong intervention type
Ref ID: 645

Chorwe-Sungani, G., Chipps, J. (2017). A systematic review of screening instruments for depression for use in antenatal services in low resource settings. BMC Psychiatry (2017) 17:112
Reason for exclusion: duplicate of Ref ID 8
Ref ID: 6502

Chorwe-Sungani, G., Chipps, J. (2018). Performance of the 3-item screener, the Edinburgh Postnatal Depression Scale, the Hopkins Symptoms Checklist-15 and the Self-Reporting Questionnaire and Pregnancy Risk Questionnaire, in screening of depression in antenatal clinics in the Blantyre district of Malawi. Malawi medical journal: the journal of Medical Association of Malawi, 30(3), 184-190
Reason for exclusion: wrong comparator, no reference standard
Ref ID: 760

Cox, J. (2017). Use and misuse of the Edinburgh Postnatal Depression Scale (EPDS): a ten point 'survival analysis'. Archives of Women's Mental Health, 20(6), 789-790
Reason for exclusion: wrong publication type
Ref ID: 5427

Di Florio, A., Putnam, K., Altemus, M., Apter, G., Bergink, V., Bilszta, J., et. al. (2017). The impact of education, country, race and ethnicity on the self-report of postpartum depression using the Edinburgh Postnatal Depression Scale. Psychological medicine, 47(5), 787-799
Reason for exclusion: no comparator
Ref ID: 907

Do, P. H., Vo, V. T., Luong-Thanh, B-Y., Nguyen, L. H., Valdebenito, S., Eisner, M., Tran, X. B., Baker, P., Dunne, M. (2021). Abstract 728: Comparative performance of WHO-5, PHQ-9 and PSS Scale for screening antenatal depression and suicide risk. International Journal of Epidemiology, 50(suppl1), i60-i61
Reason for exclusion: wrong publication type
Ref ID: 3610

Drake, E., Howard, E., Kinsey, E. (2014). Online screening and referral for postpartum depression: an exploratory study. Community mental health journal, 50(3), 305-311
Reason for exclusion: included in 2017 guideline
Ref ID: 946

Fellmeth, G., Opondo, C., Henderson, J., Redshaw, M., McNeill, J., Lynn, F., Alderdice, F. (2019). Identifying postnatal depression: Comparison of a self-reported depression item with Edinburgh Postnatal Depression Scale scores at three months postpartum. Journal of Affective Disorders, 251(2019), 8-14
Reason for exclusion: wrong comparator, no reference standard
Ref ID: 1066

Gerbasi, M. E., Eldar-Lissai, A., Acaster, S., Fridman, M., Bonthapally, V., Hodgkins, P., Kanes, S. J., Meltzer-Brody, S. (2020). Associations between commonly used patient-reported outcome tools in postpartum depression clinical practice and the Hamilton Rating Scale for Depression. Archives of Women's Mental Health, 23(5), 727-735
Reason for exclusion: wrong comparator, no reference standard
Ref ID: 1167

Gigantesco, A., Palumbo, G., Cena, L., Camoni, L., Trainini, A., Stefana, A., Mirabella, F. (2021). The limited screening accuracy of the Patient Health Questionnaire-2 in detecting depression among perinatal women in Italy. PLoS ONE, 16(11 November)
Reason for exclusion: wrong comparator, no reference standard
Ref ID: 1179

Hall, H. G., Cant, R., Munk, N., Carr, B., Tremayne, A., Weller, C., Fogarty, S., Lauche, R. (2020). The effectiveness of massage for reducing pregnant women's anxiety and depression; systematic review and meta-analysis. Midwifery, 90(2020), 102818
Reason for exclusion: wrong intervention purpose
Ref ID: 144

Haßdenteufel, K., Lingenfelder, K., Schwarze, C. E., Feisst, M., Brusniak, K., Matthies, L. M., Goetz, M., Wallwiener, M., Wallwiener, S. (2021). Evaluation of Repeated Web-Based Screening for Predicting Postpartum Depression: Prospective Cohort Study. JMIR Mental Health, 8(12): e26665
Reason for exclusion: wrong comparator, no reference standard
Ref ID: 1322

Kohlhoff, J., Hickinbotham, R., Knox, C., Roach, V., Barnett Am, B. (2016). Antenatal psychosocial assessment and depression screening in a private hospital. Australian and New Zealand Journal of Obstetrics and Gynaecology, 56(2), 173-178
Reason for exclusion: included in 2017 guideline
Ref ID: 1624

Kohlhoff, J., Tooke, S., Cibralic, S., Hickinbotham, R., Knox, C., Roach, V., Barnett, B. (2021). Antenatal psychosocial assessment and depression screening in an Australian Private Hospital setting: A qualitative examination of women's perspectives. Midwifery, 103(2021), 103129
Reason for exclusion: wrong outcome
Ref ID: 3484

Kozinszky, Z., Dudas, R. B. (2015). Validation studies of the Edinburgh Postnatal Depression Scale for the antenatal period. Journal of Affective Disorders, 176(2015), 95-105
Reason for exclusion: included in 2017 guideline
Ref ID: 28

Kozinszky, Z., Dudas, R. B. (2015). Validation studies of the Edinburgh Postnatal Depression Scale for the antenatal period. Journal of Affective Disorders, 176(2015), 95-105
Reason for exclusion: included in 2017 guideline
Ref ID: 6558

Levis, B., McMillan, D., Sun, Y., He, C., Rice, D. B., Krishnan, A., et. al. (2019). Comparison of major depression diagnostic classification probability using the SCID, CIDI, and MINI diagnostic interviews among women in pregnancy or postpartum: An individual participant data meta-analysis. International journal of methods in psychiatric research, 28(4), e1803
Reason for exclusion: wrong intervention purpose
Ref ID: 32

Lyubenova, A., Neupane, D., Levis, B., Wu, Y., Sun, Y., He, C., et. al. (2021). Depression prevalence based on the Edinburgh Postnatal Depression Scale compared to Structured Clinical Interview for DSM Disorders classification: Systematic review and individual participant data meta-analysis. International journal of methods in psychiatric research, 30(1), e1860
Reason for exclusion: wrong intervention purpose
Ref ID: 37

Matijasevich, A., Munhoz, T. N., Tavares, B. F., Barbosa, A. P. P. N., da Silva, D. M., Abitante, M. S., Dall'Agnol, T. A., Santos, I. S. (2014). Validation of the Edinburgh postnatal depression scale (EPDS) for screening of major depressive episode among adults from the general population. BMC Psychiatry, 14: 428
Reason for exclusion: wrong population
Ref ID: 1911

Matthey, S., Bilbao, F. (2018). A comparison of the PHQ-2 and MGMQ for screening for emotional health difficulties during pregnancy. Journal of Affective Disorders, 234(2018), 174-179
Reason for exclusion: wrong outcome
Ref ID: 1923

Matthey, S., Della Vedova, A. M. (2018). A comparison of two measures to screen for emotional health difficulties during pregnancy. Journal of Reproductive and Infant Psychology, 36(5), 463-475
Reason for exclusion: wrong comparator, no reference standard
Ref ID: 1924

Matthey, S., Robinson, J., Della Vedova, A. M. (2021). Women's interpretation, understanding and attribution of the anhedonia question in the PHQ-4 and modified-Whooley questions in the antenatal period. Journal of Reproductive and Infant Psychology, (2021), 1-6
Reason for exclusion: wrong outcome
Ref ID: 1925

Matthey, S., Souter, K., Valenti, B., Ross-Hamid, C. (2019). Validation of the MGMQ in screening for emotional difficulties in women during pregnancy. Journal of Affective Disorders, 256 (2019), 156-163
Reason for exclusion: wrong intervention type
Ref ID: 1927

Meijer, J. L., Beijers, C., Van Pampus, M. G., Verbeek, T., Stolk, R. P., Milgrom, J., Bockting, C. L. H., Burger, H. (2014). Predictive accuracy of Edinburgh postnatal depression scale assessment during pregnancy for the risk of developing postpartum depressive symptoms: A prospective cohort study. BJOG: An International Journal of Obstetrics and Gynaecology, 121(13), 1604-1610
Reason for exclusion: no comparator
Ref ID: 1956

O'Connor, E., Rossom, R. C., Henninger, M., Groom, H. C., Burda, B. U. (2016). Primary care screening for and treatment of depression in pregnant and postpartum women evidence report and systematic review for the US preventive services task force. JAMA - Journal of the American Medical Association, 315(4), 388-406
Reason for exclusion: included in 2017 guideline
Ref ID: 50

Olin, S-C. S., Kerker, B., Stein, R. E. K., Weiss, Dara., Whitmyre, E. D., Hoagwood, K., Horwitz, S. M. (2016). Can Postpartum Depression Be Managed in Pediatric Primary Care? Journal of Women's Health (15409996), 25(4), 381-390
Reason for exclusion: wrong outcome
Ref ID: 247

Owora, A. H., Carabin, H., Reese, J., Garwe, T. (2016). Diagnostic performance of major depression disorder case-finding instruments used among mothers of young children in the United States: A systematic review. Journal of Affective Disorders, 201(2016), 185-193
Reason for exclusion: duplicate of Ref ID 52
Ref ID: 6522

Owora, A. H., Carabin, H., Reese, J., Garwe, T. (2016). Diagnostic performance of major depression disorder case-finding instruments used among mothers of young children in the United States: A systematic review. Journal of Affective Disorders, 201(2016), 185-193
Reason for exclusion: duplicate of Ref ID 52
Ref ID: 245

Price, D. A. M., Middleton, M. M., Matthey, A. A. P. S., Goldfeld, P. S., Kemp, P. L., Orsini, M. F. (2021). A comparison of two measures to screen for mental health symptoms in pregnancy and early postpartum: the Matthey Generic Mood Questionnaire and the Depression, Anxiety, Stress Scales short-form. Journal of Affective Disorders, 281(2021), 824-833
Reason for exclusion: wrong comparator, no reference standard
Ref ID: 2315

Reilly, N., Yin, C., Monterosso, L., Bradshaw, S., Neale, K., Harrison, B., Austin, M-P. (2015). Identifying psychosocial risk among mothers in an Australian private maternity setting: A pilot study. Australian & New Zealand Journal of Obstetrics & Gynaecology, 55(5), 453-458
Reason for exclusion: included in 2017 guideline
Ref ID: 5956

Rodríguez-Muñoz, M. F., Ruiz-Segovia, N., Soto-Balbuena, C., Le, H. N., Olivares-Crespo, M. E., Izquierdo-Méndez, N. (2020). The psychometric properties of the patient health questionnaire-4 for pregnant

women International. Journal of Environmental Research and Public Health, 17(20), 1-10
Reason for exclusion: wrong comparator, no reference standard
Ref ID: 2393

Ruyak, S. L., Qeadan, F. (2018). Use of the Antenatal Risk Questionnaire to Assess Psychosocial Risk Factors Associated with Risk for Postpartum Depression: A Pilot Study. Journal of Midwifery and Women's Health, 63(5), 578-583
Reason for exclusion: wrong outcome
Ref ID: 2422

Sambrook Smith, M., Cairns, L., Pullen, L., Opondo, C., Fellmeth, Gracia, Alderdice, F. (2022). Validated tools to identify common mental disorders in the perinatal period: A systematic review of systematic reviews. Journal of Affective Disorders, 298(Part A), 634-643
Reason for exclusion: duplicate of Ref ID 57
Ref ID: 6301

San Martin Porter, M. A., Betts, K., Kisely, S., Pecoraro, G., Alati, R. (2019). Screening for perinatal depression and predictors of underscreening: Findings of the born in Queensland study. Medical Journal of Australia, 210(1), 32-37
Reason for exclusion: wrong outcome
Ref ID: 2456

Shibata, Y., Suzuki, S. (2020). Comparison of the Edinburgh Postnatal Depression Scale and the Whooley questions in screening for postpartum depression in Japan. Journal of Maternal-Fetal and Neonatal Medicine, 33(16), 2785-2788
Reason for exclusion: wrong comparator, no reference standard
Ref ID: 2547

Shrestha, S. D., Pradhan, R., Tran, T. D., Gualano, R. C., Fisher, J. R. W. (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, (2016) 16:72
Reason for exclusion: included in 2017 guideline
Ref ID: 59

Simpson, W., Glazer, M., Michalski, N., Steiner, M., Frey, B. N. (2014). Comparative efficacy of the generalized anxiety disorder 7-item scale and the Edinburgh Postnatal Depression Scale as screening tools for generalized anxiety disorder in pregnancy and the postpartum period. Canadian Journal of Psychiatry, 59(8), 434-440
Reason for exclusion: included in 2017 guideline
Ref ID: 2588

Thombs, B. D., Levis, B., Lyubenova, A., Neupane, D., Negeri, Z., Wu, Y., et. al. (2020). Overestimation of Postpartum Depression Prevalence Based on a 5-item Version of the EPDS: Systematic Review and Individual Participant Data Meta-analysis. Canadian Journal of Psychiatry, 65(12), 835-844
Reason for exclusion: wrong outcome (looking at prevalence rather than screening/diagnosis)
Ref ID: 66

Toreki, A., Andó, B., Dudas, R. B., Dweik, D., Janka, Z., Kozinszky, Z., Keresztúri, A. (2014). Validation of the Edinburgh Postnatal Depression Scale as a screening tool for postpartum depression in a clinical sample in Hungary. Midwifery, 30(8), 911-918
Reason for exclusion: included in 2017 guideline
Ref ID: 2788

Van Der Zee-Van den Berg, A. I., Boere-Boonekamp, M. M., Groothuis-Oudshoorn, C. G. M., Reijneveld, S. A. (2019). The Edinburgh Postpartum Depression Scale: Stable structure but subscale of limited value to detect anxiety. PLoS ONE, 14(9): e0221894

Reason for exclusion: wrong comparator, no reference standard
Ref ID: 2853

van der Zee-van den Berg, A. I., Boere-Boonekamp, M. M., Ijzerman, M. J., Haasnoot-Smallegange, R. M., Reijneveld, S. A. (2017). Screening for Postpartum Depression in Well-Baby Care Settings: A Systematic Review. Maternal and child health journal, 21(1), 9-20
Reason for exclusion: wrong intervention purpose
Ref ID: 69

Venkatesh, K. K., Kaimal, A. J., Castro, V. M., Perlis, R. H. (2017). Improving discrimination in antepartum depression screening using the Edinburgh Postnatal Depression Scale. Journal of Affective Disorders, 214(2017), 1-7
Reason for exclusion: no comparator
Ref ID: 2880

Vlenterie, R., van Ras, H. W. P., Roeleveld, N., Pop-Purceleanu, M., van Gelder, M. M. H. J. (2017). Epidemiological evaluation of the Patient Health Questionnaire-2 in a pregnant population. Journal of Psychosomatic Research, 101(2017), 96-103
Reason for exclusion: wrong comparator, no reference standard
Ref ID: 2906

Vogeli, J. M., Hooker, S. A., Everhart, K. D., Kaplan, P. S. (2018). Psychometric properties of the postpartum depression screening scale beyond the postpartum period. Research in Nursing & Health, 41(2), 185-194
Reason for exclusion: wrong intervention
Ref ID: 4666

Wan-Jung, H., Sbrilli, M. D., Huang, W. D., Tuyet-Mai, H., Meline, B., Laurent, H. K., TabbSee, K. M. (2021). Patients' Perceptions Of Perinatal Depression Screening: A Qualitative Study. Health Affairs, 40(10), 1612-1617
Reason for exclusion: wrong outcome
Ref ID: 3268

Wilkinson, A., Anderson, S., Wheeler, S. B. (2017). Screening for and Treating Postpartum Depression and Psychosis: A Cost-Effectiveness Analysis. Maternal and child health journal, 21(4), 903-914
Reason for exclusion: wrong outcome
Ref ID: 2971

Zee-van den Berg, A., Boere-Boonekamp, M., Ijzerman, M., Haasnoot-Smallegange, R., Reijneveld, S. (2017). Screening for Postpartum Depression in Well-Baby Care Settings: A Systematic Review. Maternal & Child Health Journal, 21(1), 9-20
Reason for exclusion: wrong population
Ref ID: 214

Zhong, Q., Gelaye, B., Rondon, M.,E. Sánchez, S.,J. García, P.,Sánchez, E.,V. Barrios, Y.,E. Simon, G.,C. Henderson, D., May Cripe, S.,A. Williams, M. (2014). Comparative performance of Patient Health Questionnaire-9 and Edinburgh Postnatal Depression Scale for screening antepartum depression. Journal of Affective Disorders, 162(2014), 1-7
Reason for exclusion: wrong comparator, no reference standard
Ref ID: 3133

2.2 Studies not taken through evidence appraisal process, with reason

Austin, M. P. V., Mule, V., Hadzi-Pavlovic, D., Reilly, N. (2022). Screening for anxiety disorders in third trimester pregnancy: a comparison of four brief measures. Archives of Women's Mental Health, 25, 389-397
Reason for exclusion: reference standard not consistent with the PICO *(see **Appendix 5** for study summary)*
Ref ID: 455

Chorwe-Sungani, G.,Chipps, J. (2017). A systematic review of screening instruments for depression for use in antenatal services in low resource settings. BMC Psychiatry, 17(1)
Reason for exclusion: Setting not applicable to Australia.
Ref ID: 8

Darwin, Z.,McGowan, L.,Edozien, L. C. (2015). Antenatal mental health referrals: review of local clinical practice and pregnant women's experiences in England. Midwifery, 31(3), e17-e22
Reason for exclusion: Only relevant outcome was referrals but data are not comparative.
Ref ID: 850

Darwin, Z.,McGowan, L.,Edozien, L. C. (2016). Identification of women at risk of depression in pregnancy: using women's accounts to understand the poor specificity of the Whooley and Arroll case finding questions in clinical practice. Archives of Women's Mental Health, 19(1), 41-49
Reason for exclusion: Whooley and Arroll questions are not used in the Australian setting, and data were collected in this study over a decade ago (2010-2011).
Ref ID: 851

Edward, K. L.,Giandinoto, J. A.,Stephenson, J.,Mills, C.,McFarland, J.,Castle, D. J. (2019). Self-screening using the Edinburgh post natal depression scale for mothers and fathers to initiate early help seeking behaviours. Archives of psychiatric nursing, 33(4), 421-427
Reason for exclusion: RCT compared EPDS self-screening to EPDS self-screening with the addition of a referral pathway pamphlet (i.e. examined effectiveness of providing the referral pathway pamphlet). Comparison and outcomes not relevant to PICO.
Ref ID: 975

Fellmeth, G.,Harrison, S.,Opondo, C.,Nair, M.,Kurinczuk, J. J.,Alderdice, F. (2021). Validated screening tools to identify common mental disorders in perinatal and postpartum women in India: a systematic review and meta-analysis. BMC Psychiatry, 21(1)
Reason for exclusion: Setting not applicable to Australia.
Ref ID: 18

Gollan, J. K.,Wisniewski, S. R.,Luther, J. F.,Eng, H. F.,Dills, J. L.,Sit, D.,Ciolino, J. D.,Wisner, K. L. (2017). Generating an efficient version of the Edinburgh Postnatal Depression Scale in an urban obstetrical population. Journal of Affective Disorders, 208, 615-620
Reason for exclusion: Validation of a shorter version of the EPDS (7-item EPDS) focused solely on depression (not anxiety or anhedonia); the need for a shorter tool is not of particular relevance to the Australian setting where screening is administered online.
Ref ID: 1199

Harel, D.,Levis, B.,Ishihara, M.,Levis, A. W.,Vigod, S. N.,Howard, et. al. (2021). Shortening the Edinburgh postnatal depression scale using optimal test assembly methods: Development of the EPDS-Dep-5. Acta Psychiatrica Scandinavica, 143(4), 348-362
Reason for exclusion: Validation of a shorter version of the EPDS (5-item EPDS) focused solely on depression (not anxiety or anhedonia); the need for a shorter tool is not of particular relevance to the Australian setting where screening is administered online.
Ref ID: 1302

Howard, L. M.,Ryan, E. G.,Trevillion, K.,Anderson, F.,Bick, D.,Bye, A.,Byford, S.,O'Connor, S.,Sands, P.,Demilew, J.,Milgrom, J.,Pickles, A. (2018). Accuracy of the Whooley questions and the Edinburgh Postnatal Depression Scale in identifying depression and other mental disorders in early pregnancy. British Journal of Psychiatry, 212(1), 50-56
Reason for exclusion: Enrolled a sample of Whooley negative participants; assessed performance of EPDS in Whooley negative and Whooley positive participants.
Ref ID: 1384

Lind, A.,Richter, S.,Craft, C.,Shapiro, A. C. (2017). Implementation of Routine Postpartum Depression Screening and Care Initiation Across a Multispecialty Health Care Organization: An 18-Month Retrospective Analysis. Maternal and child health journal, 21(6), 1234-1239

Reason for exclusion: Only relevant outcome was effectiveness but data are not comparative.
Ref ID: 1774

Marcano-Belisario, J. S.,Gupta, A. K.,O'Donoghue, J.,Ramchandani, P.,Morrison, C.,Car, J. (2017). Implementation of depression screening in antenatal clinics through tablet computers: results of a feasibility study. BMC medical informatics and decision making, 17(1), 59
Reason for exclusion: Comparison of screening using different tablet layouts is not relevant to the Australian setting where screening **has** progressed to mobile-friendly formats.
Ref ID: 1877

McGlone, C.,Hollins Martin, C. J.,Furber, C. (2016). Midwives' experiences of asking the Whooley questions to assess current mental health: a qualitative interpretive study. Journal of Reproductive and Infant Psychology, 34(4), 383-393
Reason for exclusion: <10 participants (8 midwives) from one maternity hospital in northwest England.
Ref ID: 1944

Nelson, H. D.,Cantor, A.,Pappas, M.,Weeks, C. (2020). Screening for anxiety in adolescent and adult women a systematic review for the women's preventive services initiative. Annals of Internal Medicine, 173(1), 29-41
Reason for exclusion: Focus of SR is adolescent and adult women rather than pregnant or postpartum women.
Ref ID: 47

Neupane, D.,Levis, B.,Bhandari, P. M.,Thombs, B. D.,Benedetti, A. (2021). Selective cutoff reporting in studies of the accuracy of the Patient Health Questionnaire-9 and Edinburgh Postnatal Depression Scale: Comparison of results based on published cutoffs versus all cutoffs using individual participant data meta-analysis. International journal of methods in psychiatric research, 30(3), e1873
Reason for exclusion: Search dates pre-dated 2017 Guideline searches.
Ref ID: 48

Owora, A. H.,Hélène, Carabin,Reese, J.,Garwe, T. (2016). Diagnostic performance of major depression disorder case-finding instruments used among mothers of young children in the United States: A systematic review. Journal of Affective Disorders, 201, 185-193
Reason for exclusion: Not restricted to within 1 year postpartum and search date may have pre-dated 2017 Guideline searches.
Ref ID: 52

Owora, Arthur H.,Carabin, Hélène,Reese, Jessica,Garwe, Tabitha (2016). Summary diagnostic validity of commonly used maternal major depression disorder case finding instruments in the United States: A meta-analysis. Journal of Affective Disorders, 205, 335-343
Reason for exclusion: Not restricted to within 1 year postpartum and search date may have pre-dated 2017 Guideline searches.
Ref ID: 246

Reilly, N., Hadzi-Pavlovic, D., Loxton, D., Black, E., Mule, V., Austin, M. P. (2022). Supporting routine psychosocial assessment in the perinatal period: The concurrent and predictive validity of the Antenatal Risk Questionnaire-Revised. Women and Birth, 35, e118-e124
Reason for exclusion: reference standard not consistent with the PICO *(see **Appendix 5** for study summary)*
Ref ID: 2368

Reinstein, Sarah,Lieb, Kate,Bernstein, Peter S.,Karkowsky, Chavi Eve (2020). The best perinatal depression screening: Is self-administered PHQ2 more feasible than a nurse-administered one? Perspectives in Psychiatric Care, 56(1), 81-85
Reason for exclusion: Effectiveness of administration methods (self-administered vs nurse-administered) not relevant to the Australian setting where screening is self-administered online.
Ref ID: 3846

Sambrook Smith, M.,Cairns, L.,Pullen, L. S. W.,Opondo, C.,Fellmeth, G.,Alderdice, F. (2022). Validated tools to identify common mental disorders in the perinatal period: A systematic review of systematic reviews. Journal of Affective Disorders, 298, 634-643
Reason for exclusion: Umbrella review with potential for overlap with primary studies, and reference standard not restricted to formal psychiatric diagnosis.
Ref ID: 57

Smith, E. K.,Gopalan, P.,Glance, J. B.,Azzam, P. N. (2016). Postpartum depression screening: A review for psychiatrists. Harvard Review of Psychiatry, 24(3), 173-187
Reason for exclusion: Search date may have pre-dated 2017 Guideline searches (search date not reported in methods).
Ref ID: 61

Venkatesh, K. K.,Nadel, H.,Blewett, D.,Freeman, M. P.,Kaimal, A. J.,Riley, L. E. (2016). Implementation of universal screening for depression during pregnancy: feasibility and impact on obstetric care. American Journal of Obstetrics and Gynecology, 215(4), 517.e1-517.e8
Reason for exclusion: Effectiveness outcomes not comparative, no other relevant outcomes.
Ref ID: 2881

Appendix 3 Included studies list

The updated literature search covered the period from the literature search dates for the 2017 Australian Guideline to **07 March 2022**.

Note: This list includes a total of **21** unique studies that met the eligibility criteria and were taken through the evidence appraisal process. It does not include studies that were included in the Technical Report Part B for the 2017 Australian Guideline.

3.1 Psychosocial assessment

Kalra, H., Reilly, N., Austin, M. P. (2018). An evaluation of routine antenatal depression screening and psychosocial assessment in a regional private maternity setting in Australia. Australian and New Zealand. Journal of Obstetrics and Gynaecology, 58(6), 629-635
Ref ID: 1521

Kingston, D., Austin, M. P., Veldhuyzen van Zanten, S., Harvalik, P., Giallo, R., McDonald, S. D., MacQueen, G., Vermeyden, L., Lasiuk, G., Sword, W., Biringer, A. (2017). Pregnant Women's Views on the Feasibility and Acceptability of Web-Based Mental Health E-Screening Versus Paper-Based Screening: A Randomized Controlled Trial. Journal of medical Internet research, 19(4), e88
Ref ID: 1603

Marley, J. V., Kotz, J., Engelke, C., Williams, M., Stephen, D., Coutinho, S., Trust, S. K. (2017). Validity and acceptability of Kimberley Mum's mood scale to screen for perinatal anxiety and depression in remote aboriginal health care settings. PLoS ONE, 12(1)
Ref ID: 1885

3.2 Screening for depression

Avalos, L. A., Raine-Bennett, T., Chen, H., Adams, A. S., Flanagan, T. (2016). Improved perinatal depression screening, treatment, and outcomes with a universal obstetric program. Obstetrics and Gynecology, 127(5), 917-925
Ref ID: 458

BenDavid, Dina N. (2016). Uncovering the Golden Veil: Applying the Evidence for Telephone Screening to Detect Early Postpartum Depression. Journal of Perinatal Education, 25(1), 37-45
Ref ID: 5840

Blackmore, R., Gibson-Helm, M., Melvin, G., Boyle, J. A., Fazel, M., Gray, K. M. (2022). Validation of a Dari translation of the Edinburgh Postnatal Depression Scale among women of refugee background at a public antenatal clinic. Australian and New Zealand Journal of Psychiatry, 56(5), 525-534
Ref ID: 595

Chan, A. W., Reid, C., Skeffington, P., Gorman, E., Marriott, R. (2022). Experiences of using the Edinburgh Postnatal Depression Scale in the context of antenatal care for Aboriginal mothers: Women and midwives' perspectives. Women and Birth, 35(4), 367-377
Ref ID: 714

Chan, A. W., Reid, C., Skeffington, P., Marriott, R. (2021). A systematic review of EPDS cultural suitability with Indigenous mothers: a global perspective. Archives of Women's Mental Health, 24(3), 353-365
Ref ID: 6

Ezirim, N., Younes, L. K., Barrett, J. H., Kauffman, R. P., Macleay, K. J., Newton, S. T., Tullar, P. (2021). Reproducibility of the Edinburgh Postnatal Depression Scale during the Postpartum Period. American Journal of Perinatology
Ref ID: 1023

Kotz, J., Marriott, R., Reid, C. (2021). The EPDS and Australian Indigenous women: A systematic review of the literature. Women and Birth, 34(2), e128-e134
Ref ID: 27

Lawson, A., Dalfen, A., Murphy, K. E., Milligan, N., Lancee, W. (2019). Use of text messaging for postpartum depression screening and information provision. Psychiatric Services, 70(5), 389-395
Ref ID: 1710

Levis, B., Negeri, Z., Sun, Y., Benedetti, A., Thombs, B. D. (2020). Accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for screening to detect major depression among pregnant and postpartum women: Systematic review and meta-analysis of individual participant data. The BMJ, 371
Ref ID: 33

Logsdon, M. C., Vogt, K., Davis, D. W., Myers, J., Hogan, F., Eckert, D., Masterson, K. (2018). Screening for Postpartum Depression by Hospital-Based Perinatal Nurses MCN. The American journal of maternal child nursing, 43(6), 324-329
Ref ID: 1793

Nithianandan, N., Gibson-Helm, M., McBride, J., Binny, A., Gray, K. M., East, C., Boyle, J. A. (2016). Factors affecting implementation of perinatal mental health screening in women of refugee background. Implementation science: IS, 11(1), 150
Ref ID: 2141

Vik, K., Aass, I. M., Willumsen, A. B., Hafting, M. (2021). Experiences with the routine use of the Edinburgh Postnatal Depression Scale from health visitors' and midwives' perspectives - An exploratory qualitative study. Midwifery, 100, 103017
Ref ID: 2895

Wang., L., Kroenke, K., Stump, T. E., Monahan, P. O. (2021). Screening for perinatal depression with the Patient Health Questionnaire depression scale (PHQ-9): A systematic review and meta-analysis. General Hospital Psychiatry, 68, 74-82
Ref ID: 72

Yapp, E., Howard, L. M., Kadicheeni, M., Telesia, L. A., Milgrom, J., Trevillion, K. (2019). A qualitative study of women's views on the acceptability of being asked about mental health problems at antenatal booking appointments. Midwifery, 74, 126-133
Ref ID: 4221

3.3 Screening for anxiety

Blackmore, R., Gibson-Helm, M., Melvin, G., Boyle, J. A., Fazel, M., Gray, K. M. (2022). Validation of a Dari translation of the Edinburgh Postnatal Depression Scale among women of refugee background at a public antenatal clinic. Australian and New Zealand Journal of Psychiatry, 56(5), 525-534
Ref ID: 595

Chan, A. W., Reid, C., Skeffington, P., Gorman, E., Marriott, R. (2022). Experiences of using the Edinburgh Postnatal Depression Scale in the context of antenatal care for Aboriginal mothers: Women and midwives' perspectives. Women and Birth, 35(4), 367-377
Ref ID: 714

Fairbrother, N., Corbyn, B., Thordarson, D. S., Ma, A., Surm, D. (2019). Screening for perinatal anxiety disorders: Room to grow. Journal of Affective Disorders, 250, 363-370
Ref ID: 1030

Lieb, K., Reinstein, S., Xie, X., Bernstein, P. S., Karkowsky, C. E. (2020). Adding perinatal anxiety screening to depression screening: is it worth it? American Journal of Obstetrics and Gynecology MFM, 2(2)
Ref ID: 1764

Nithianandan, N., Gibson-Helm, M., McBride, J., Binny, A., Gray, K. M., East, C., Boyle, J. A. (2016). Factors affecting implementation of perinatal mental health screening in women of refugee background.

Implementation science: IS, 11(1), 150
Ref ID: 2141

3.4 Economic analyses

Chambers, G.M., Botha, W., Reilly, N., Black, E., Kingston, D., Austin, M-P. (2022). The clinical performance and cost-effectiveness of two psychosocial assessment models in maternity care: The Perinatal Integrated Psychosocial Assessment study. Women and Birth, 35(2), e133-e141

*(Nb. see **Appendix 5.1.2** for a summary of this economic analysis)*

Appendix 4 Characteristics of included studies

4.1 Psychosocial assessment

Table App. 8 lists the 'new' evidence for psychosocial assessment in reverse chronological order. The table summarises the characteristics of studies that were ultimately included in the results section of this report, and studies that were not taken through the evidence appraisal process. The reasons for not progressing to full evidence appraisal are documented in **Appendix 2.2**.

Table App. 8 New evidence identified in the literature search update: Psychosocial assessment

Ref ID	Author & year	Study type, timeframe and location	Aim	Population	Intervention	Relevant outcomes
Included						
2368	Reilly 2022	Prospective study with self-reported reference standard (SAGE-SR), conducted as part of a larger project comparing two models of integrated psychosocial care during pregnancy Recruitment Mar 2017 to May 2019 Australia	To examine the psychometric performance of the ANRQ-R by determining its concurrent and predictive validity when used across the perinatal period	Pregnant women who had completed a psychosocial assessment as part of routine antenatal care N=1565	Intervention: ANRQ-R, in 2nd and 3rd trimesters, and 3 months following birth Reference standard: Series of Assessments for Guiding Evaluation - Self-Report (SAGE-SR) depression and anxiety information	Tool performance: concurrent validity, predictive validity, sensitivity, specificity
1521	Kalra 2018	Mixed-methods study comprising retrospective medical records audit (1 Aug 2015 to 31 Jul 2016) and prospective online survey of women's experiences (10 Jul 2016 to 3 Oct 2016) Ballarat, Australia	To explore women's experience of receiving depression screening and psychosocial risk assessment as part of routine antenatal care	Women attending a regional private maternity hospital who received antenatal psychosocial assessment and depression screening within an integrated psychosocial model of care N=455 in medical record review N=101/109 completed feedback survey	ANRQ and EPDS administered at a woman's first antenatal booking visit, with localised protocols for referral for additional assessment or support as required	Clinical usefulness: acceptability to pregnant women
1603	Kingston 2017	RCT Recruitment Aug 2013 to Jan 2015 Alberta, Canada	To evaluate the feasibility and acceptability of web-based mental health e-screening compared with paper-based screening among pregnant women	Pregnant women recruited from community-based physician-led maternity clinics, high-risk antenatal unit in a tertiary care centre, or community hospital-based prenatal classes N=636 (305 intervention, 331 control)	Intervention: ALPHA and EPDS, web-based e-screening on a tablet (unassisted) Control: ALPHA and EPDS, paper-based screening, completed independently	Clinical usefulness: feasibility and acceptability to pregnant women (quantitative analysis)

Ref ID	Author & year	Study type, timeframe and location	Aim	Population	Intervention	Relevant outcomes
1885	Marley 2017	Cross-sectional study Data collection May 2013 to June 2014 Community setting Kimberley, Australia	To determine if the Kimberley Mum's Mood Scale (KMMS) is a reliable, valid and acceptable tool for identifying Kimberley Aboriginal women at risk of perinatal anxiety or depressive disorders	At 15 Kimberley sites, Aboriginal women aged 16 years and older with ongoing pregnancy or birth within the previous 12 months N=97	Intervention: KMMS, comprising Part 1 covering same areas and scoring as EPDS, that is moderated by Part 2 psychosocial tool based on SAFESTART Guidelines and provision of a mental health brief intervention (support, action plans, referral etc) Reference standard: GP assessment using DSM-IV and severity based on Australian GP Mental State Examination	Tool performance: validity, reliability, sensitivity, specificity, accuracy (correctly classified) Clinical usefulness: acceptability to perinatal women and study personnel

Abbreviations: ALPHA, Antenatal Psychosocial Health Assessment; ANRQ, Antenatal Risk Questionnaire; ANRQ-R, Antenatal Risk Questionnaire -Revised; EPDS, Edinburgh Postnatal Depression Scale; GP, General Practitioner; KMMS, Kimberley Mum's Mood Scale; RCT, randomised controlled trial ; SAGE-SR, Series of Assessments for Guiding Evaluation - Self-Report.

4.2 Screening for depression and/or anxiety

Table App. 9 lists the 'new' evidence for depression and anxiety screening in reverse chronological order. The table summarises the characteristics of studies that were ultimately included in the results section of this report, and studies that were not taken through the evidence appraisal process. The reasons for not progressing to full evidence appraisal are documented in **Appendix 2.2**.

Table App. 9 New evidence identified in the literature search update

Ref ID	Author & year	Study type, timeframe and location	Aim	Population	Intervention	Relevant outcomes
Included						
455	Austin 2022	Cross-sectional study Recruitment Mar 2017 to May 2019 Community setting Sydney, Australia	To compare the test performance of the GAD-2, GAD-7 and EPDS-3A and the two 'anxiety' items of the ANRQ-2A in a large sample of women in the third trimester of pregnancy	Pregnant women (3rd trimester) attending their first antenatal appointment at a tertiary teaching hospital in metropolitan Sydney N=954	Intervention: Direct comparison of GAD-2, GAD-7, EPDS-3A, ANRQ-2A Reference standard: DSM-V via SAGE-SR (online self-report)	Tool performance: sensitivity, specificity, PPV, NPV, LR+, LR-, correctly classified, Youden's index, AUROC
595	Blackmore 2022	Cross-sectional study Recruitment Jul 2016 to Nov 2018 Community setting Melbourne, Australia	To investigate the screening properties of the EPDS with a sample of Dari-speaking women seeking antenatal care	Adult Dari-speaking pregnant women who arrived to Australia from Afghanistan on a humanitarian visa, asylum seeker or spousal visa attending a public antenatal clinic that operates as a designated refugee antenatal clinic one day per week N=52	Intervention: EPDS (iCOPE digital platform) translated to Dari Reference standard: SCID-5-RV (Research Version)	Tool performance: sensitivity, specificity, PPV, NPV, AUROC
714	Chan 2022	Qualitative study *Timeframe not reported* Perth, Australia	To explore cultural validity of EPDS through understanding experiences of Aboriginal women and midwives	Perth-based Aboriginal antenatal women and non-Aboriginal midwives N=13 women and 10 midwives	Intervention: EPDS	Clinical usefulness: acceptability by women and midwives
6	Chan 2021	SR with descriptive analysis Search dates: to Jan 2020 Included studies: 14 studies (11 in Australia, 2 in Canada, 1 in US)	To elucidate how culturally suitable the EPDS is and how variations of EPDS implementation have been used with Indigenous populations across the world	Indigenous pregnant or postpartum women and/or their HCPs	Intervention: EPDS	Tool performance psychometric validity Clinical usefulness acceptability (via tool rejection)
27	Kotz 2021	SR with descriptive analysis Search dates: 1990 to Jan 2019 Included studies: 4 studies	To review the effectiveness, validity, reliability and cultural safety of the EPDS in the Indigenous Australian context	Perinatal Indigenous Australian women	Intervention: EPDS (including modifications in an Indigenous Australian context)	Tool performance construct, standardisation, reliability, validity criteria, accuracy, replicability Clinical usefulness: cultural safety

Ref ID	Author & year	Study type, timeframe and location	Aim	Population	Intervention	Relevant outcomes
72	Wang 2021	SR Search dates: Jan 2001 to Jun 2020 Included studies: 35 studies (10 criterion validity, 25 convergent validity)	To determine the criterion and convergent validity of the PHQ-9 in perinatal depression screening and to compare the performance against the EPDS	Perinatal population	Intervention: PHQ-9 Reference standard: structured psychiatric interview (criterion validity) or another validated depression measure (convergent validity)	Tool performance sensitivity, specificity, AUROC, validity
1023	Ezirim 2021	Longitudinal study Recruitment Oct 2016 to Sep 2018 Community setting Texas, US	To determine degree of consistency in EPDS scores in immediate postpartum period compared with 6-week postpartum visit	Postpartum women who completed screening at 3-24 hours postpartum in an academic hospital in the US and were screened again at an outpatient clinic at 6 weeks postpartum N=848	Intervention: EPDS Reference standard: DSM-IV evaluation if EPDS >9	Tool performance: sensitivity, specificity, LR+, LR-, AUROC
2895	Vik 2021	Exploratory qualitative study (focus group interviews) with thematic analysis Data collection Apr 2018 Community setting Norway	To examine how health visitors and midwives perceive and practice routine screening for PND at a municipal health care centre	Health visitors and midwives who routinely use EPDS for PND screening N=10 health visitors, 2 midwives	Intervention: EPDS	Clinical usefulness: acceptability (thoughts, feelings, expectations, experiences)
33	Levis 2020	SR and MA of individual participant data (the DEPRESsion Screening Data [DEPRESSED] EPDS Collaboration) Search dates: inception to Oct 2018 Included studies: 58 (25 included pregnant women, 30 postpartum women, 3 both)	To use individual participant data MA to evaluate EPDS screening accuracy for detection of major depression in pregnant and postpartum women	Adult women who completed EPDS assessments during pregnancy or within 12 months of giving birth, who were not receiving psychiatric assessment or care and not identified as possible depression	Intervention: EPDS Reference standard: diagnostic classification for current MDD or MDE using DSM or ICD criteria based on a validated semi-structured (SCID, CIS, DIGS) or fully structured interview (CIDI, CIS-R, MINI)	Tool performance: sensitivity, specificity, AUROC
1764	Lieb 2020	Comparative historical control study *Timeframe not reported* Community setting Bronx, New York, US	To assess whether adding perinatal anxiety screening to perinatal depression screening in an urban prenatal care clinic would increase the identification of women who would benefit from mental health support	Pregnant women (24 to <29 weeks' gestation) attending an appointment at an urban prenatal care clinic in the US (low socioeconomic status) N=225	Intervention: PHQ-2 plus GAD-2 Control: PHQ-2 alone (historical control) Any woman referred to a social worker for a positive PHQ-2 or GAD-2 (score of ≥3) underwent more extensive screening using PHQ-7 and GAD-7	Clinical usefulness: effectiveness - change in positive screening results due to adding the anxiety screening; social work referrals and mental healthcare actions
1030	Fairbrother 2019	Cross-sectional study Recruitment Nov 2007 to Nov 2010 Community setting British Columbia, Canada	To assess the accuracy of the most commonly used and/or recommended screening tools for perinatal anxiety disorders	Pregnant women in Canada participating in a study of the prevalence of perinatal anxiety disorders N=310 (N=115 completed reference standard)	Intervention: EPDS, EPDS-3A, GAD-7, GAD-2 at approximately 35 weeks' gestation and again at 3 months' postpartum (postpartum results only) Reference standard: SCID-IV if scored at/above threshold on one or more instruments	Tool performance: sensitivity, specificity, PPV, NPV, LR+, LR-, correctly classified, Youden's index, AUROC

Ref ID	Author & year	Study type, timeframe and location	Aim	Population	Intervention	Relevant outcomes
1710	Lawson 2019	Feasibility study Recruitment Jul 2015 to Jan 2017 Toronto, Canada	To evaluate the feasibility of using text messages to enhance postpartum depression screening and education of women in the immediate postpartum	Adult pregnant women (2nd or 3rd trimester) attending antenatal clinic in Toronto N=937	Intervention: PHQ-2 by text message repeated biweekly until 12 weeks postpartum, plus 3 informational texts per week designed to bolster awareness and self-care behaviours Reference standard: EPDS via telephone in screen positive women	Tool performance: sensitivity, specificity, PPV, NPV, AUROC Clinical usefulness: effectiveness - referrals to perinatal psychiatrist; attended appointment with psychiatrist; psychiatrist diagnosis of mental health problem; acceptability to women
4221	Yapp 2019	Qualitative study (semi-structured interviews) Recruitment Nov 2014 to Jun 2016 Community setting London, UK	To explore women's views on the acceptability of being asked about mental health problems at antenatal booking	Pregnant women who completed an antenatal booking appointment (including Whooley questions) at an inner-city UK service N=52	Intervention: Whooley questions	Clinical usefulness: acceptability, experiences and satisfaction of women
1793	Logsdon 2018	Comparative descriptive study *Timeframe not reported* Louisville, US	To determine whether the screening and education of women prior to hospital discharge were acceptable to new mothers, and if screening for PPD continues for new mothers in the community after hospital discharge	Adult new mothers screened the night before hospital discharge at a mother-baby unit in the US N=101	Intervention: EPDS ≥10 and referral for a home visiting program Control: EPDS <10	Clinical usefulness: acceptability to mothers
458	Avalos 2016	Population-based retrospective cohort study with historical control Universal Perinatal Depression Screening Program Progressive implementation Time period 2007 to 2014 Northern California	To evaluate whether universal prenatal and early postnatal screening for depression leads to increased detection, subsequent intervention, and improved depressive symptom outcomes	Adult pregnant women with at least one obstetric visit during pregnancy (first 20 weeks; 20 weeks' gestation to delivery) and postpartum (3 months) as part of Universal Perinatal Depression Screening Program in US N=97,678	Intervention: PHQ-9 administered as part of Universal Perinatal Depression Screening Program Historical control: pre-implementation	Clinical usefulness: effectiveness - treatment rates for a new depression diagnosis (at least 1 antidepressant or at least 1 individual counselling session or attendance at group class to 6 months postpartum), improvement in depressive symptoms
2141	Nithianandan 2016	Qualitative study (semi-structured interviews and thematic analysis) Recruitment Apr to Jul 2015 Melbourne, Australia	To investigate barriers and enablers to implementing evidence-based, nationally recommended PNMH screening, and inform sustainable implementation of a screening and referral program, in women of refugee background	Health professionals and researchers with specialist knowledge in PNMH, refugee health or both, or who were involved in the Monash Health refugee antenatal clinic N=28 HCPs (midwives, obstetricians, PNMH and refugee experts) & interpreters; N=9 women of refugee background	Intervention: Perinatal mental health screening in women of refugee background	Clinical usefulness: acceptability (and barriers and enablers) to women and HCPs

Ref ID	Author & year	Study type, timeframe and location	Aim	Population	Intervention	Relevant outcomes
5840	BenDavid 2016	Longitudinal study Community setting Data collection Aug to Nov 2013 Pittsburgh, US	To demonstrate how universal PPD screening implemented in primary care could result in increased PPD detection and treatment	Postpartum women (2-3 weeks) who were patients at an inner-city health centre in Southeastern Massachusetts (low-income mothers with limited social support) N=27	Intervention: EPDS delivered via telephone	Clinical usefulness: acceptability to women, effectiveness - acceptance of provider and support referrals
Not taken through evidence appraisal						
57	Sambrook Smith 2022	Umbrella review (SR of SRs) Search dates: to May 2021 Included studies: 30 SRs (76 different screening tools, 33 validated tools)	To provide a summary of psychometric properties of tools for the identification of common mental disorders in the perinatal period	Pregnant or postpartum (up to 12 months) women	Intervention: Screening tool for common mental disorders Reference standard: Diagnostic interview or 'best practice' in that setting (e.g. another screening tool)	Tool performance: sensitivity, specificity
18	Fellmeth 2021	SR Search dates: to Apr 2020 Included studies: 7 studies	To synthesise evidence on the validation of screening tools for common mental disorders in perinatal women in India	Pregnant (any trimester) or postpartum (up to 12 months) women living in India	Intervention: Screening tools for common mental disorders Reference standard: Diagnostic interview or another screening tool	Tool performance: sensitivity, specificity, AUROC
48	Neupane 2021	SR and MA of individual participant data (the DEPRESsion Screening Data [DEPRESSED] EPDS Collaboration) Search dates: inception to Jun 2016 (EPDS) and Jan 2000 to Feb 2015 (PHQ-9) Included studies: 19 EPDS (14 reported multiple cutt-offs; 30 PHQ-9 (23 reported multiple cut-offs)	To use individual participant data MA to indirectly compare accuracy of EPDS and PHQ-9 based on published cutt-offs and all relevant cut-offs from the full dataset	EPDS: Pregnant or postpartum (within 12 months of giving birth) women, not recruited from psychiatric settings or with symptoms of depression PHQ-9: Adults, not recruited from school-based settings or psychiatric settings or with symptoms of depression	Intervention: EPDS Reference standard: diagnostic classification for current MDD or MDE using DSM or ICD criteria based on a validated diagnostic interview	Tool performance: sensitivity, specificity
1302	Harel 2021	Primary study using individual participant data from SR/MA (the DEPRESsion Screening Data [DEPRESSED] EPDS Collaboration) Search dates: inception to Oct 2018 Included studies: 22 primary studies (5157 participants, 765 major depression cases)	To develop a reliable and valid shortened form of the EPDS (EPDS-Dep-5) for depression screening in pregnancy and postpartum using individual participant data MA	Adult pregnant or postpartum (within 1 year) women, from published datasets (multiple countries) N=5157	Intervention: EPDS Reference standard: SCID	Tool performance: sensitivity, specificity
47	Nelson 2020	SR Search dates: 1996 to Nov 2019 Included studies: 9 studies and 1 poor-quality SR (Meades 2011, used as Foundation Review for 2017 Guideline) For US Women's Preventive Services Initiative	To evaluate evidence on the effectiveness of screening for anxiety disorders in primary care in improving symptoms, function and quality of life; harms of screening, accuracy of screening instruments; and effectiveness and harms of treatments	Women and adolescent girls aged 13 years or older (>50% female participants) not currently diagnosed with anxiety disorders, including pregnant and postpartum women	Intervention: Screening tools for anxiety disorders, applicable to primary care settings in US, such as brief report or clinician-administered questionnaires Reference standard: Diagnostic interview or another screening tool	Tool performance: Sensitivity, specificity, LR+, LR-, AUROC, effectiveness (clinical response, reduction in anxiety symptoms, QoL)

Ref ID	Author & year	Study type, timeframe and location	Aim	Population	Intervention	Relevant outcomes
3846	Reinstein 2020	Longitudinal comparative historical control study Community setting Data collection May to Aug 2016 (nurse-administered), Jul to Sep 2016 (self-administered) Bronx, New York, US	To assess PND screening via self-administered PHQ-2 versus nurse-administered PHQ-2	Pregnant women who presented for a prenatal appointment between 24-29 weeks' gestation at a large, urban medical centre in the US N=227 (n=127 nurse-administered, n=100 self-administered)	Intervention: Self-administered PHQ-2, handed out by front desk at appointment checking in Comparator: Nurse-administered (verbal) PHQ-2	Clinical usefulness: effectiveness - referral and utilisation of social work and other mental health services, postpartum depression
975	Edward 2019	RCT, 12 months' follow-up Data collection Apr 2014 to Feb 2015 Melbourne, Australia	To examine the effectiveness of a self-screening tool (EPDS) and referral pathway pamphlet for expectant women and their partners	Expectant mother and father dyad in 3rd trimester in a large private hospital in Melbourne N=140 (70 dyads)	Intervention: EPDS (self-completed) and K10 at baseline (3rd trimester) and postpartum (at 12 months follow-up) Reference standard: EPDS	Clinical usefulness: effectiveness (EPDS and K10 scores at follow-up)
1384	Howard 2018	Cross-sectional study Community setting Recruitment Nov 2014 to Jun 2016 South-east London	To investigate the prevalence of mental disorders in early pregnancy and the diagnostic accuracy of depression screening using Whooley questions compared with EPDS	Pregnant women attending booking appointment at inner-city maternity service in London and asked the Whooley questions N=545 (258 Whooley negative, 287 Whooley positive)	Intervention: Index test was Whooley questions, followed by EPDS Reference standard: SCID-I-Research Version	Tool performance: sensitivity, specificity, PPV, NPV, LR+, LR-, AUROC
8	Chorwe-Sungani 2017	SR Search dates: 2000 to Sep 2015 Included studies: 11 (7 different screening tools)	To appraise the best available evidence on screening instruments suitable for detecting depression in antenatal care in low resource settings	Pregnant women at any stage of pregnancy attending antenatal care in low resource settings, defined as low income and LMIC and some UMIC such as South Africa	Intervention: Screening instrument for depression Reference standard: Formal psychiatric assessment (SCID, MINI, CIDI, ICD-10 or DSM-IV by psychiatrist)	Tool performance: sensitivity, specificity, AUROC
1199	Gollan 2017	Cross-sectional study *Timeframe not reported* Community setting Pennsylvania, US	To identify the specific latent structure of the EPDS in a large-scale community setting, and to establish new threshold scores for a shortened version (7-item one factor [depression] version instead of 10-item two factor [depression + anxiety] version)	Adult postpartum (4-6 weeks) women in US N=15,172 (n=5,055 used for exploratory factor analysis and n=10,117 used for confirmatory factor analysis)	Intervention: EPDS administered via telephone Reference standard: SIGH-ADS (comprised of 21-item HAM-D and additional 8 items to ascertain neurovegetative symptoms) via home visit if EPDS ≥10	Tool performance: sensitivity, specificity, ROC
1774	Lind 2017	Retrospective analysis Recruitment Jan 2013 to Jul 2014 Community setting Midwestern US	To evaluate the implementation of a PND screening and treatment initiation process to identify and treat women at risk for PND in a large multispecialty healthcare organisation	Women presenting for routine postpartum care (well child visits) from birth to 4 months N=23,398 eligible visits	Intervention: EPDS	Clinical usefulness: effectiveness (SSRI prescriptions; mental health department visits)

Ref ID	Author & year	Study type, timeframe and location	Aim	Population	Intervention	Relevant outcomes
1877	Marcano-Belisario 2017	RCT Recruitment Oct 2015 to May 2016 Community setting England, UK	To assess the feasibility of using tablets in the waiting area of antenatal clinics for implementing antenatal depression screening	Adult pregnant women (any gestational age) attending an antenatal clinic appointment, and who had not been diagnosed with depression of GAD in the past 12 months and were not receiving treatment for these disorders N=597	Intervention: Whooley questions and EPDS administered on a tablet in the waiting area using a scrolling layout (single screen allowing vertical scrolling) Comparator: Whooley questions and EPDS administered on a tablet in the waiting area using a paging layout (one question displayed on a screen at any given time)	Clinical usefulness: completion time, feasibility
52	Owora 2016a	SR Search dates: 1994 to 2014 Included studies: 14 articles (21 different instruments)	To identify the most valid maternal MDD case-finding instrument in the US	Mothers of young children (0-5 years) in the US	Intervention: Case-finding instruments for MDD Reference standard: Any	Tool performance: sensitivity, specificity
61	Smith 2016	SR (qualitative) Search dates: Jan 1987 to search date not reported Included studies: 61 articles	To perform a qualitative review on screening for PPD as applicable to the general psychiatrist	Postpartum population	Intervention: Screening instruments for PPD Reference standard: No restrictions specified	Tool performance: sensitivity, specificity, PPV, NPV
246	Owora 2016b	SR Search dates: Jan 1994 to Dec 2015 Included studies: 12 articles (6 different instruments)	To estimate the diagnostic validity of commonly used maternal MDD case-finding instruments in the US	Mothers of young children (0-5 years) in the US	Intervention: Case-finding instruments for MDD Reference standard: Any	Tool performance: sensitivity, specificity
851	Darwin 2016	Mixed-method cohort study Data collection Jun 2010 to Oct 2011 North of England	To provide the first validation of the Whooley and Arroll questions completed at booking in UK practice and explore women's views and experiences of antenatal mental health assessment	Pregnant women attending first formal antenatal appointment at a large, inner-city hospital in north of England N=191 (22 interviewed)	Intervention: Whooley and Arroll questions (self-completed) Reference standard: EPDS	Tool performance: sensitivity, specificity, PPV, NPV Clinical usefulness: acceptability (women's views and experiences)
1944	McGlone 2016	Qualitative interpretive study (semi-structured interviews) *Timeframe not reported* Community setting England, UK	To explore midwives' experiences of asking the Whooley questions at the antenatal booking interview	Midwives who regularly conducted antenatal booking visits N=8	Intervention: Whooley questions asked at antenatal booking visit	Clinical usefulness: acceptability to midwives (thoughts, attitudes towards, and experiences of asking Whooley questions)
2881	Venkatesh 2016	Prospective observational study Study period Jul 2010 to Jun 2014 Community setting Boston, US	To assess the feasibility of large-scale implementation of universal screening for depression at 24-28 weeks' gestation and at 6 weeks' postpartum, followed by EPDS for those who screen positive	Pregnant women screened at 24-28 weeks' gestation and again at 6 weeks' postpartum N=8840 antenatal screening; N=7780 postpartum screening	Intervention: EPDS	Clinical usefulness: effectiveness - referral to mental health services, psychiatric diagnosis, treatment initiation with antidepressant

Ref ID	Author & year	Study type, timeframe and location	Aim	Population	Intervention	Relevant outcomes
850	Darwin 2015	Mixed-method cohort study *Timeframe not reported* North of England	To investigate consistency and completeness of mental health assessment, subsequent management of pregnant women at risk of mental health problems, and women's experiences of referral process	Pregnant women attending first formal antenatal appointment at a large, inner-city hospital in north of England N=191	Intervention: Whooley and Arroll questions (self-completed)	Clinical usefulness: effectiveness measured via referrals

Abbreviations: ALPHA, Antenatal Psychosocial Health Assessment; ANRQ-2A, 2 'anxiety' items from the Antenatal Risk Questionnaire; AUROC, area under the receiver-operating characteristics curve; CIDI, Composite International Diagnostic Interview; CIS, Clinical Interview Schedule; CIS-R, Clinical Interview Schedule - Revised; DIGS, Diagnostic Interview for Genetic Studies; DSM, Diagnostic and Statistical Manual of Mental Disorders; DSM-IV, Diagnostic and Statistical Manual of Mental Disorders, 4th edition; DSM-V, Diagnostic and Statistical Manual of Mental Disorders; 5th Edition; EPDS, Edinburgh Postnatal Depression Scale; EPDS-3A, Edinburgh Postnatal Depression Scale - Anxiety subscale; GAD-2, Generalized Anxiety Disorder 2-item Scale; GAD-7, Generalized Anxiety Disorder 7-item Scale; HAM-D, Hamilton Depression Rating Scale; HCP, health care professional; ICD, International Statistical Classification of Diseases and Related Health Problems; ICD-10, International Statistical Classification of Diseases and Related Health Problems, 10th revision; iCOPE, Centre of Perinatal Excellence digital screening platform; K10, Kessler Psychological Distress Scale (10 item); LMIC, Low Middle Income Countries; LR-, negative likelihood ratio; LR+, positive likelihood ratio; MA, meta-analysis; MDD, major depressive disorder; MDE, major depressive episode; MINI, Mini International Neuropsychiatric Interview; NPV, negative predictive value; PHQ, Patient Health Questionnaire; PHQ-2, first 2 items of the PHQ-9; PHQ-7, Patient Health Questionnaire-7; PHQ-9, Patient Health Questionnaire-9; PND, postnatal depression; PNMH, perinatal mental health; PPD, postpartum depression; PPV, positive predictive value; ROC, receiver operating characteristics; SAGE-SR, Series of Assessments for Guiding Evaluation – Self-Report; SCID, Structured Clinical Interview for DSM Disorders; SCID-I-Research version, Structured Clinical Interview for DSM Axis I Disorders Research Version; SCID-5-RV, Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders 5th edition Research Version; SIGH-ADS, Structured Interview Guide for the Hamilton Depression Rating Scale with Atypical Depression Supplement; SR, systematic review; SSRI, selective serotonin reuptake inhibitor; UK, United Kingdom; UMIC, Upper Middle Income Countries; US, United States.

Appendix 5 Emerging evidence of relevant new tools

5.1 ANRQ-R

Two studies examining the ANRQ-R (Reilly 2022, Chambers 2022) were identified in the Evidence Review Update. Reilly 2022 was excluded by the EWG as the reference standard did not meet the eligibility criteria in the research protocol. Chambers 2022 was identified in the search for Australian economic analyses of psychosocial assessment tools.

The ANRQ-R was identified by the EWG as an emerging tool, and further information from these two studies is therefore included below.

5.1.1 Technical performance

The **Reilly 2022** study formed part of the Perinatal Integrated Psychosocial Assessment (PIPA) Project, underway at the Royal Hospital for Women in Randwick, NSW. The study examined the concurrent and predictive validity of the (then newly revised) ANRQ-R tool in the perinatal period in pregnant women at a large Sydney teaching hospital. Test-retest reliability was reported by the same group in a separate publication (**Reilly 2021**).

The tool was administered alongside a self-reported reference standard (Series of Assessments for Guiding Evaluation – Self-Report [SAGE-SR]) to 1565 women who consented to participate in one or more visits as follows: second trimester (N=1166), third trimester (N=957) and at three months postpartum (N=796). Thus, rather than a series of consecutive visits, the subjects in each group may or may not have completed the tool at the other visits. The majority of women completed both the tool and reference standard in the same day (second trimester 91.7%, third trimester 90.8%; three months postpartum 87.9%). Results of the study are summarised in Table App. 10 and Table App. 11.

In the second trimester, third trimester and three months postpartum, 6.5%, 5.6% and 6.2% of subjects met the reference criteria for any depressive or anxiety disorder (respectively). Women who met the reference case criteria at each time point had higher mean ANRQ-R scores than 'non-case' subjects (second trimester, 21.5±8.0 vs 13.1±6.6; third trimester, 22.7±8.7 vs 13.7±6.9; postpartum, 22.0±7.5 vs 13.7±6.8, respectively).

The ANRQ-R (cut-off ≥18) correctly classified 75.5% of concurrent subjects during the second trimester, 72.2% in the third trimester, and 73.4% at three months postpartum. In terms of predictive use, the ANRQ-R worked best to predict status at three months postpartum when administered in the second trimester (AUC 0.789 [95%CI: 0.681, 0.896]) rather than in the third trimester (AUC 0.705 [95%CI: 0.600, 0.810]). Using a ≥18 cut-off, the sensitivity and specificity of the ANRQ-R were comparable (between ~0.7 – 0.8) when used in the second trimester for identifying current cases or predicting future cases of depression or anxiety meeting the reference standard criteria.

Table App. 10 Study design and psychometric properties: Technical performance of ANRQ-R (Reilly 2022)

Study characteristics (N)	Content validity	Reliability	Applicability
Study design: Prospective study with self-reported reference standard (SAGE-SR) to compare effectiveness of two alternative models of integrated psychosocial care. *Study population:* Pregnant women who completed a psychosocial assessment as part of routine antenatal care in 2nd and 3rd trimesters, and 3mo following birth. N=1565 n = 1166 (second trimester), n = 957 (third trimester) and n = 796 (3-month postpartum).	*Included domains:* • Friends and family supports • Partner support • History of abuse by partner • Major stressors; impact of major stressors • Mother's experience of parenting in childhood • Childhood experience of abuse • Impact of childhood experiences • History of depression; any recent episodes • Impact of depression on psychosocial function • Whether treatment was sought or recommended, and a diagnosis made • Whether medication has been prescribed and if so whether the mother is still taking it • Whether the mother has a problem with drugs or alcohol • Trait anxiety • Obsessional/compulsive traits *Method of development:* • Update to ANRQ to incorporate questions regarding domestic violence and substance abuse. Actual development methodology reported in separate manuscript (Reilly (2021) Journal of Affective Disorders 293. 43-50).	*Test-retest reliability:* Reilly 2021 reported test-retest reliability in a community sample of pregnant women at first antenatal appointment and within 4 weeks ICC for ANRQ-R total score = 0.77 ('good reliability') ICC for individual likert-type items =0.65-0.80 ('moderate to good')	*Country:* Australia *Setting:* Outpatient. Royal Hospital for Women, Randwick NSW (large metropolitan teaching hospital) *Normative data:* Yes; describes sociodemographic factors and psychosocial profile relative to scores

Abbreviations: ANRQ, Antenatal Risk Questionnaire; ANRQ-R, Antenatal Risk Questionnaire – Revised; ICC, intraclass correlation coefficient; KMMS, Kimberley Mum's Mood Scale; SAGE-SR, Series of Assessments for Guiding Evaluation - Self-Report.

Table App. 11 Evidence profile table for ANRQ-R

Evidence base		Performance[3]		
k (N)	**Study ID(s)**	**Predictive accuracy**	**Concurrent validity**	
1 (1565)	Reilly 2022	6.5%, 5.6% and 6.2% of women met SAGE-SR criteria for any depressive or anxiety disorder at 2nd trimester, 3rd trimester and 3 months postpartum, respectively. *Concurrent Use (does mother have depression/anxiety now?)* Correctly classified 72–76% of 'cases' (true positives) and 'non-cases' (true negatives *Predictive Use (will mother have depression/anxiety postpartum?)* Correctly predicted 74–78% of postnatal 'cases' and 'non-cases'	LR+ of a case at 3mo (cut-off ≥18): 2nd trimester: 3.8 3rd trimester: 2.2	
			Concurrent Use: Sensitivity: 0.70–0.74 Specificity: 0.72–0.76 AUROC: 0.789–0.798	*Predictive Use:* Sensitivity: 0.52–0.72 Specificity: 0.75–0.79 AUROC: 0.705–0.789

Abbreviations: ANRQ-R, Antenatal Risk Questionnaire – Revised; AUROC, area under the receiver-operating characteristics curve; LR+, positive likelihood ratio; SAGE-SR, Series of Assessment for Guiding Evaluation – Self-Report.

5.1.2 Economic analysis

Chambers 2022 compared two models of psychosocial assessment in a primary care maternity setting: the PIPA model, consisting of the ANRQ-R and the EPDS, versus the usual care model (SAFE START), comprising SAFE START psychosocial questions and the EPDS. The PIPA model measures cumulative risk, auto-scores both the ANRQ-R and EPDS (to eliminate manual scoring errors) and allows specific clinician concerns to be documented. In addition, PIPA generates six psychosocial risk levels (no risk; no risk on ANRQ-R; low risk; medium risk; medium-high risk; high risk) and referral pathways tailored to the risk levels. In the SAFE START model, psychosocial risk is measured in three levels (Level 1, Level 2, Level 3).

The study was a prospective cohort that recruited women attending their first antenatal visit at Royal Hospital for Women in Sydney. The study recruited 3,673 women into the usual care model and 3,132 women into the PIPA model. The PIPA model was found to have a marginally lower sensitivity (0.78 vs 0.82) but a higher specificity (0.89 vs 0.74) than the usual care model. The positive predictive value (PPV) was higher in the PIPA model than the usual care model (0.69 vs 0.41) and the negative predictive value (NPV) of both PIPA model and usual care model were very similar (0.93 and 0.95 respectively).

The cost analysis was performed from a healthcare perspective over a one-year time horizon and focused solely on screening and triage activities (the time commitment of the screening midwives, administrative staff and Triage Committee clinicians). The reference standard for effectiveness was midwives' opinion, measured in terms of True Positives (i.e. the screening midwife agreed with the positive 'at-risk' flag and the woman was referred to the Triage Committee) and False Positives (i.e. the screening midwife did not agree with the positive 'at-risk' flag and the woman was not referred to the Triage Committee). The mean cost of integrated psychosocial assessment and depression screening was marginally lower in the PIPA model than the usual care model ($13.63 vs $14.38). The economic evaluation showed that the cost per correct referral was lower by $3.97 per woman in the PIPA model compared with usual care model, and the cost saved per false positive avoided was $5.80 per woman. Overall, the results indicate that the PIPA model was cost-saving; however, the authors acknowledge that studies that address overall cost-effectiveness and longer-term health outcomes are warranted.

5.2 Screening for anxiety

Austin 2021 presents a study that forms part of the PIPA Project, underway at the Royal Hospital for Women in Randwick, NSW. The authors present a comparison of the psychometric characteristics of four commonly used anxiety screening tools. The four tools were: the EPDS anxiety subscale (EPDS-3A), the GAD-7 and GAD-2 scales, and the anxiety items of the ANRQ (ANRQ-2A). The EPDS-3A and GAD-2 were administered as part of the respective EPDS and GAD-7 tools of which they are a subset.

The study evaluated the tools in 954 women attending an antenatal appointment during their third trimester at a large Sydney teaching hospital. The authors noted the limited empirical evidence to support the current consensus-based recommendation to use EPDS-3A or ANRQ-2A as part of universal clinical assessment during the perinatal period (for anxiety screening in Australia).

The authors pre-specified the same technical performance criteria as reported in Fairbrother 2019 to define a tool suitable for widespread clinical use: AUC ≥ 0.8; Youden's index ≥ 0.5; NPV ≥ 0.8, and LR+ ≥ 4.0. All four tools met the criteria for AUC, Youden's index and NPV but only EPDS-3A (at a ≥6 cut-off) met the LR+ criterion.

Importantly, although all women completed the reference standard within two weeks of the screening measures, the majority (90.5%) actually completed it on the same day. The reference standard, however, did not include specific phobia among the anxiety disorder modules, which likely contributed to low observed prevalence of anxiety disorders (3.0%) in the sample compared to other studies. The authors posit that this may have resulted in risk of specificity bias.

Compared to non-participants (also having antenatal appointments at the hospital), the sample subjects were slightly older, less likely to have an unplanned pregnancy, majority university educated (82.0%) and partnered (i.e. in a relationship)(99.1%). Thus, the sample was not considered representative of pregnant women in Australia.

The paper concludes that the findings support the use of the EPDS-3A and (to a lesser extent) ANRQ-2A in a perinatal clinical setting, but the results require replication in more socioeconomically diverse populations with higher prevalence rates for anxiety disorders.

Results of the study are presented in Table App. 12 through Table App. 16.

Table App. 12 Evidence Summary: tools for detection of anxiety in perinatal women (Austin 2022)

Population	N	Cut-off	Reference Standard	LR+ (95% CI)	LR– (95% CI)	Consistency Cronbach's α	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	Youden index
ANRQ-2A										
Antenatal, Australia	954	≥6	DSM-V via SAGE-SR	3.335 [CI ranges: NR]	0.303 [CI ranges: NR]	0.672	0.767 [CI ranges: NR]	0.839 [CI ranges: NR]	0.843 (0.772–0.913)	0.536
EPDS-3A										
Antenatal, Australia	954	≥5	DSM-V via SAGE-SR	4.348 [CI ranges: NR]	0.358 [CI ranges: NR]	0.757	0.700 [CI ranges: NR]	0.839 [CI ranges: NR]	0.809 (0.724–0.893)	0.540
GAD-7										
Antenatal, Australia	954	≥4	DSM-V via SAGE-SR	2.721 [CI ranges: NR]	0.364 [CI ranges: NR]	0.830	0.800 [CI ranges: NR]	0.706 [CI ranges: NR]	0.818 (0.739–0.897)	0.506
GAD-2										
Antenatal, Australia	954	≥2	DSM-V via SAGE-SR	3.977 [CI ranges: NR]	0.283 [CI ranges: NR]	0.777	0.700 [CI ranges: NR]	0.824 [CI ranges: NR]	0.834 (0.758–0.910)	0.524

Abbreviations: ANRQ-2A, 2 'anxiety' items from the Antenatal Risk Questionnaire; AUC, area under the curve; CI, confidence interval; DSM-V, Diagnostic and Statistical Manual of Mental Disorders, 5th edition; EPDS-3A, Edinburgh Postnatal Depression Scale – Anxiety subscale; GAD-2, Generalized Anxiety Disorder 2-item scale; GAD-7, Generalized Anxiety Disorder 7-item scale; LR+, positive likelihood ratio; LR-, negative likelihood ratio; mo, months; NR, not reported; SAGE-SR, Series of Assessments for Guiding Evaluation – Self-Report.

Table App. 13 Summary of Findings: ANRQ-2A for detection of anxiety in antenatal women (Austin 2022)

Tool; Condition; Cutt-off	No. of studies (participants)	Critical outcomes			Important outcomes		
		LR+ (95% CI)	LR– (95% CI)	AUROC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Youden index
ANRQ-2A; anxiety disorders (GAD, panic disorder, agoraphobia, social anxiety disorder, OCD); ≥6	1 (954)	3.335 [CI ranges: NR]	0.303 [CI ranges: NR]	0.843 (0.772–0.913)	0.767 [CI ranges: NR]	0.839 [CI ranges: NR]	0.536

Abbreviations: ANRQ-2A, 2 'anxiety' items from the Antenatal Risk Questionnaire; AUROC, area under the receiver-operating characteristics curve; CI, confidence interval; GAD, generalised anxiety disorder; LR+, positive likelihood ratio; LR-, negative likelihood ratio; NR, not reported; OCD, obsessive compulsive disorder.

Table App. 14 Summary of Findings: EPDS-3A for detection of anxiety in perinatal women (Austin 2022)

Tool; Condition; Cutt-off	No. of studies (participants)	Critical outcomes			Important outcomes		
		LR+ (95% CI)	LR– (95% CI)	AUROC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Youden index
EPDS-3A; anxiety disorders (GAD, panic disorder, agoraphobia, social anxiety disorder, OCD); ≥5	1 (954)	4.348 [CI ranges: NR]	0.358 [CI ranges: NR]	0.809 (0.724–0.893)	0.700 [CI ranges: NR]	0.839 [CI ranges: NR]	0.540

Abbreviations: AUROC, area under the receiver-operating characteristics curve; CI, confidence interval; EPDS-3A, Edinburgh Postnatal Depression Scale – Anxiety subscale; GAD, generalised anxiety disorder; LR+, positive likelihood ratio; LR-, negative likelihood ratio; NR, not reported; OCD, obsessive compulsive disorder.

Table App. 15 Summary of Findings: GAD-7 for detection of anxiety in perinatal (Austin 2022)

Tool; Condition; Cutt-off	No. of studies (participants)	Critical outcomes			Important outcomes		
		LR+ (95% CI)	LR– (95% CI)	AUROC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Youden index
GAD-7; anxiety disorders (GAD, panic disorder, agoraphobia, social anxiety disorder, OCD); ≥4	1 (954)	2.721 [CI ranges: NR]	0.364 [CI ranges: NR]	0.818 (0.739–0.897)	0.800 [CI ranges: NR]	0.706 [CI ranges: NR]	0.506

Abbreviations: AUROC, area under the receiver-operating characteristics curve; CI, confidence interval; GAD, generalised anxiety disorder; GAD-7, Generalized Anxiety Disorder 7-item scale; LR+, positive likelihood ratio; LR-, negative likelihood ratio; NR, not reported; OCD, obsessive compulsive disorder.

Table App. 16 Summary of Findings: GAD-2 for detection of anxiety in perinatal women (Austin 2022)

Tool; Condition; Cutt-off	No. of studies (participants)	Critical outcomes			Important outcomes		
		LR+ (95% CI)	LR– (95% CI)	AUROC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Youden index
GAD-2; anxiety disorders (GAD, panic disorder, agoraphobia, social anxiety disorder, OCD); ≥2	1 (954)	3.977 [CI ranges: NR]	0.283 [CI ranges: NR]	0.834 (0.758–0.910)	0.700 [CI ranges: NR]	0.824 [CI ranges: NR]	0.524

Abbreviations: AUROC, area under the receiver-operating characteristics curve; CI, confidence interval; GAD, generalised anxiety disorder; GAD-2, Generalized Anxiety Disorder 2-item scale; LR+, positive likelihood ratio; LR-, negative likelihood ratio; NR, not reported; OCD, obsessive compulsive disorder.

Appendix 6 QUADAS-2 quality assessment

6.1 Screening for depression

Review question	What is the performance (defined as reliability, sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio) of validated tools for perinatal depression screening?
Patients:	Pregnant or postnatal women (birthing parent)
Index test(s):	Validated screening tools to identify people with depression in the perinatal period. Limited to tools investigated in the 2017 Australian Perinatal Mental Health Guideline: EPDS, PHQ-2/PHQ-9, K10, Whooley questions, DASS-21, HADS
Reference standard and target condition:	Any type of standardised diagnostic interview, defined as a structured interview (such as the SCID, CIDI or MINI) delivered by trained staff, or an ICD mental health diagnosis by a psychiatrist or clinical psychologist
Outcomes:	Critical outcomes: LR+, LR-, AUROC Important outcomes: sensitivity, specificity

Abbreviations: AUROC, area under the receiver operating characteristic curve; CIDI, Composite International Diagnostic Interview; DASS-21, Depression Anxiety Stress Scales; DSM, Diagnostic and Statistical Manual of Mental Disorders; EPDS, Edinburgh Postnatal Depression Scale; HADS, Hospital Anxiety and Depression Scale; ICD, International Classification of Diseases and Related Health Problems; K10, Kessler Psychological Distress Scale (10 item); LR+, positive likelihood ratio; LR-, negative likelihood ratio; MINI, Mini-International Neuropsychiatric Interview; PHQ-2, first 2 items of the PHQ-9; PHQ-9, Patient Health Questionnaire-9; SCID, Structured Clinical Interview for DSM Disorders.

Table App. 17 Quality assessment: Ezirim 2021

QUADAS-2 assessment	Ezirim 2021
Domain 1: Patient selection	
Risk of bias	
Description of patient selection methods:	All participants at a single academic hospital were screened between 3 and 24 hours postpartum as per hospital policy and again at 6-week follow-up visit. Participants without both results were excluded as were women with fetal demise.
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Risk: High – women at higher risk of depression might be less likely to attend the follow-up visit and would therefore be excluded.
Concerns regarding applicability	
Description of included patients:	Women who were between 3 and 24 hours postpartum and 6-weeks postpartum.
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	EPDS administered as routine clinical care. 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?	Unclear
If a threshold was used, was it pre-specified?	Yes – based on existing literature
Could the conduct or interpretation of the index test have introduced bias?	Risk: Unclear – administering clinician likely to have access to patient history
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	Women with EPDS scores >9 during the immediate postpartum interval were evaluated by a resident physician using the DSM-V criteria.
Is the reference standard likely to correctly classify the target condition?	Yes

QUADAS-2 assessment	Ezirim 2021
Were the reference standard results interpreted without knowledge of the results of the index test	No
Could the reference standard, its conduct, or its interpretation have introduced bias?	Risk: High – only undertaken for those with elevated EPDS (partial verification bias)
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	Patients without both results, including women who refused screening or failed to attend the post-natal visit, were excluded as were women with fetal demise. Women who did not have elevated EPDS scores perinatally did not receive the reference standard.
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?	Unclear
Did all patients receive the reference standard?	No
Did patients receive the same reference standard?	No
Were all patients included in the analysis?	No
Could the patient flow have introduced bias?	Risk: High

Table App. 18 Quality assessment: Blackmore 2022

QUADAS-2 assessment	Blackmore, 2022
Domain 1: Patient selection	
Risk of bias	
Description of patient selection methods:	A single health service with a clinic designated for refugee's antenatal care. Women attending a clinic providing antenatal care for refugees who listed their preferred language as Dari were identified and contacted prior to their appointment. Seventy-three women were approached in the clinic and 20 (27%) declined to participate.
Was a consecutive or random sample of patients enrolled?	No
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Risk: Low – although some women refused to participate, these women were screened and there was no significant difference between mean EPDS scores of those who did or did not participate. Reasons for non-participation were recorded.
Concerns regarding applicability	
Description of included patients:	Women were between 13 and 39 weeks gestation. Women aged 18 years and over; currently pregnant; and arrived to Australia on a humanitarian visa, asylum seeker or spousal visa were invited to attend. Women presenting with an acute psychotic episode, intellectual impairment or any serious complications with their pregnancy resulting in termination were excluded from participating.
Is there concern that the included patients do not match the review question?	Concern: Low – but is a specific sub-population (low generalisability)
Domain 2a: Index test - EPDS	
Risk of bias	
Description of the index test and how it was conducted and interpreted	Women completed the EPDS in Dari on a tablet device using the digital platform iCOPE, which was developed by the Centre of Perinatal Excellence. The process used to translate the Dari version was described in Shafiei, 2015.
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

QUADAS-2 assessment	**Blackmore, 2022**
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	Women attended an interview with the researcher, a registered psychologist, who was blinded to their EPDS screening scores, in order to complete the SCID-5-RV (American Psychiatric Association, 2013). During the interview, the depressive, anxiety and trauma disorder modules were administered, along with the Harvard Trauma Questionnaire (HTQ) (Mollica et al., 1992). Two other members of the project team, both psychologists, joined to form an expert panel to review all clinical interview material for each participant, in order to determine consensus DSM-5 diagnoses.
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	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 and all were included in the analysis.
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?	Yes
Could the patient flow have introduced bias?	Risk: Low

6.2 Screening for anxiety

Review question	What is the performance (defined as reliability, sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio) of validated tools for perinatal anxiety screening?
Patients:	Pregnant or postnatal women (birthing parent)
Index test(s):	Validated screening tools to identify people with anxiety in the perinatal period. Limited to tools investigated in the 2017 Australian Perinatal Mental Health Guideline: EPDS, K10, DASS-21, GAD-2/GAD-7, GHQ, HADS, HADS-A, STAI, ANRQ [2 anxiety items]
Reference standard and target condition:	Any type of standardised diagnostic interview, defined as a structured interview (such as the SCID, CIDI or MINI) delivered by trained staff, or an ICD mental health diagnosis by a psychiatrist or clinical psychologist
Outcomes:	Critical outcomes: LR+, LR-, AUROC Important outcomes: sensitivity, specificity

Abbreviations: AUROC, area under the receiver operating characteristic curve; CIDI, Composite International Diagnostic Interview; DASS-21, Depression Anxiety Stress Scales; DSM, Diagnostic and Statistical Manual of Mental Disorders; EPDS, Edinburgh Postnatal Depression Scale; GAD-2, Generalized Anxiety Disorder 2-item scale; GAD-7, Generalized Anxiety Disorder 7-item scale; GHQ, General Health Questionnaire; HADS, Hospital Anxiety and Depression Scale; HADS-A, Hospital Anxiety and Depression Scale – Anxiety subscale; ICD, International Classification of Diseases and Related Health Problems; K10, Kessler Psychological Distress Scale (10 item); LR+, positive likelihood ratio; LR-, negative likelihood ratio; MINI, Mini-International Neuropsychiatric Interview; STAI, State-Trait Anxiety Inventory; SCID, Structured Clinical Interview for DSM Disorders.

Note: Blackmore 2022 is included in screening for depression and anxiety. See **Appendix 6.1** for QUADAS-2 assessment.

Table App. 19 Quality assessment: Fairbrother 2019

QUADAS-2 assessment	Fairbrother, 2019
Domain 1: Patient selection	
Risk of bias	
Description of patient selection methods:	Participants were recruited in pregnancy via prenatal clinic visits, physician offices and midwifery clinics at three different hospitals, through community outreach at events and through word of mouth. The primary method of recruitment was direct-approach (i.e., approaching women as they waited for their appointments). The remainder were recruited through the use of posters and pamphlets.
Was a consecutive or random sample of patients enrolled?	No
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Risk: Unclear – participants opt-in to the study
Concerns regarding applicability	
Description of included patients:	Women were screened six to eight weeks postpartum. Eligibility: fluent English and lived in City of Vancouver.
Is there concern that the included patients do not match the review question?	Concern: Low
Domain 2a: Index test – GAD-7/GAD-2	
Risk of bias	
Description of the index test and how it was conducted and interpreted	The GAD-7 is a seven item self-report measure designed to assess for generalized anxiety disorder (GAD). Within the GAD-7, the first two items comprise a distinct subscale that assess the core anxiety symptoms, called the GAD-2. Women completed the screening tool and returned them by mail to the investigators.
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
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 – EPDS-3-A	
Risk of bias	

QUADAS-2 assessment	**Fairbrother, 2019**
Description of the index test and how it was conducted and interpreted	The EPDS is the most widely used screening tool for postpartum depression, items 3, 4 and 5 of the EPDS have been identified, via factor analytic studies, as a distinct anxiety subscale, namely the EPDS-3-A. Women completed the screening tool and returned them by mail to the investigators.
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
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	The Structured Clinical Interview for DSM-IV (SCID-IV; (First et al., 1996) is a reliable and valid semi-structured diagnostic interview designed for the assessment of a wide range of psychiatric problems. Interviews were conducted by PhD students in clinical and counselling psychology degree programs. Case defined as: (a) the full composite of the AD as well as (b) the five core anxiety disorders (i.e., panic disorder, agoraphobia, specific phobia, generalized anxiety disorder and social anxiety disorder) only.
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	No
Could the reference standard, its conduct, or its interpretation have introduced bias?	Risk: High– only undertaken for those with elevated screening results (partial verification bias)
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	Only participants with positive index test results had 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	Reference standard was conducted up to 7 weeks after the index tests.
Was there an appropriate interval between index test(s) and reference standard?	No
Did all patients receive the reference standard?	No
Did patients receive the same reference standard?	No
Were all patients included in the analysis?	Unclear
Could the patient flow have introduced bias?	Risk: High