Australian Perinatal Mental Health Guideline Evidence Review

Technical Report Part A

Overall approach and governance

Prepared by

hereco
clinical evidence for the real world

June 2017

TABLE OF CONTENTS

A1 SCOPE AND PURPOSE 5

A1.1 Objectives 5

A1.2 Questions 5

A1.2.1 Psychosocial assessment 6

A1.2.2 Screening 6

A1.2.3 Effectiveness of interventions 7

A1.2.4 Harms associated with selected interventions 8

A1.3 Population 10

A2 STAKEHOLDER INVOLVEMENT 11

A2.1 Group membership 11

A2.1.1 Identification of key stakeholders and establishment of the Expert Working Group 11

A2.1.2 Methodologists 14

A2.1.3 Technical writer 14

A2.2 Target population preferences and views 14

A2.3 Target users 15

A3 RIGOUR OF DEVELOPMENT 16

A3.1 Search methods 16

A3.2 Evidence selection criteria 18

A3.2.1 Psychosocial assessment and screening 18

A3.2.2 Effectiveness of interventions 18

A3.2.3 Harms of interventions 18

A3.3 Strengths and limitations of the evidence 18

A3.4 Formulation of recommendations 19

A3.5 Consideration of benefits and harms 19

A3.6 Link between recommendations and evidence 20

A3.7 External review 21

A3.8 Updating procedure 21

A4 CLARITY OF PRESENTATION 22

A4.1 Specific and unambiguous recommendations 22

A4.2 Management options 22

A4.3 Identifiable key recommendations 22

A5 APPLICABILITY 23

A5.1 Facilitators and barriers to application 23

A5.2 Implementation advice/tools 24

A5.2.1 Resources for health professionals 24

A5.2.2 Consumer and carer resources 24
A5.3 Resource implications 24
A5.4 Monitoring/ auditing criteria 25
A6 EDITORIAL INDEPENDENCE 26
A6.1 Funding body 26
A6.2 Competing interests 26

TABLE OF TABLES

Table A2-1 Perinatal guideline Expert Working Group Members 11
Table A2-2 Perinatal guideline Proxy Representatives 12
Table A2-3 Members of the Expert Committees 13
Table A3-1 Literature searches 17
Table A3-2 Cross-referencing to Technical Report for Evidence Based Recommendations 20

A1 SCOPE AND PURPOSE

A1.1 OBJECTIVES

The objective of the Perinatal Mental Health Guideline is to guide best practice in the identification, prevention and treatment/management of mental health disorders that may occur during pregnancy or in the first year following the birth of a baby (the perinatal period).

Health intents

The Guideline is designed to guide health professionals in the identification of the more common disorders (depression and anxiety), together with the prevention and treatment of these conditions through a range of treatment approaches that includes psychosocial and psychological therapies, pharmacological, complementary and physical therapies.

In addition, the Guideline addresses the management of low prevalence, more severe mental illnesses - namely schizophrenia, bipolar disorder, postpartum psychosis, and borderline personality disorder (BPD). For each of these conditions the Guideline provides guidance in the provision of psychosocial and psychological therapies, pharmacological and physical therapies.

Expected benefits or outcomes

Through undertaking a review of the latest evidence, the first aim of the Guideline was to identify current and effective tools for the detection of those most at risk of perinatal mental health conditions (psychosocial assessment) as well as those likely to be experiencing symptoms of the more common conditions (screening tools). The second aim of the Guideline was to assess the evidence of interventions for managing mental health disorders, with a particular focus on the impact to the offspring of *in utero* exposure to systemically active treatments (i.e., medications, complementary therapies and some physical therapies).

The intention is that these evidence-based findings can inform local, state and national policy surrounding the timely implementation of appropriate tools to ensure early identification of needs and timely, safe (for mother and baby) and effective intervention. Early detection and management of perinatal mental health disorders will serve to have significant health and economic benefits for the woman, her family and the broader community.

Target

The target audience for this Guideline is primary health professionals caring for women as they plan pregnancy, and throughout the perinatal period. This includes but is not limited to: midwives, Child and Family Health Nurses, Mental Healthcare Workers, General Practitioners, and obstetricians.

A1.2 QUESTIONS

There were three main topics under investigation for this Guideline:

- Identification of mental health problems during the perinatal period using psychosocial assessment and screening [assessed in **Part B** of the **Technical Report**].
- The effectiveness of treatment or prevention of mental health problems during the perinatal period using various interventions [assessed in **Part C** of the **Technical Report**]
- Harms to the offspring or mother related to the use of selected interventions [assessed in **Part D** of the **Technical Report**].

To address these three topics, eight main questions were formulated, with each question being broken down into multiple sub-questions based on population, intervention or outcome.

In addition to these clinical questions, the resource implications of perinatal mental health screening have been addressed via targeted consideration of published cost-effectiveness analyses.

A1.2.1 Psychosocial assessment

The main question relating to psychosocial assessment was broken down into five sub-questions based on different outcomes. It should be noted that each sub-question could be broken down further into individual psychological instruments/tool and outcomes.

Main question:

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

Sub-questions:

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

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

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

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

1e. 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; see **Part B Technical Report**]

A1.2.2 Screening

The two main questions relating to the depression and anxiety screening for pregnant or postpartum women were each broken down into five sub-questions based on different outcomes. It should be noted that each sub-question could be broken down further into individual tools/in and outcomes.

A1.2.2.1 Depression screening

The main question relating to depression screening was broken down into five sub-questions based on different outcomes.

Main question:

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

Sub-questions:

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

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

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

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

2e. 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; see **Part B Technical Report**]

A1.2.2.2 Anxiety screening

The main question relating to anxiety screening was broken down into five sub-questions based on different outcomes.

Main question:

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

Sub-questions:

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

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

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

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

3e. 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; see **Part B Technical Report**]

A1.2.3 Effectiveness of interventions

The two main questions relating to the effectiveness of interventions for the treatment of mental health problems in pregnant or postpartum women, or prevention of mental health problems in pregnant or postpartum women identified as being at risk of developing mental health problems, were each broken down into five sub-questions based on different intervention types. It should be noted that each sub-question could be broken down further into individual interventions and outcomes. The detailed definitions associated with these interventions and outcomes can be found in Section C2.2 of the Part C Technical Report.

A1.2.3.1 Treatment interventions

Main question:

4. What is the efficacy and safety of interventions for the treatment of mental health problems in women in the antenatal or postnatal period?

Sub-questions:

4a. What is the efficacy and safety of psychosocial interventions for the treatment of mental health problems in women in the antenatal or postnatal period? [addressed via systematic review; see **Part C Technical Report**]

4b. What is the efficacy and safety of psychological interventions for the treatment of mental health problems in women in the antenatal or postnatal period? [addressed via systematic review; see **Part C Technical Report**]

4c. What is the efficacy and safety of pharmacological interventions for the treatment of mental health problems in women in the antenatal or postnatal period? [addressed via systematic review; see **Part C Technical Report**]

4d. What is the efficacy and safety of complementary interventions for the treatment of mental health problems in women in the antenatal or postnatal period? [addressed via systematic review; see **Part C Technical Report**]

4e. What is the efficacy and safety of physical interventions for the treatment of mental health problems in women in the antenatal or postnatal period? [addressed via systematic review; see **Part C Technical Report**]

A1.2.3.2 Prevention interventions

Main question:

5. What is the efficacy and safety of interventions for the prevention of mental health problems in women in identified as being at risk of developing a mental health problem in the antenatal or postnatal period?

Sub-questions:

5a. What is the efficacy and safety of psychosocial interventions for the prevention of mental health problems in women identified as being at risk of developing a mental health problem in the antenatal or postnatal period? [addressed via systematic review; see **Part C Technical Report**]

5b. What is the efficacy and safety of psychological interventions for the prevention of mental health problems in women identified as being at risk of developing a mental health problem in the antenatal or postnatal period? [addressed via systematic review; see **Part C Technical Report**]

5c. What is the efficacy and safety of pharmacological interventions for the prevention of mental health problems in women identified as being at risk of developing a mental health problem in the antenatal or postnatal period? [addressed via systematic review; see **Part C Technical Report**]

5d. What is the efficacy and safety of complementary interventions for the prevention of mental health problems in women identified as being at risk of developing a mental health problem in the antenatal or postnatal period? [addressed via systematic review; see **Part C Technical Report**]

5e. What is the efficacy and safety of physical interventions for the prevention of mental health problems in women identified as being at risk of developing a mental health problem in the antenatal or postnatal period? [addressed via systematic review; see **Part C Technical Report**]

The detailed PICO definitions associated with these questions can be found in Section C2.2 of the Part C Technical Report.

A1.2.4 Harms associated with selected interventions

The four main questions relating to the harms associated with interventions for the treatment of mental health problems in pregnant or postpartum women, or prevention of mental health problems in pregnant or postpartum women identified as being at risk of developing mental health problems, were each broken down into four sub-questions based on the different populations that are harmed. It should be noted that each sub-question could be broken down further into individual interventions and outcomes. The detailed

definitions associated with these interventions and outcomes can be found in Section D2.2 of the Part D Technical Report.

A1.2.4.1 Pharmacological interventions

Main question:

6. What are the harms that occur as a result of perinatal exposure to pharmacological interventions used for the treatment of mental health problems?

Sub-questions:

6a. What are the harms that occur to the fetus as a result of perinatal exposure to a pharmacological intervention used for the treatment of mental health problems? [malformations; addressed via systematic review; see **Part D Technical Report**]

6b. What are the harms that occur to the infant as a result of perinatal exposure to a pharmacological intervention used for the treatment of mental health problems? [pregnancy and birth outcomes; addressed via systematic review; **Part D Technical Report**]

6c. What are the harms that occur to the child as a result of perinatal exposure to a pharmacological intervention used for the treatment of mental health problems? [neurodevelopmental outcomes; addressed via systematic review; **Part D Technical Report**]

6d. What are the harms that occur to the mother as a result of perinatal exposure to a pharmacological intervention used for the treatment of mental health problems? [postpartum haemorrhage; addressed via systematic review; see **Part D Technical Report**]

A1.2.4.2 Complementary interventions

Main question:

7. What are the harms that occur as a result of perinatal exposure to a complementary intervention used for the treatment of mental health problems?

Sub-questions:

7a. What are the harms that occur to the fetus as a result of perinatal exposure to a complementary intervention used for the treatment of mental health problems? [malformations; addressed via systematic review; see **Part D Technical Report**]

7b. What are the harms that occur to the infant as a result of perinatal exposure to a complementary intervention used for the treatment of mental health problems? [pregnancy and birth outcomes; addressed via systematic review; see **Part D Technical Report**]

7c. What are the harms that occur to the child as a result of perinatal exposure to a complementary intervention used for the treatment of mental health problems? [neurodevelopmental outcomes; addressed via systematic review; see **Part D Technical Report**]

7d. What are the harms that occur to the mother as a result of perinatal exposure to a complementary intervention used for the treatment of mental health problems? [postpartum haemorrhage; addressed via systematic review; see **Part D Technical Report**]

A1.2.4.3 Physical interventions

Main question:

8. What are the harms that occur as a result of perinatal exposure to a physical intervention used for the treatment of mental health problems?

Sub-questions:

8a. What are the harms that occur to the fetus as a result of perinatal exposure to a physical intervention used for the treatment of mental health problems? [malformations; addressed via systematic review; see **Part D Technical Report**]

8b. What are the harms that occur to the infant as a result of perinatal exposure to a physical intervention used for the treatment of mental health problems? [pregnancy and birth outcomes; addressed via systematic review; see **Part D Technical Report**]

8c. What are the harms that occur to the child as a result of perinatal exposure to a physical intervention used for the treatment of mental health problems? [neurodevelopmental outcomes; addressed via systematic review; see **Part D Technical Report**]

8d. What are the harms that occur to the mother as a result of perinatal exposure to a physical intervention used for the treatment of mental health problems? [postpartum haemorrhage; addressed via systematic review; see **Part D Technical Report**]

A1.3 POPULATION

The population to whom the Guideline applies includes all pregnant or postnatal women, with the postnatal period being defined as the 12 months following birth. As this guideline also provide an assessment of the harms associated with interventions used for the treatment or prevention of perinatal mental health issues, the population also encompasses the offspring of these women.

Attention is also given to women with a history of mental health issues who might be planning a pregnancy.

A2 STAKEHOLDER INVOLVEMENT

A2.1 GROUP MEMBERSHIP

A2.1.1 Identification of key stakeholders and establishment of the Expert Working Group

On the commissioning of this Guideline, the Executive Director of COPE wrote to all company members, inviting their respective College or Organisation to nominate a representative for the Guideline Expert Working Group (EWG). In doing so the College was asked to consider the expertise and representation of the College in the area of perinatal mental health specifically.

Company Members are as follows:

- Australian College of Mental Health Nurses (ACMHN)
- Australian College of Midwives (ACM)
- Australian Psychological Society (APS)
- Maternal Child and Family Health Nursing Association (MCaFNA)
- Post and Antenatal Depression Association (PANDA)
- Royal Australian College of General Practitioners (RACGP)
- Royal Australian New Zealand College of Obstetricians and Gynaecologists (RANZCOG)
- Royal Australian and New Zealand College of Psychiatrists (RANZCP)
- Congress of Aboriginal and Torres Strait Islander Nurses and Midwives (CATSINaM)

The role of EWG members was to provide oversight regarding the scope of the Guideline and agreement on the clinical questions, including the mental health disorders to be included and the outcomes of most relevance to women in the perinatal period and different users of the Guideline. The inclusion of two representatives from PANDA on the EWG is one way in which the Guideline captures consumer (patients') views and preferences.

The nominated members assigned to the EWG from their respective Colleges and Organisations is detailed in **Table A2-1**

Table A2-1 Perinatal guideline Expert Working Group Members

Representative	Expertise	Organisation Representing	Institutional Affiliation(s)	Geographical Location
Professor Marie-Paule Austin (Chair)	Perinatal Psychiatrist, Former Chair beyondblue Clinical Guideline, researcher and clinician working across private and public perinatal settings.	Royal Australian College of Psychiatrists (RANZCP)	University of New South Wales, St John of God Healthcare, Royal Hospital for Women, Black Dog Institute.	Sydney, NSW
Dr Nicole Highet (Co-chair)	Former Co-Chair & Director beyondblue perinatal Guideline, online training programs & resources. Expertise in consumer/carer research, advocacy, policy & implementation.	Centre of Perinatal Excellence (COPE)	Centre of Perinatal Excellence (COPE)	Flemington, Vic.
Dr James Best	General Practitioner with specialist training and expertise in perinatal mental health.	Royal Australian College of General Practitioners (RACGP)	Your Doctors (Medical Practice)	Leichhardt & Summer Hill, NSW

Representative	Expertise	Organisation Representing	Institutional Affiliation(s)	Geographical Location
Mr Andrew Davis	Carer Representative and volunteer at PANDA.	Carer Representative	None	Melbourne, Vic.
Ms Suzanne Higgins	Mental health nurse and social worker with expertise clinical and educational expertise in perinatal mental health.	Australian College of Mental Health Nurses (ACMHN)	St John of God Healthcare.	Geelong, Vic.
Dr Helen Lindner	Health psychologist and former member of the EWG for beyondblue perinatal guideline.	Australian Psychological Society (APS)	Australian Psychological Society (APS)	Melbourne, Vic.
Ms Creina Mitchell	Clinician, researcher and educator in maternal and child health with expertise and interest in perinatal mental health.	Maternal & Child Family Health Australia (MCaFHNA)	Griffith University	Brisbane, Qld.
Ms Jenni Richardson	Consumer representative involved in the management of PANDA helpline with expertise in consumer needs, experiences and advocacy.	Consumer representative Perinatal Anxiety and Depression Association (PANDA)	Perinatal Anxiety and Depression Association (PANDA)	Fitzroy, Vic
Ms Terri Smith	Chief Executive Officer of PANDA	Consumer representative Perinatal Anxiety and Depression Association (PANDA)	Perinatal Anxiety and Depression Association (PANDA)	Fitzroy, Vic
Dr Vijay Roach	Obstetrician with dedicated expertise in perinatal mental health. Chair of the Gidget Foundation (perinatal mental health support organisation) and carer.	Royal Australian College of Obstetricians and Gynaecologists (RANZCOG)	Royal North Shore Hospital	Sydney, NSW
Prof. Rhonda Marriott	Midwife, researcher and specialist in Aboriginal and Torres Strait Islander perinatal mental health.	ATSI Representative	Murdoch University	Perth, WA
Dr Jan Taylor	Clinician, researcher and educator in midwifery with expertise in perinatal mental health. Former member of the beyondblue EWG.	Australian College of Midwives (ACM)	Canberra University	Canberra, ACT

In order to ensure adequate representation at all meetings, if assigned EWG members are not able to attend a meeting, they are asked to nominate a proxy for the meeting from their respective College/organisation – again reflecting the interests and expertise of the College/organisation in perinatal mental health. Proxys attending over the course of the Guideline development process are detailed below.

Table A2-2 Perinatal guideline Proxy Representatives

Representative	Expertise	Organisation Representing	Institutional Affiliation(s)	Geographical Location
Dr Anne Sved Williams	Perinatal Psychiatrist. Head, Medical Unit.	Royal Australian College of Psychiatrists (RANZCP)	Helen Mayo House	Adelaide, SA

Representative	Expertise	Organisation Representing	Institutional Affiliation(s)	Geographical Location
Julie Ferguson	Mental health nurse practitioner with expertise clinical and educational expertise in perinatal mental health.	Australian College of Mental Health Nurses (ACMHN)	St John of God Healthcare.	Geelong, Vic.
Dr Louise Roufiele	Psychologist and researcher in perinatal mental health.	Australian Psychological Society (APS)	Australian Psychological Society (APS)	Melbourne, Vic.
Ms Terri Smith	Consumer representative involved in the management of PANDA helpline with expertise in consumer needs, experiences and advocacy.	Consumer representative Perinatal Anxiety and Depression Association (PANDA)	Perinatal Anxiety and Depression Association (PANDA)	Fitzroy, Vic
Dr Agnes Wilson	Senior policy advisor, Royal Australian College of Obstetricians and Gynaecologists (RANZCOG)	Royal Australian College of Obstetricians and Gynaecologists (RANZCOG)	Royal North Shore Hospital	Sydney, NSW
Dr Catherine Chamberline	Senior Research Fellow, College of Science, Health and Engineering, School of Nursing and Midwifery Department of Public Health	ATSI Representative	La Trobe University	Melbourne, Vic

In addition to the Expert Working Group, two additional committees were formed to provide opportunity for a broader representation of specialist clinical expertise.

Expert Committee #1: Harms expert committee

The first of these pertained to with respect addressing the potential harms associated with treatment for perinatal mental health (particularly pharmacological treatments).

Expert Committee #2: Low prevalence disorder expert committee

Following the initial EWG meeting when the scope of the Guideline was discussed, there was strong support for the scope of the perinatal mental health Guideline to include a broader range of mental health disorders. As such, the scope of the Guideline was increased to include Borderline Personality Disorder and Schizophrenia. In line with this change in scope, a second committee with specialist expertise in these illness areas was also formed.

Both expert committees contained representation of recognised experts (identified by the EWG) and contained representatives from perinatal psychiatry and pharmacotherapy. A list of the members of each of these committees is detailed in **Table A2-3**.

Table A2-3 Members of the Expert Committees

Representative	Expertise	Institutional Affiliation	Geographical location	Expert committee
Professor Marie-Paule Austin (Chair)	Chair Perinatal Mental Health Unit, Professorial Fellow and Consultant Psychiatrist	University of New South Wales, Black Dog Institute, University of New South Wales.	Sydney, NSW	Harms Low prevalence disorders
Professor Phillip Boyce	Professor of Psychiatry, Perinatal Psychiatrist	University of Sydney and Westmead Hospital.	Wentworthville, NSW	Harms Low prevalence disorders
Professor Megan Gallbally	Foundation Chair in Perinatal Psychiatry & Perinatal Psychiatrist.	University of Notre Dame, Fiona Stanley Hospital.	Perth, Western Australia.	Harms

Representative	Expertise	Institutional Affiliation	Geographical location	Expert committee
Professor Anne Buist	Perinatal Psychiatrist & Director, North-East Women's Mental Health Parent Infant Program	Austin Hospital and University of Melbourne	Heidelberg West, Victoria.	Low prevalence disorders
Dr Debra Kennedy	Director, Mothersafe.	Royal Hospital for Women	Sydney, NSW.	Harms
Dr Tram Nguyen	Consultant Psychiatrist. Centre for Women's Mental Health.	The Royal Women's Hospital	Melbourne, Vic.	Harms Low prevalence disorders
Dr Anne Sved-Williams	Perinatal Psychiatrist. Head, Medical Unit.	Helen Mayo House Family Unit.	Glenside, South Australia.	Harms Low prevalence disorders
Dr Sylvia Lim-Gibson	Consultant Psychiatrist, Royal Hospital for Women. Director of Postgraduate Psychiatry Training for the South-East Sydney and Illawarra.	Royal Hospital for Women, NSW. Brain and Mind Research Institute.	NSW	Low prevalence disorders
Professor Louise Newman	Director, Centre of Women's Mental Health.	The Royal Women's Hospital	Melbourne, Vic.	Low prevalence disorders

A2.1.2 Methodologists

In addition, the Guideline developer (COPE) engaged the skills and expertise of a team of Guideline methodologists led by Dr Sarah Norris (Health Research Consulting; hereco). The methodologists were engaged by the Guideline developer due to their knowledge and experience in guideline development, and in particular, their authorship of the systematic reviews associated with the initial (Beyondblue) Guideline.

A2.1.3 Technical writer

A technical writer, Jenny Ramson (Ampersand), was also engaged to draft the Guideline. The contractor was selected due their experience in the writing of the current National Antenatal Care Guideline as well as the writing of the development of the initial Commonwealth/beyondblue Guideline. This technical writer will also be invited to undertake the writing of the companion documents that will be produced following submission to (and approval by) the NHMRC.

A2.2 TARGET POPULATION PREFERENCES AND VIEWS

All COPE members were informed of the development of the Guideline through the COPE Annual General Meeting (November 2016). COPE collaborated with each of the professional bodies identified in Section A2.1.1 (who are members of COPE) to disseminate inform and engage College members and consumer groups for consultation. This ensured widespread consultation with health professionals involved in the delivery of primary, maternity and mental health care, as well as those involved in the education, screening, identification and provision of treatment across both primary and specialist care settings. In addition, the peak bodies/leaders in perinatal mental health working with Aboriginal and Torres Strait Islander and culturally and linguistically diverse (CALD) populations will be approached in the consultation phase to provide feedback on the Clinical Guideline.

In putting the Guideline out for public consultation, individuals or organisations will be invited to provide feedback in writing to the Expert Working Group for Consideration. All information gathered from the Consultation period will be reviewed by the EWG in consultation with the Expert Committees for consideration in the formation of recommendations and finalisation of the Guideline.

On the finalisation of the Guideline, companion documents will be developed for consumers, carers and health professionals. Feedback on drafted documents will be obtained through consultation with

representatives from all professional bodies as well as consumers and carers to ensure relevance and appropriate presentation of information.

A2.3 TARGET USERS

The target populations for this guideline can be considered threefold:

1) **Primary healthcare professionals**

   In order to ensure inclusion of views and preferences from each target group, health professionals were identified through approaching their respective colleges and asking them to nominate a representative (as per Section A2.1.1). This approach to the Colleges (as opposed to selecting individuals) ensured a transparent approach to nomination onto the expert working group. In addition to seeking representatives from each professional body, a representative from the Aboriginal and Torres Strait Islander community was selected through an approach to the Congress of Aboriginal and Torres Strait Islander Nurses and Midwives, and inclusion of a representative and proxy onto the EWG.

2) **Consumers of health services**

   In order to ensure the views of consumers and carers are reflected in the development of the Guideline, the Perinatal Anxiety and Depression Association (PANDA) was approached to nominate a consumer representative to the Expert working group.

3) **Carers**

   Perinatal mental health conditions can have a significant impact upon carers and family members, and this was reflected in the inclusion of a carer representative on the Expert Working Group, also identified through PANDA.

The Guideline will be used by each of the professional groups in accordance with their role in the management of perinatal health. For example, those involved at the front-end of maternity care provision (GPs, midwives and obstetricians) will be informed about best practice screening and assessment tools to identify and respond to identified mental health problems in pregnancy, whilst those professionals involved in the provision of treatment for mental health conditions (psychiatrists, psychologists, GPs) will likely refer to the information surrounding safe and effective treatments for perinatal mental health conditions. Consumers and carers will also refer to the Guideline to obtain information surrounding the assessment of risk and symptom detection, as well and the recommended safe and effective treatments for perinatal mental health. Specifically, this will include the development of tailored factsheets and resources for consumers and carers as well as health professionals (see **Section A5**).

A3 RIGOUR OF DEVELOPMENT

A3.1 SEARCH METHODS

A summary of the searches performed, databases used and search dates is presented in **Table A3-1**; further details are provided below. Full details of all searches can be found in the Appendices for Part B (**Section B8.1**), Part C (**Section AppC1**) and Part D (**Section AppD1)** of the Technical Report.

Searches were conducted in the MEDLINE, Embase and PsychINFO databases, and also in CINAHL for psychosocial assessment and screening (via the OVID and/or Embase.com interfaces), various databases of the Cochrane Library, and included examination of the reference lists of included SRs and individual studies. Searches were conducted between June 2016 and April 2017.

It should be noted that 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.

A3.1.1.1 Psychosocial and screening searches

A two-tiered search strategy was undertaken as follows:

1. An initial systematic review search (SR search) identified systematic reviews that assessed various treatments for the main mental health disorders seen during the perinatal period; these included depression, anxiety, puerperal psychosis, bipolar disorder and schizophrenia. It should be noted that this search was conducted to identify studies not only for the assessment of effectiveness and harms, but also for screening.

   From this search, an initial list was assembled of SRs that reported on various aspects of psychosocial assessment and screening (as well as effectiveness and harms of interventions used for the treatment or prevention of mental health problems in pregnant or postpartum women). The individual studies included in each SR were identified and, where possible, a 'foundation review' was identified. The foundation review was defined as the SR that included the most recent and comprehensive set of data for a particular psychosocial assessment or screening tool, and if suitable could be included in the evidence review; if not suitable for inclusion, the foundation review could be used to identify relevant individual studies.

2. Supplementary searches for individual studies were then conducted as required. Circumstances for supplementary searches included updating literature searches from foundation reviews that were more than 3 years old (for psychosocial assessment) or re-specifying literature searches from foundation reviews because they did not completely align with the questions within the current Guideline (anxiety screening).

A3.1.1.2 Intervention searches

A two-tiered search strategy was also undertaken as follows:

3. An initial SR search identified systematic reviews that assessed various treatments for the main mental health disorders seen during the perinatal period; these included depression, anxiety, puerperal psychosis, bipolar disorder and schizophrenia. It should be noted that this search was conducted to identify studies not only for the assessment of effectiveness and harms, but also for screening.

From this search, an initial list was assembled of SRs that assessed effectiveness (and harms for the systemically-active interventions) associated with the psychosocial, psychological, pharmacological, complementary and physical therapies included in the review. The individual studies included in each SR were identified and, where possible, a 'foundation review' was identified. The foundation review was defined as the SR that included the most recent and comprehensive set of data for a particular intervention and outcome, and if suitable could be included in the evidence review; if not suitable for inclusion, the foundation review could be used to identify relevant individual studies.

4. Based on the findings of the SR search, a second series of literature searches were carried out for the systemically-active interventions[1] and online interventions ('updated searches'). These searches aimed to identify additional SRs, and individual RCTs and observational studies, and were based on the interventions of interest as follows:
    - Where a suitable foundation review was identified, the search was limited from the year of the foundation review's literature search up to October 2016. Date-limited searches were conducted for all pharmacological agents except z-drugs, and the complementary therapy omega-3 fatty acids.
    - Where no suitable foundation review was identified, no initial date limit was set, and the search was conducted up to October 2016. Extended date searches were conducted for z-drugs, the complementary therapies St John's Wort and Gingko biloba, and the physical therapies electroconvulsive therapy (ECT) and transcranial magnetic stimulation (TMS).

Table A3-1 Literature searches

Search	Question type	Study type/s	Databases	Date
Psychosocial assessment	Psychosocial risk factor assessment	**SR search**	MEDLINE, Embase and PsychINFO (via OVID) CDSR, DARE and HTA (via Cochrane Library)	01 Jun 16
		Supplementary search	Medline, EMBASE, PsychInfo, CINAHL (via OVID)	15 Dec 16
Depression screening	Screening	**SR search**	MEDLINE, Embase and PsychINFO (via OVID)	01 Jun 16
Anxiety screening	Screening	**SR search**	MEDLINE, Embase and PsychINFO (via OVID)	01 Jun 16
		Supplementary search	Medline, EMBASE, PsychInfo, CINAHL (via OVID)	21 Dec 16
Psychosocial and psychological interventions	Effectiveness	**SR search** SR (RCTs)	MEDLINE and Embase (via Embase.com) CDSR, DARE and HTA (via Cochrane Library)	01 Jun 16 29 Jul 16
Online interventions	Effectiveness	**SR search** SR (RCTs)	MEDLINE and Embase (via Embase.com) CDSR, DARE and HTA (via Cochrane Library)	01 Jun 16 29 Jul 16
		Updated search SR (RCTs), RCT	MEDLINE, Embase and PsychINFO (via OVID) All Cochrane databases (Cochrane Library)	12 Apr 17 12 Apr 17
Pharmacological agents Omega-3 fatty acids St John's Wort and Gingko biloba ECT and TMS	Effectiveness and harms	**SR search** SR (RCTs)	MEDLINE and Embase (via Embase.com) CDSR, DARE and HTA (via Cochrane Library)	01 Jun 16 29 Jul 16
		Updated search SR (RCTs/OBS), RCT, OBS	MEDLINE (PubMed) Embase and PsychINFO (OVID) All Cochrane databases (Cochrane Library)	11 Oct 16 12 Oct 16 13 Oct 16

Abbreviations: ECT, electroconvulsive therapy; OBS, observational studies; RCT, randomised control trial; SR, systematic review; TMS, transcranial magnetic stimulation.

[1] Includes pharmacological therapies (antidepressants, antipsychotics, anticonvulsants, benzodiazepines and z-drugs, and lithium), complementary therapies (omega-3 fatty acids, St John's Wort and Gingko biloba) and selected physical therapies (electroconvulsive therapy and transcranial magnetic stimulation).

A3.2 EVIDENCE SELECTION CRITERIA

Inclusion/exclusion criteria were formulated based on the PICO (Population, Intervention, Comparator, Outcome) criteria used to define the research questions. The PICO criteria for each question type can be found in the following sections of the Technical Report:

- Psychosocial assessment – **Section B3.1.2** in Part B of the Technical Report.
- Depression screening and anxiety screening – **Section B3.1.1** in Part B of the Technical Report.
- Effectiveness of interventions – **Section C2.2** in Part C of the Technical Report
- Harms of interventions – **Section D2.2** in Part D of the Technical Report.

The main inclusion/exclusion criteria for each of the research question types were as follows:

A3.2.1 Psychosocial assessment and screening

- Target population – all pregnant or postnatal women (psychosocial assessment), or pregnant or postnatal women with no known diagnosis of depression or anxiety (screening)
- Study design – prospective, controlled studies reporting predictive accuracy (psychosocial assessment) or diagnostic accuracy (screening)
- Comparisons – subsequent manifestation of mental health issues (psychosocial assessment), orany standard clinical/diagnostic interview as a reference standard (screening)
- Language – limited to English.

A3.2.2 Effectiveness of interventions

- Target population – pregnant or postnatal women diagnosed with a mental health problem, or considered to be at risk of developing a mental health problem.
- Study design – SRs of RCTs, or individual RCTs if no SR or SR out of date.
- Interventions – Psychosocial, psychological, pharmacological, complementary or physical therapies used to treat or prevent mental health problems in pregnant or postnatal women.
- Comparisons – no treatment/placebo/treatment as usual or active treatment
- Language – limited to English.

A3.2.3 Harms of interventions

- Target population – pregnant or postnatal women diagnosed with a mental health problem, or considered to be at risk of developing a mental health problem, or a fetus, infant or child of a mother exposed to a pharmacological, complementary or physical therapy.
- Study design – SRs of RCTs (if available), SRs of observational studies, or individual observational studies if no SR or SR out of date or unsuitable.
- Comparisons – no treatment/exposure or active treatment
- Language – limited to English.

A3.3 STRENGTHS AND LIMITATIONS OF THE EVIDENCE

The strengths and limitations of the evidence have been considered from the perspective of the individual studies and the body of evidence aggregated across all the studies. Wherever possible validated methods have been used to assess:

- Study design(s)
- Study methodology limitations (sampling, blinding, allocation concealment, analytical methods)
- Appropriateness/relevance of primary and secondary outcomes considered
- Consistency of results across studies
- Direction of results across studies

- Magnitude of benefit versus magnitude of harm
- Applicability to practice context

GRADE methodology was used to determine the quality of the evidence available for each intervention/outcome. The majority of the evidence for fetal harms was considered generally to be of very low (●○○○) or inadequate (○○○○) quality. It should be noted that the category 'inadequate' was added for this review to better reflect the broad range of quality that would have been considered very low if GRADE methods had been adhered to. A discussion of this adaptation of GRADE methodology can be found in **Part D Technical Report Section D2.5.1**.

In addition, no GRADE methods could be identified for the assessment of psychometric instruments. Consequently, a hybrid method was developed for quality appraisal of psychosocial assessment instruments. This method was based on accepted psychometric properties and QUADAS-2 principles and is described in detail in **Part B Technical Report Sections B4.1** and **B5.2.**

A3.4 FORMULATION OF RECOMMENDATIONS

As groups of evidence reviews were completed they were considered by the EWG and the relevant Expert Committee(s) as appropriate. General discussion of the interpretation and implications of the review findings were discussed, and then Evidence Based Recommendations developed once consensus was reached. The strength of the EBRs was agreed at this point. Once a group of related EBRs and CBRs was developed, the EWG then deliberated on the need for Practice Points to highlight important aspects of care.

The Expert Committees were engaged to provide specific expertise to support the EWG. The Harms Expert Committee were the first to review the Harms systematic reviews. This Committee then drafted proposed Recommendations, for consideration and approval by the EWG.

Once Recommendations had been developed across all types of intervention, the Low Prevalence Expert Committee then used their expertise to apply the Recommendations to perinatal women with bipolar disorder, postpartum psychosis, schizophrenia or borderline personality disorder. This process involved explicit consideration of relevant, recent Australian Guidelines for mood disorder, schizophrenia, and borderline personality disorder in general populations.

A3.5 CONSIDERATION OF BENEFITS AND HARMS

The evidence reviews present an explicit consideration of health benefits and harms. The trade-off between benefits and harms is articulated in the rationale for each Recommendation.

Recommendations regarding the use of psychosocial and psychological interventions were based primarily on evidence of the effectiveness, because they do not cause direct harm to the fetus, infant or child.

Recommendations regarding the use of pharmacological, complementary and selected physical interventions were to be based on a trade-off between effectiveness and harm; however, there was very little evidence of effectiveness for these interventions in the pregnant and postpartum population. The only evidence available was for antidepressants (suggesting it may improve postnatal depression) and omega-3 fatty acids (where it appeared to have no effect on depression).

The harms most likely to impact on recommendations were major and cardiac malformations, and neurodevelopmental harms. Due to its strong association with major and cardiac malformation, and adverse cognitive outcome, as well as a lack of evidence of effectiveness in pregnant or postpartum women with, or at risk of developing, a mental health problem, the prescribing of sodium valproate in all women of childbearing age, was strongly recommended against. The evidence of harm associated with

carbamazepine, and the lack of evidence for lamotrigine led to a consensus-based recommendation to prescribe anticonvulsants with great caution during pregnancy.

While there were a number of pregnancy and birth outcomes found to be associated with pharmacological therapies (including miscarriage, preterm birth, poor neonatal adaptation syndrome, respiratory distress, convulsions and persistent pulmonary hypertension), these were not directly captured in any recommendations; instead, a Practice Point notes that the potential risks of treatment (including the risk of relapse), as well as the benefits, should be discussed with women.

There was little evidence available of the side effects experienced by the mother of pharmacological, complementary and physical interventions assessed; these treatments are all used regularly in clinical practice and as such their side effect profiles are well established. However, based the known side effect of clozapine, agranulocytosis, a Consensus-based Recommendation states that its use should not be initiated during pregnancy due to a theoretical potential harm to the infant.

A3.6 LINK BETWEEN RECOMMENDATIONS AND EVIDENCE

An explicit link has been made between the evidence and the recommendations arising from that evidence. In circumstances where evidence was sought and not found, or where evidence was relied on from other populations (e.g. a general depressed population, not a perinatal depressed population) then Consensus Based Recommendations were developed. Aspects of care that were not within scope of the evidence reviews are then captured in Practice Points.

An explicit Evidence to Decision framework was developed for psychosocial assessment and screening to capture evidence beyond related to predictive or diagnostic accuracy. Consideration of Recommendations regarding treatment options explicitly considered the effectiveness of the interventions and the harms to the fetus.

The links between the EBRs and the evidence reviews are shown in **Table A3-2.**

Table A3-2 Cross-referencing to Technical Report for Evidence Based Recommendations

Evidence-based Recommendations		Strength	Location of evidence	
			Benefit	Harm
Screening and assessment				
Screening for depression				
1	Use the EPDS to screen women for a possible depressive disorder in the perinatal period	Strong	Part B Tech Report – **Table B5-20**	Discussed but not formally assessed
2	Arrange further assessment of perinatal woman with an EPDS score of 13 or more.	Strong	Part B Tech Report – **Table B5-20**	Discussed but not formally assessed
Assessing psychosocial risk				
3	Use the ANRQ to assess the presence of psychosocial risk.	Strong	Part B Tech Report – **Table B4-13**	Discussed but not formally assessed
Prevention and treatment				
Depressive and anxiety disorders				
Psychosocial support and psychological approaches				
4	Provide structured psychoeducation to women with symptoms of depression in the perinatal period.	Strong	Part C Tech Report – **Table C3-1**	Not assessed
5	Advise women with symptoms of depression in the postnatal period of the potential benefits of a social support group.	Conditional	Part C Tech Report – **Table C3-4**	Not assessed
6	Recommend individual structured psychological interventions (cognitive behavioural therapy or interpersonal psychotherapy) to women with mild to moderate depression in the perinatal period.	Strong	Part C Tech Report – **Table C3-19**	Not assessed
7	Advise women with depression or anxiety disorder in the postnatal period of the possible benefits of directive counselling.	Conditional	Part C Tech Report – **Table C3-22**	Not assessed

Evidence-based Recommendations		Strength	Location of evidence	
			Benefit	Harm
Complementary therapies				
8	Advise women who enquire about omega-3 fatty acid supplementation that it does not appear to improve depression symptoms but is not harmful to the offspring when taken during pregnancy or while breastfeeding.	Conditional	Part C Tech Report – **Table C3-38 and Table C4-28 to 29**	Part D Tech Report – **Table D3.40**
Pharmacological				
9	Consider the use of SSRIs as first-line treatment for moderate to severe depression in pregnant women.	Conditional	No evidence available	Part D Tech Report – **Table D3-2**
10	Recommend the use of SSRIs as first-line treatment for moderate to severe depression in postnatal women	Strong	Part C Tech Report – **Table C3-31**	Part D Tech Report – **Table D3-2**
Severe mental illness				
Antipsychotics				
11	Consider the use of antipsychotics for treating psychotic symptoms in pregnant women	Conditional	No evidence available	Part D Tech Report – **Tables D3-14 to 25**
Anticonvulsants				
12	Do not prescribe sodium valproate to women of childbearing age	Strong	No evidence available	Part D Tech Report – **Table D3-27**

Abbreviations: ANRQ, antenatal risk questionnaire; EPDS, Edinburgh Postnatal Depression Scale; SSRI, selective serotonin reuptake inhibitor;

A3.7 EXTERNAL REVIEW

Methodology used to conduct external review will be completed after public consultation, and will cover:

- *Purpose and intent of the external review (e.g., to improve quality, gather feedback on draft recommendations, assess applicability and feasibility, disseminate evidence)*
- *Methods taken to undertake the external review (e.g., rating scale, open-ended questions)*
- *Description of the external reviewers (e.g., number, type of reviewers, affiliations)*
- *Outcomes/information gathered from the external review (e.g., summary of key findings)*
- *How the information gathered was used to inform the guideline development process and/or formation of the recommendations (e.g., guideline panel considered results of review in forming final recommendations)*

A3.8 UPDATING PROCEDURE

The procedure to be used to update the guidelines will be completed after public consultation, and will cover:

- *A statement that the guideline will be updated*
- *Explicit time interval or explicit criteria to guide decisions about when an update will occur*
- *Methodology for the updating procedure*

A4 CLARITY OF PRESENTATION

A4.1 SPECIFIC AND UNAMBIGUOUS RECOMMENDATIONS

In developing the Recommendations and Practice Points the developers have adhered to the following principles:

- A succinct statement of the recommended action,
- In a clearly stated relevant population (e.g., pregnant women, postnatal women, or perinatal women)
- At a specific timing, if appropriate.

The rationale for each Recommendation and Practice Point covers: the intent or purpose of the recommended action (e.g., to improve quality of life, to decrease side effects); any caveats or qualifying statements (e.g., patients or conditions for whom the recommendations would not apply); and if there is uncertainty about the best care option(s), a description of the nature of that uncertainty.

A4.2 MANAGEMENT OPTIONS

The Guideline addresses multiple management options and these are clearly articulated via the structure of the Guideline and the wording of the Recommendations and Practice Points.

A4.3 IDENTIFIABLE KEY RECOMMENDATIONS

The key recommendations are presented in a way that is easy to identify, and to differentiate between Evidence-Based Recommendations (EBRs), Consensus Based Recommendations (CBRs) and Practice Points (PPs). The strength of the EBRs is also clearly identified as either ‘Strong’ or ‘Conditional’.

A5 APPLICABILITY

A5.1 FACILITATORS AND BARRIERS TO APPLICATION

There are a number of facilitators to guideline application which include:

- Engagement of key stakeholders in the Guideline Development
  Peak bodies provide varying aspects of perinatal health and mental health care, each of whom have been involved in the development of the Guideline from the outset.
- Infrastructure of peak bodies
  Each of the Colleges will play a key role in communicating the Guideline to their members and advocating for their implementation through communication with College members in newsletters, academic publications in journals and presentation at conferences.
- The infrastructure of the health system
  The framework of maternity, postnatal and primary care provision provides a vehicle for all aspects of guideline implementation from consumer education through to screening and assessment and treatment provision. The health and community care landscape has been taken into account when considering the Guideline application across maternity, postnatal, general practice, public and private healthcare settings as well as the range of services available across jurisdictions.
- The history of the National Perinatal Depression Initiative (NPDI)
  The Commonwealth Government's investment into the NPDI with States and Territories (2008-15) has provided some valuable history and infrastructure to implementation of the Guideline. Current investment is variable across States and Territories. For example, whilst some States (eg. NSW) have state-wide policies surrounding in relation to screening, in other states this has been discontinued in the absence of funding. Awareness of the state of play across each jurisdiction and ongoing relationships and collaboration with key Commonwealth and State Government and Policy Stakeholders since the NPDI, will serve to provide an opportunity to continue to advocate and seek support for national Guideline implementation.
- The development of a perinatal mental health website to house all information for consumers, carers and health professions
  Since the release of the initial (beyondblue) Guideline, COPE (Guideline developer) has been established to provide a dedicated focus on perinatal mental health. As part of this work, an extensive website has been developed to provide best practice information for consumers, carers and health professionals (www.cope.org.au). The website will be updated to reflect the latest evidence for existing disorders, and be expanded to include the additional mental health disorders that have been specifically addressed in the current Guideline (schizophrenia and borderline personality disorder). In addition, this dedicated website will include all factsheets and screening aids (companion documents) and house the online training program (see below).
- The development of a free, online, accredited training program for health professionals
  To support implementation, a free online training program will accompany the release of the Guideline. This will facilitate education for health professionals in include coverage of all guideline recommendations and good practice points. In addition, all companion document that have been developed for health professionals and consumers/carers will be embedded into the online program to direct people to specific information on each topic.
- Innovative guide for consumers and carers
  In addition to the website and fact sheets for consumers and carers, as much of the Guideline focus on the need for education and information provision for consumers. In response to this, a series of

fortnightly emails for expectant and new parents will provide emotional and mental health information relative to each stage in the perinatal period, whilst providing information and links to further information and factsheets derived from the Guideline.

- Innovative technology to facilitate screening in accordance with the Guideline
  As one of the greatest barriers to screening is time taken to do screening within tight maternity and postnatal appointments, the Guideline developer (COPE) has developed a Digital screening platform that allows screening to be undertaken electronically on an iPad (http://cope.org.au/health-professionals-3/icope-digital-screening/). The feasibility trials and subsequent implementation across a range of primary, maternity and postnatal healthcare settings demonstrates the ability of the platform (iCOPE) to save time, reduce language barriers, improve screening rates in accordance with the national guideline. Through the programming of any additional Guideline recommended scales onto the iCOPE Platform this will also facilitate their application. Furthermore, the automated production of clinical reports at the time of screening serves to guide health professional in best practice with respect to screening outcomes and referral pathways. Consumers also can also access a tailored report (via email or SMS) detailing outcomes and referring to more information on the COPE website.

Barriers to application include:

- Low screening in the private sector
  The greatest barriers to implementation are likely to be found in the private system, whereby many specialist obstetricians do not prioritise perinatal mental health, but rather tend to focus on physical health. In response to this work is being led by the Royal Australian and New Zealand College of Obstetricians (RANZCOG) to include screening in accordance with the Guidelines as a Medicare item number.
- Lack of time to undertake screening and assessment
  As detailed above time is a barrier and hence this is addressed through the selection of brief assessment tools and the digitisation of screening to improve screening rates, times, accuracy and inclusiveness.

A5.2 IMPLEMENTATION ADVICE/TOOLS

A5.2.1 Resources for health professionals

As part of the Guideline, an online, fully accredited training program will be developed and made free available for health professionals. This will cover all information contained in the Guideline to support their implementation in practice. This will involve evaluation in order for professionals to receive accreditation and be supported throughout the development of a series of factsheets to guide and support assessment, management and treatment of perinatal mental health disorders.

A5.2.2 Consumer and carer resources

The free e-newsletter (Ready to COPE) will be made freely available for consumers and contain within it specifically designed factsheets.

A5.3 RESOURCE IMPLICATIONS

These have been addressed in the body of the Technical Report and the Guideline. This section will be completed after public consultation and will cover:

- *Types of cost information that were considered (e.g., economic evaluations, drug acquisition costs)*

- *Methods by which the cost information was sought (e.g., a health economist was part of the guideline development panel, use of health technology assessments for specific drugs, etc.)*
- *Information/description of the cost information that emerged from the inquiry (e.g., specific drug acquisition costs per treatment course)*
- *How the information gathered was used to inform the guideline development process and/or formation of the recommendations*

A5.4 MONITORING/ AUDITING CRITERIA

As the peak body for perinatal mental health in Australia, COPE will continue to consult with service providers nationally to ensure the dissemination and application of the Clinical Guideline across the country. For those utilizing the digital screening, this will enable the monitoring of screening rates and outcomes across sites and setting in real time. Further the integration of clinical advice into the clinical reporting facilitated by the iCOPE platform will serve to inform and guide best practice by the health professionals.

Further to this, COPE will continue to liaise with representatives of all states and territories involved in the implementation of perinatal mental health initiatives.

A6 EDITORIAL INDEPENDENCE

A6.1 FUNDING BODY

The funding body was the Commonwealth Government of Australia. The Government had no direct or indirect influence on the content of the Guideline. Specifically, the Commonwealth Government were not involved in any of the committees or content reviews, rather were only engaged in reporting of progress (as opposed to the specific content).

A6.2 COMPETING INTERESTS

All group members have declared whether they have any competing interests at the outset of the Guideline development process and at each subsequent meeting.

At the outset of the Guideline development process, all representatives were informed of the importance of managing competing interests and ensuring that any potential conflicts of interest were identified in advance of any meeting (as evidenced in meeting minutes). Processes put in place to manage any potential conflicts of interest are as follows:

- All EWG members and proxies involved in the Guideline development process were required to complete a Declaration of Interest Form (as per the NHMRC requirements). These signed and scanned forms were reviewed by the Co-Chairs of the EWG and are held by the Guideline developer.
- On sending out agenda papers, EWG members were to be informed of the arising agenda items and asked to notify the Chairperson in advance of the meeting of any potential conflicts of interest that had arisen since the most recent meeting.
- Any arising conflicts of interest were to be adjudicated by the Chair and Co-Chair. In the event of a conflict of interest held by the Chair, this was to be managed by the Co-Chair and the area of conflict clearly stated. In this instance, as with other conflicts of interest declared by other EWG members, members were to be invited to take part and contribute discussions, however were asked to leave the room when forming recommendations. In the instance where the Chair has a declared conflict of interest, this will be managed by the Co-Chair.
- If a conflict of interest was deemed to be material prior to a meeting, the member was asked to continue to contribute to the committee, with the above measures taken to limit the introduction of bias.

There was only one instance of a possible competing interest, and that was the review of a clinical scale which was developed by one of the expert working group members. This was made known to all members of the expert working group at the outset of these discussions. To address this competing issue when reviewing the evidence, the member of the group was involved in the discussion, however when deciding on its inclusion the member was asked to remove him or herself from the discussion and decision-making process.

Australian Perinatal Mental Health Guideline Evidence Review

Technical Report Part B

Psychosocial assessment and screening for depression or anxiety

Prepared by

hereco
clinical evidence for the real world

June 2017

TABLE OF CONTENTS

B1 INTRODUCTION 1

B2 CLINICAL QUESTIONS 3

B2.1 Psychosocial assessment 3

B2.2 Depression screening 4

B2.3 Anxiety screening 5

B3 SEARCH METHODS 6

B3.1 Overall approach to searches 6

B3.1.1 Screening 7

B3.1.2 Psychosocial assessment 10

B3.1.3 Acceptability, Effectiveness and Implementation 12

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

B4 PSYCHOSOCIAL ASSESSMENT 13

B4.1 Relevant outcomes and critical appraisal methods 13

B4.2 Characteristics of studies of technical performance 13

B4.3 Evidence profile tables 18

B4.4 Non-technical characteristics of relevant tools 20

B4.5 Clinical usefulness of relevant tools 20

B4.5.1 ALPHA 20

B4.5.2 ANRQ 21

B4.5.3 PRQ 23

B4.6 General evidence of clinical usefulness 23

B4.7 Overall summary of findings 25

B5 SCREENING FOR DEPRESSION 27

B5.1 Relevant outcomes of technical performance 27

B5.2 Critical appraisal of technical performance 28

B5.2.1 Quality assessment of individual studies 28

B5.2.2 Overall certainty of the evidence by outcome 29

B5.3 Evidence of technical performance 29

B5.3.1 Characteristics of studies in foundation review 29

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

B5.3.3 Summary of findings regarding technical performance 41

B5.4 Non-technical characteristics of relevant tools 50

B5.5 Clinical usefulness of relevant tools 50

B5.5.1 Acceptability of depression screening 50

B5.5.2 Effectiveness of depression screening 51

B5.5.3 Implementation of depression screening 52

B5.6 Cost-effectiveness of perinatal depression screening 52

B5.7 Overall summary of findings 53

B6 SCREENING FOR ANXIETY 55

B6.1 Relevant outcomes of technical performance 55

B6.2 Critical appraisal of technical performance 55

B6.3 Evidence of technical performance 55

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

B6.3.2 Evidence summaries from included studies 58

B6.3.3 Summary of findings regarding technical performance 64

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

B6.5 Clinical usefulness of relevant tools 68

B6.6 Overall summary of findings 68

B7 REFERENCES 70

B8 APPENDICES 72

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

B8.1.1 Search strings 72

B8.1.2 Exclusion of studies 77

B8.1.3 Excluded studies list 78

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

B8.2.1 Depression screening SRs 78

B8.3 Lists of included studies from supplementary searches 82

B8.3.1 Anxiety screening 82

B8.3.2 Psychosocial assessment 82

B8.4 Details of QUADAS-2 assessments 82

B8.4.1 Questions for quality assessment of diagnostic studies 82

B8.4.2 Checklists for additional screening studies 83

TABLE OF TABLES

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

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

TABLE OF FIGURES

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

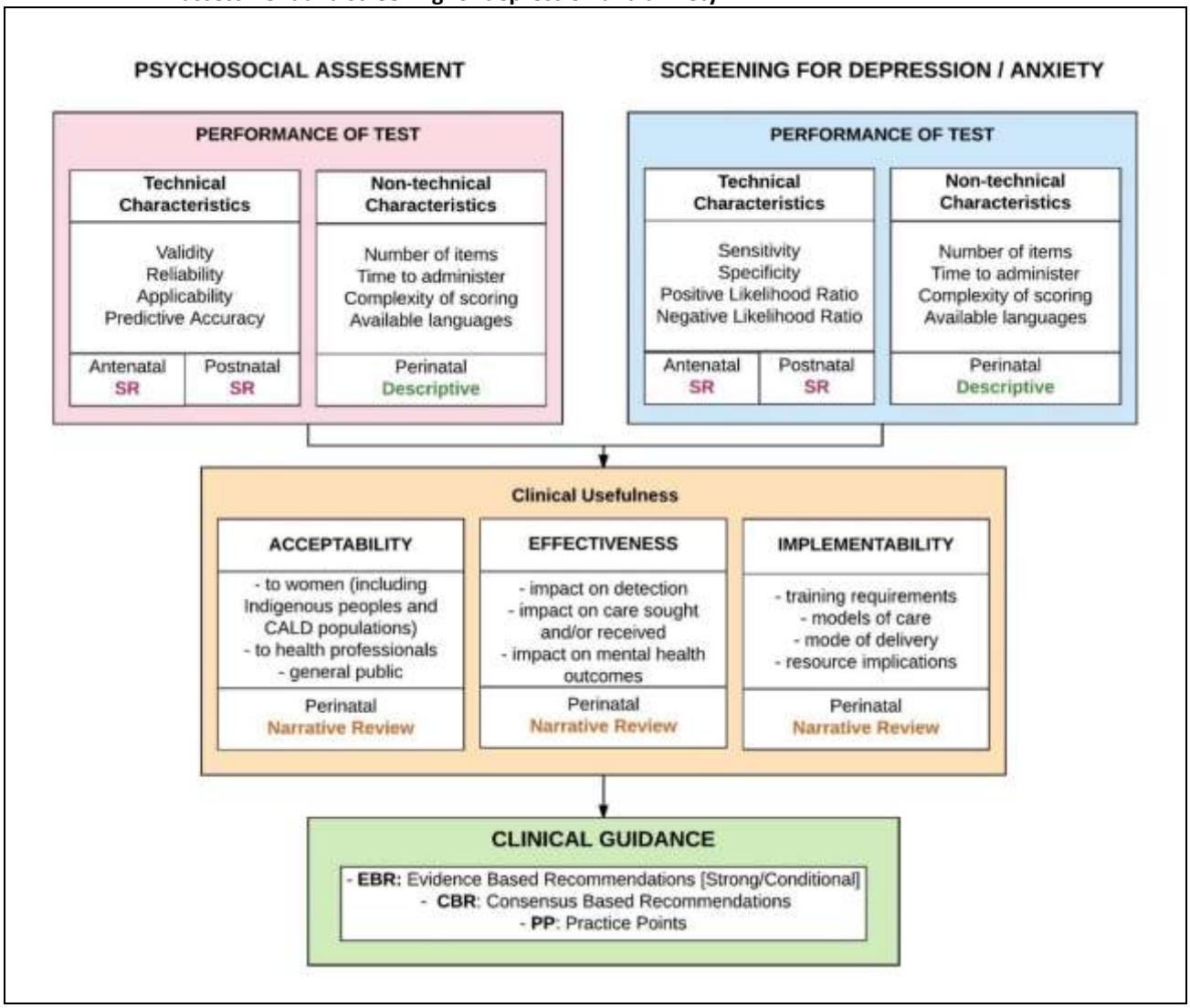
B1 INTRODUCTION

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

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

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

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



Abbreviations: SR, systematic review.

B2 CLINICAL QUESTIONS

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

B2.1 PSYCHOSOCIAL ASSESSMENT

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

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

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

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

Main question:

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

Sub-questions:

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

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

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

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

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

B2.2 DEPRESSION SCREENING

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

Main question:

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

Sub-questions:

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

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

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

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

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

B2.3 ANXIETY SCREENING

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

Main question:

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

Sub-questions:

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

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

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

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

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

B3 SEARCH METHODS

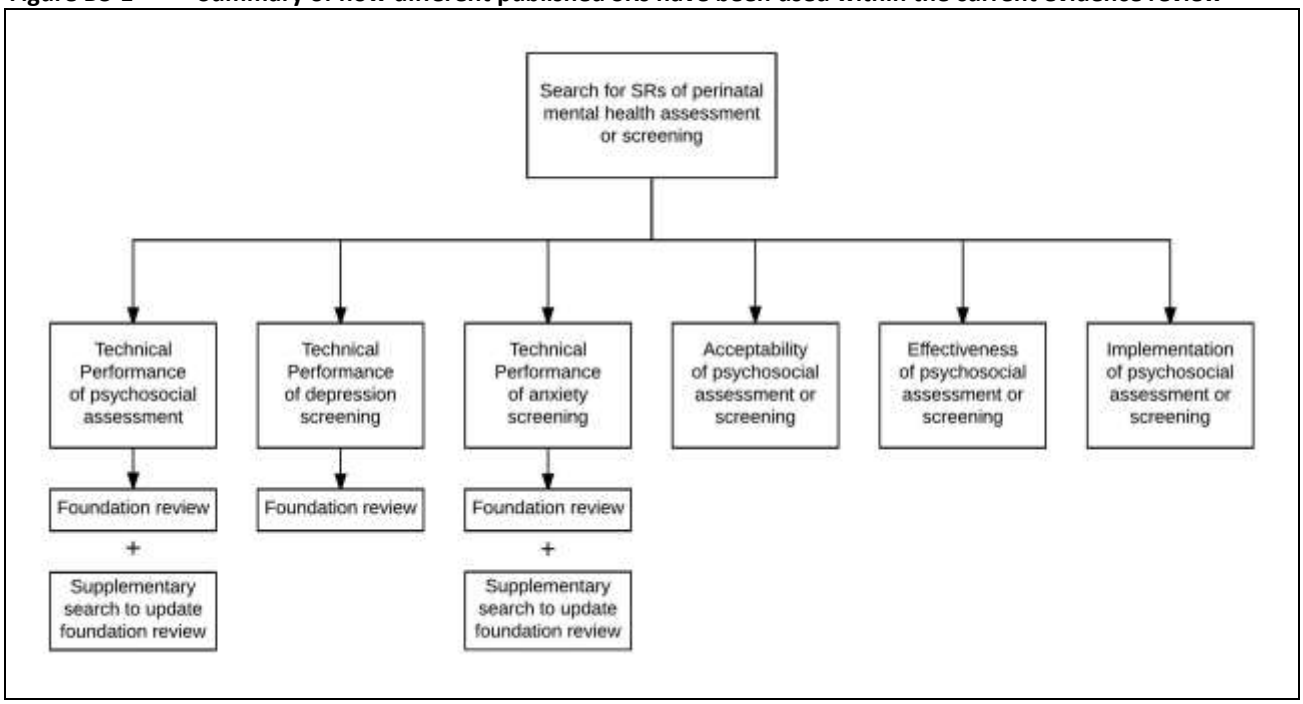
B3.1 OVERALL APPROACH TO SEARCHES

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

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

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

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



Abbreviations: SR, systematic review.

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

B3.1.1 Screening

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

B3.1.1.1 Selection of foundation review for depression screening

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

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

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

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

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

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

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

B3.1.1.2 Supplementary search for individual studies of depression screening

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

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

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

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

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

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

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

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

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

B3.1.1.3 Selection of foundation review for anxiety screening

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

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

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

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

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

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

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

B3.1.1.4 Supplementary search for individual studies for anxiety screening

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

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

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

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

- Conducted in a perinatal population

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

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

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

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

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

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

B3.1.2 Psychosocial assessment

B3.1.2.1 Selection of foundation review

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

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

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

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

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

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

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

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

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

B3.1.2.2 Supplementary search and inclusion criteria for individual studies

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

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

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

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

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

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

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

B3.1.3 Acceptability, Effectiveness and Implementation

B3.1.3.1 Approach to evidence review

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

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

B3.1.4 Cost-effectiveness of perinatal mental health screening

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

B4 PSYCHOSOCIAL ASSESSMENT

B4.1 RELEVANT OUTCOMES AND CRITICAL APPRAISAL METHODS

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

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

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

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

B4.2 CHARACTERISTICS OF STUDIES OF TECHNICAL PERFORMANCE

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

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

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

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

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

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

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

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

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

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

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

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

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

B4.3 EVIDENCE PROFILE TABLES

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

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

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

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

Evidence Statements:

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

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

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

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

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

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

B4.4 NON-TECHNICAL CHARACTERISTICS OF RELEVANT TOOLS

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

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

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

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

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

B4.5 CLINICAL USEFULNESS OF RELEVANT TOOLS

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

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

B4.5.1 ALPHA

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

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

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

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

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

Abbreviations: ALPHA, Antenatal Psychosocial Health Assessment.

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

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

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

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

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

B4.5.2 ANRQ

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

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

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

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

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

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

Abbreviations: ANRQ, Antenatal Risk Questionnaire.

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

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

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

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

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

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

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

B4.5.3 PRQ

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

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

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

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

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

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

Abbreviations: PRQ; Pregnancy Risk Questionnaire

B4.6 GENERAL EVIDENCE OF CLINICAL USEFULNESS

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

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

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

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

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

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

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

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

B4.7 OVERALL SUMMARY OF FINDINGS

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

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

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

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

Footnotes

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

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

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

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

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

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

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

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

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

B5 SCREENING FOR DEPRESSION

B5.1 RELEVANT OUTCOMES OF TECHNICAL PERFORMANCE

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

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

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

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

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

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

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

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

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

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

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

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

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

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

B5.2 CRITICAL APPRAISAL OF TECHNICAL PERFORMANCE

B5.2.1 Quality assessment of individual studies

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

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

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

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

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

B5.2.2 Overall certainty of the evidence by outcome

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

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

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

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

B5.3 EVIDENCE OF TECHNICAL PERFORMANCE

B5.3.1 Characteristics of studies in foundation review

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

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

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

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

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

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

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

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

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

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

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

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

B5.3.2 Evidence summaries based on studies in foundation review

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

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

B5.3.2.1 EPDS

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

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

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

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

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

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

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

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

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

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

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

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

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

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

B5.3.2.2 PHQ

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

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

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

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

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

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

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

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

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

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

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

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

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

B5.3.2.3 Whooley questions

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

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

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

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

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

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

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

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

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

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

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

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

B5.3.2.4 K-10

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

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

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

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

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

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

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

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

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

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

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

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

B5.3.3 Summary of findings regarding technical performance

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

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

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

B5.3.3.1 Depression screening in the antenatal period

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

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

Evidence statements:

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

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

Footnotes:

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

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

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

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

[5] Multiple studies with a range of quality.

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

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

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

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

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

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

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

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

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

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

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

B5.3.3.2 Depression screening in the postnatal period

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

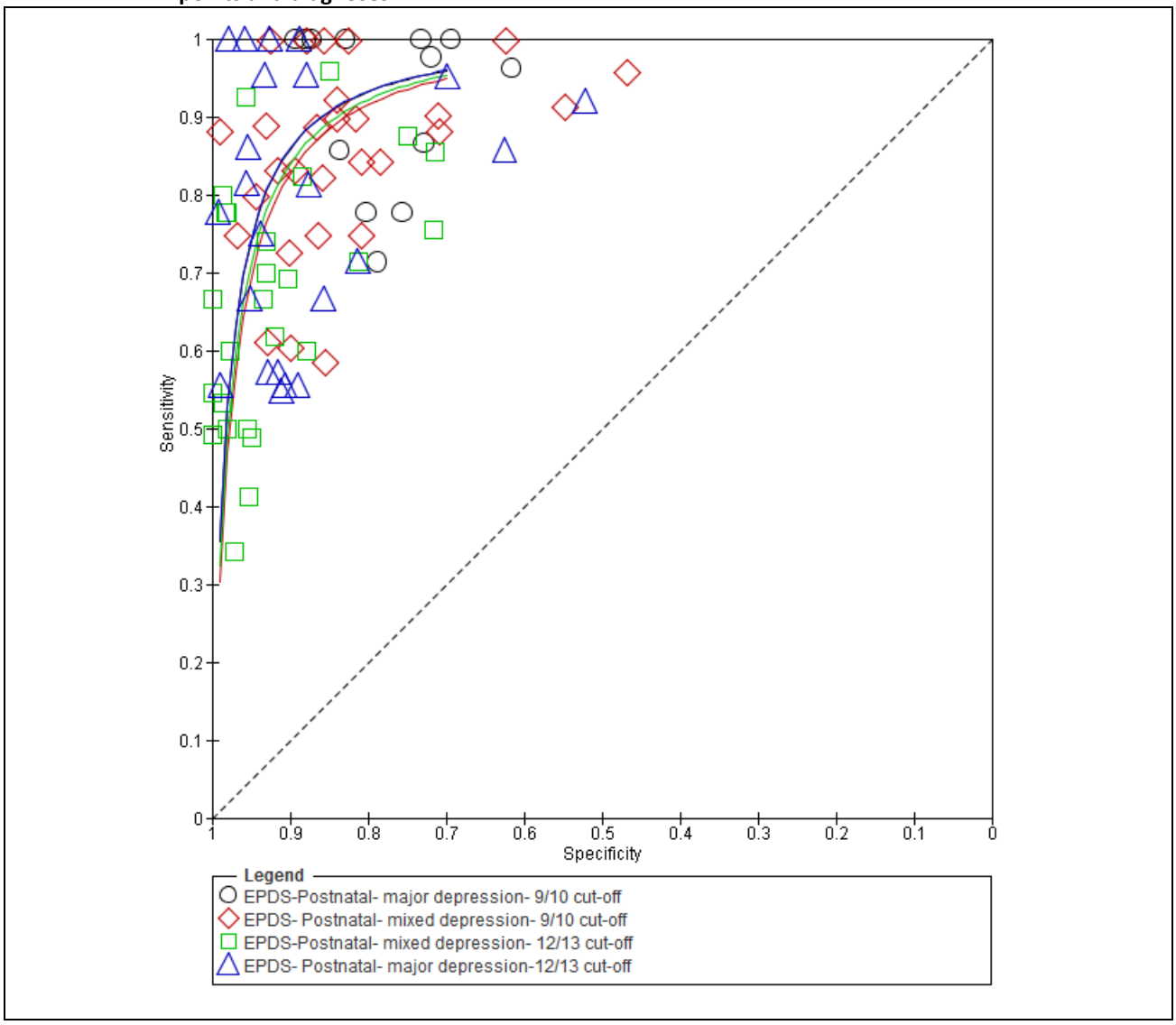
Tool; Condition; Cut-off	No of studies[1] (participants)	Important outcomes		Critical outcomes			Overall certainty
		Pooled sensitivity (95% CI) [1]	Pooled specificity (95% CI) [1]	Positive Likelihood Ratio [2]	Negative Likelihood Ratio [2]	Pooled AUC	
EPDS; mixed depression[3]; 9/10	29 (5,463)	0.83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

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

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

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

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



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

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

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

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

Footnotes:

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

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

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

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

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

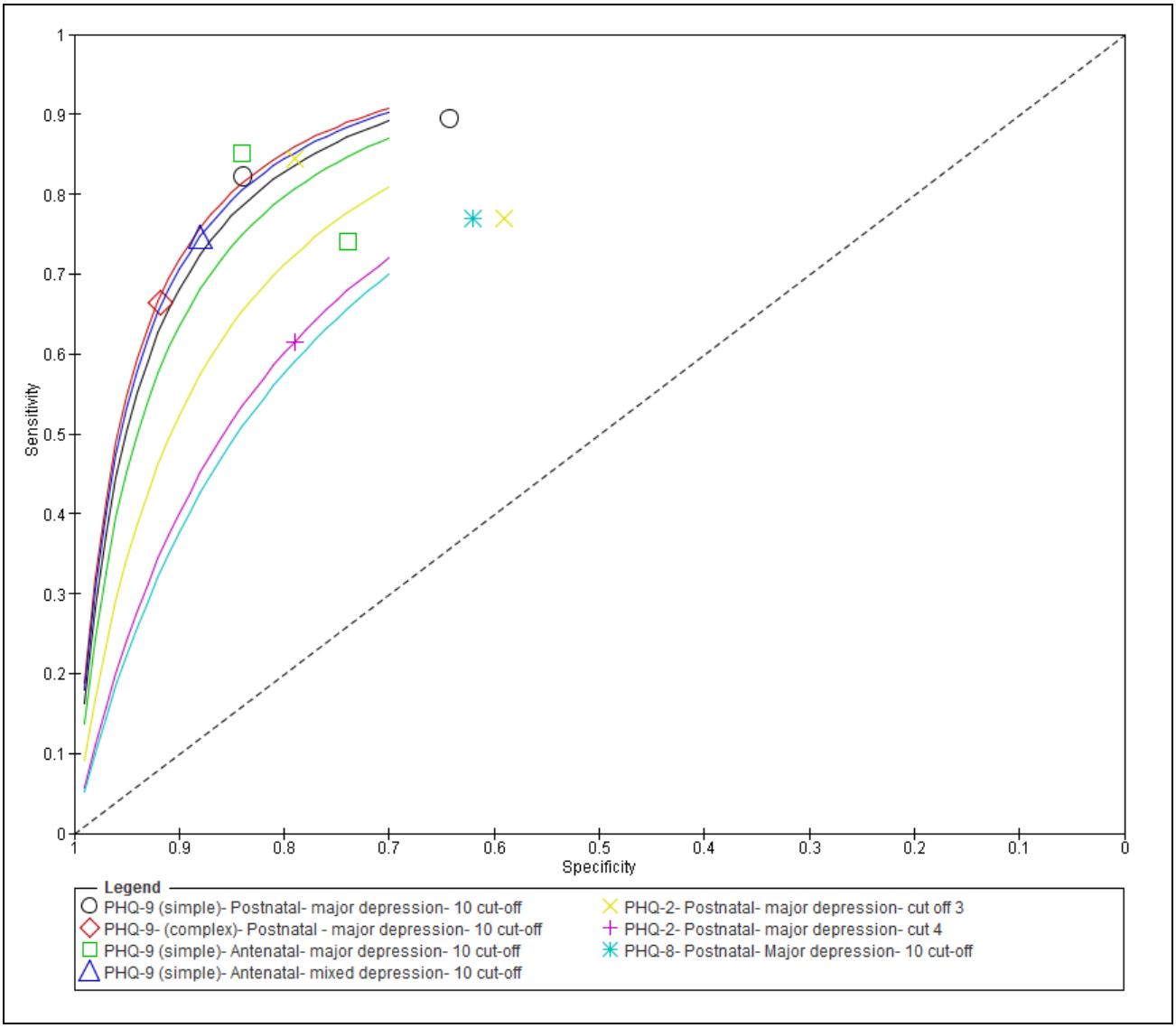
[6] Two studies of very low quality

[7] One study of very low quality

[8] Two studies of low to very low quality

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

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



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

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

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

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

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

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

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

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

B5.4 NON-TECHNICAL CHARACTERISTICS OF RELEVANT TOOLS

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

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

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

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

B5.5 CLINICAL USEFULNESS OF RELEVANT TOOLS

B5.5.1 Acceptability of depression screening

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

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

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

B5.5.2 Effectiveness of depression screening

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

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

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

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

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

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

B5.5.3 Implementation of depression screening

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

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

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

B5.6 COST-EFFECTIVENESS OF PERINATAL DEPRESSION SCREENING

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

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

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

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

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

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

B5.7 OVERALL SUMMARY OF FINDINGS

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

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

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

Footnotes:

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

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

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

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

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

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

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

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

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

B6 SCREENING FOR ANXIETY

B6.1 RELEVANT OUTCOMES OF TECHNICAL PERFORMANCE

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

B6.2 CRITICAL APPRAISAL OF TECHNICAL PERFORMANCE

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

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

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

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

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

B6.3 EVIDENCE OF TECHNICAL PERFORMANCE

B6.3.1 Characteristics of individual studies included for anxiety screening

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

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

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

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

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

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

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

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

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

Footnotes:

Italics indicate divergence from data reported in foundation review

[1] Compiled from information presented in Meades 2011.

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

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

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

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

B6.3.2 Evidence summaries from included studies

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

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

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

B6.3.2.1 EPDS

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

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

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

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

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

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

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

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

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

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

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

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

B6.3.2.2 GAD-7

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

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

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

B6.3.2.3 GHQ

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

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

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

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

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

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

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

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

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

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

B6.3.2.4 HADS

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

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

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

B6.3.2.5 K-10

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

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

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

B6.3.2.6 STAI

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

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

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

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

B6.3.3 Summary of findings regarding technical performance

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

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

B6.3.3.1 Anxiety screening in antenatal women

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

B6.4 NON-TECHNICAL CHARACTERISTICS OF RELEVANT INCLUDED TOOLS

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

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

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

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

B6.5 CLINICAL USEFULNESS OF RELEVANT TOOLS

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

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

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

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

B6.6 OVERALL SUMMARY OF FINDINGS

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

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

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

Footnotes

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

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

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

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

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

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

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

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

[9] Inferred from evidence of depression screening

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

B7 REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Simpson, W., M. Glazer, N. Michalski, M. Steiner and B. N. Frey (2014). "Comparative efficacy of the generalized anxiety disorder 7-item scale and the Edinburgh Postnatal Depression Scale as screening tools for generalized anxiety disorder in pregnancy and the postpartum period." Can J Psychiatry 59(8): 434-440.

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

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

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

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

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

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

B8 APPENDICES

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

B8.1.1 Search strings

B8.1.1.1 Systematic review search

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

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

B8.1.1.2 Psychosocial assessment search

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

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

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

B8.1.1.3 Anxiety screening search

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

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

B8.1.1.4 Economic search

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

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

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

B8.1.2 Exclusion of studies

B8.1.2.1 Systematic review search

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

B8.1.2.2 Psychosocial assessment search

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

B8.1.2.3 Anxiety screening search

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

B8.1.2.4 Economic search

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

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

TOTAL	**Included**		**0**

B8.1.3 Excluded studies list

B8.1.3.1 Systematic review search

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

B8.1.3.2 Psychosocial assessment search

Full list of excluded studies can be provided on request.

B8.1.3.3 Anxiety screening search

Full list of excluded studies can be provided on request.

B8.1.3.4 Economic search

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

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

B8.2.1 Depression screening SRs

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

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

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

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

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

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

B8.3 LISTS OF INCLUDED STUDIES FROM SUPPLEMENTARY SEARCHES

B8.3.1 Anxiety screening

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

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

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

B8.3.2 Psychosocial assessment

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

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

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

B8.4 DETAILS OF QUADAS-2 ASSESSMENTS

B8.4.1 Questions for quality assessment of diagnostic studies

Patient selection - Risk of Bias:

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

Patient selection – Applicability concerns:

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

Index test(s) – Risk of Bias:

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

Index test(s) – Applicability concerns:

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

Reference standard – Risk of Bias:

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

Reference standard – Applicability concerns:

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

Flow and Timing – Risk of Bias:

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

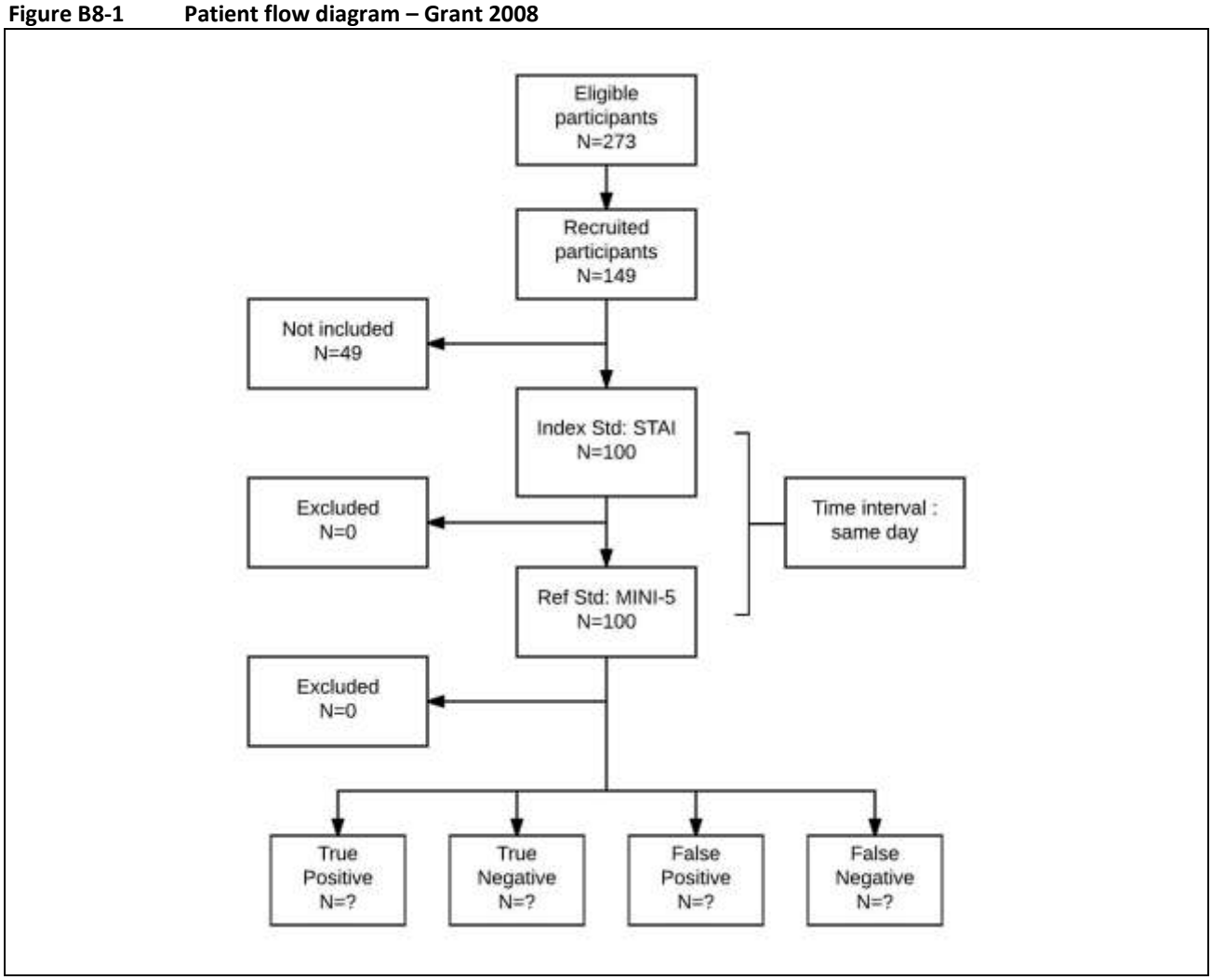
B8.4.2 Checklists for additional screening studies

Table B8-4 Key aspects of the review question

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

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

Figure B8-1 Patient flow diagram – Grant 2008



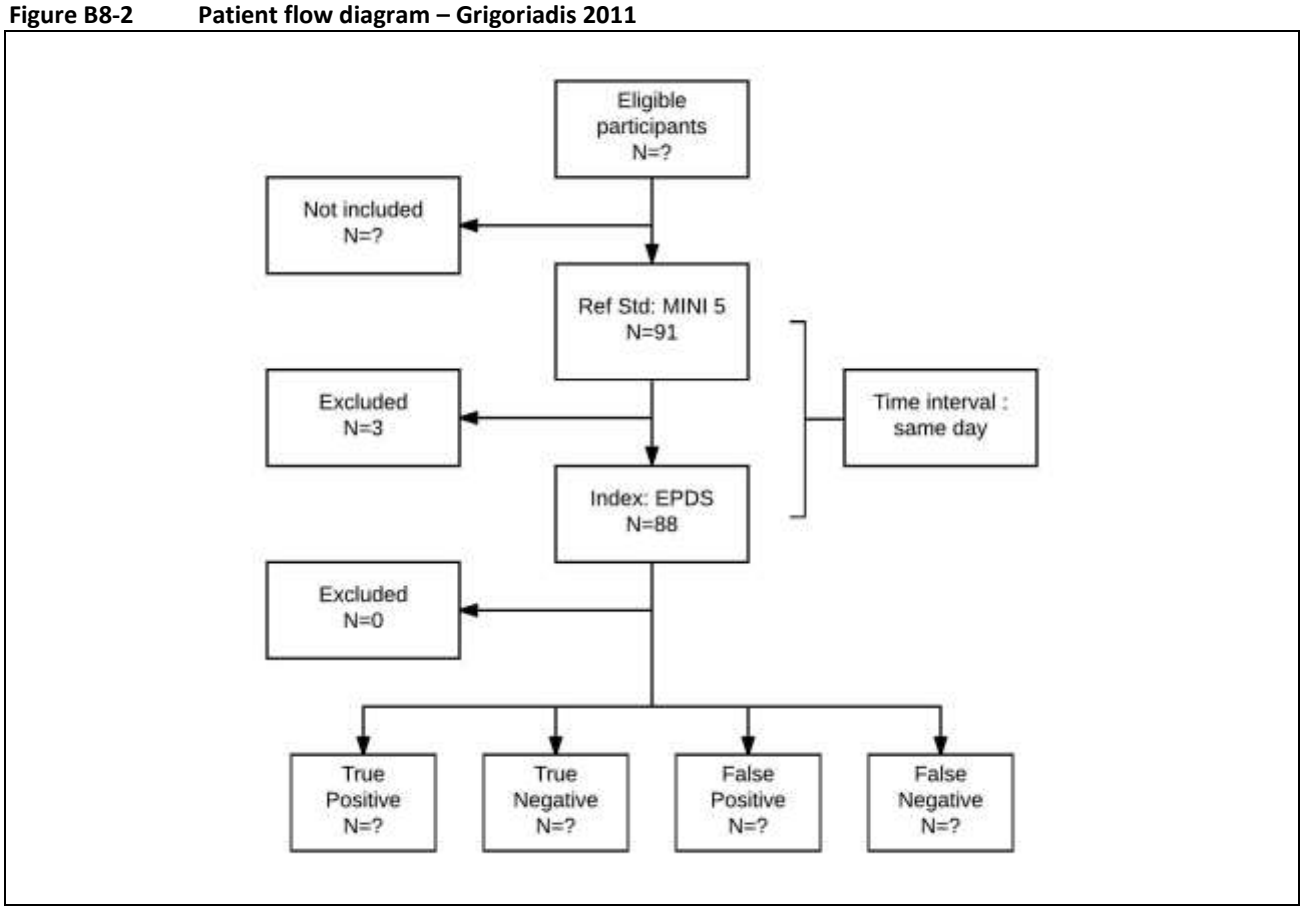
Abbreviations: MINI, Mini-international neuropsychiatric interview; STAI, State-Trait Anxiety Inventory; Std, standard.

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

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

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

Figure B8-2 Patient flow diagram – Grigoriadis 2011



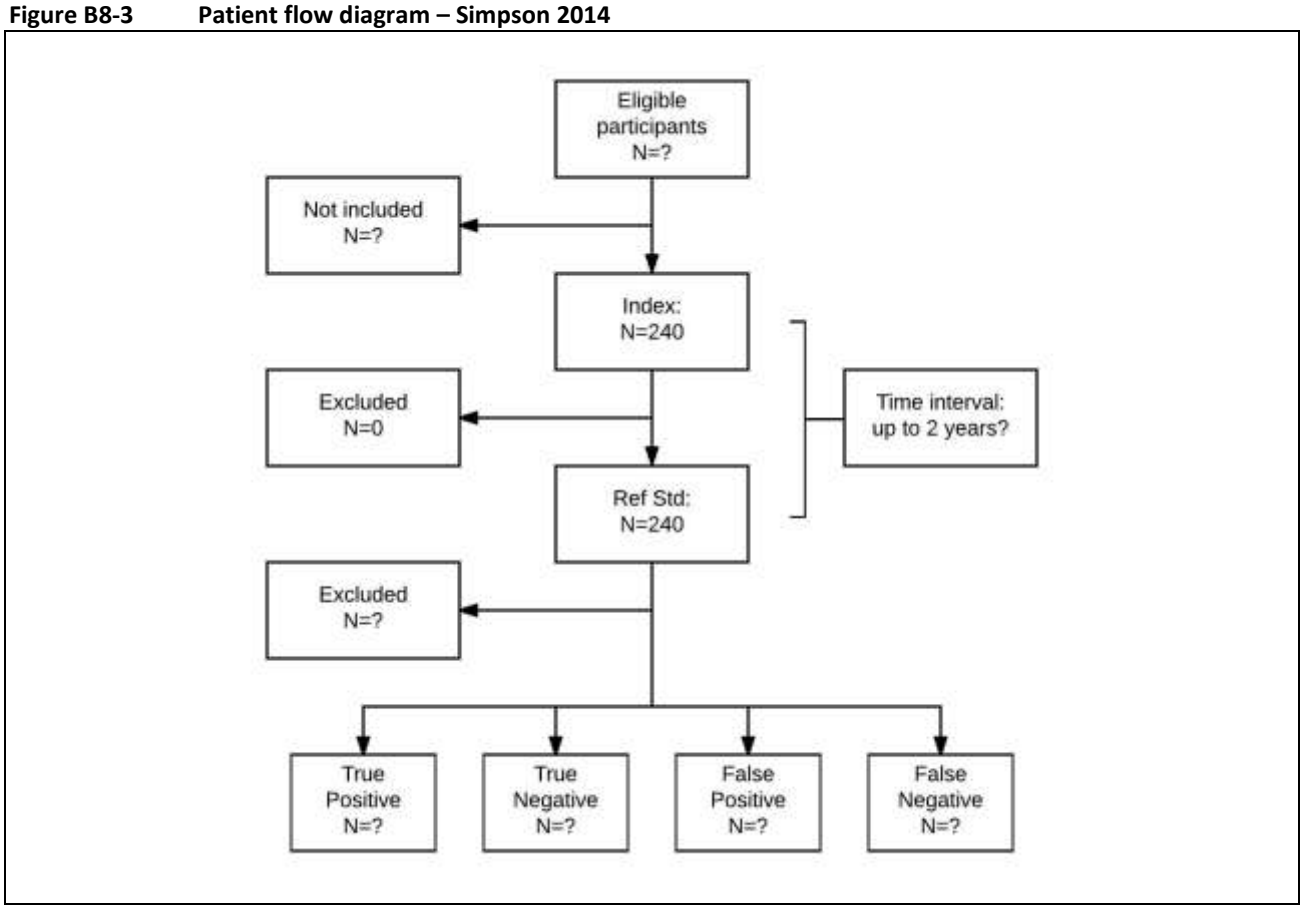
Abbreviations: EPDS, Edinburgh Postnatal Depression Scale; MINI, Mini-international neuropsychiatric interview; Std, standard.

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

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

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

Figure B8-3 Patient flow diagram – Simpson 2014



Abbreviations: Std, standard.

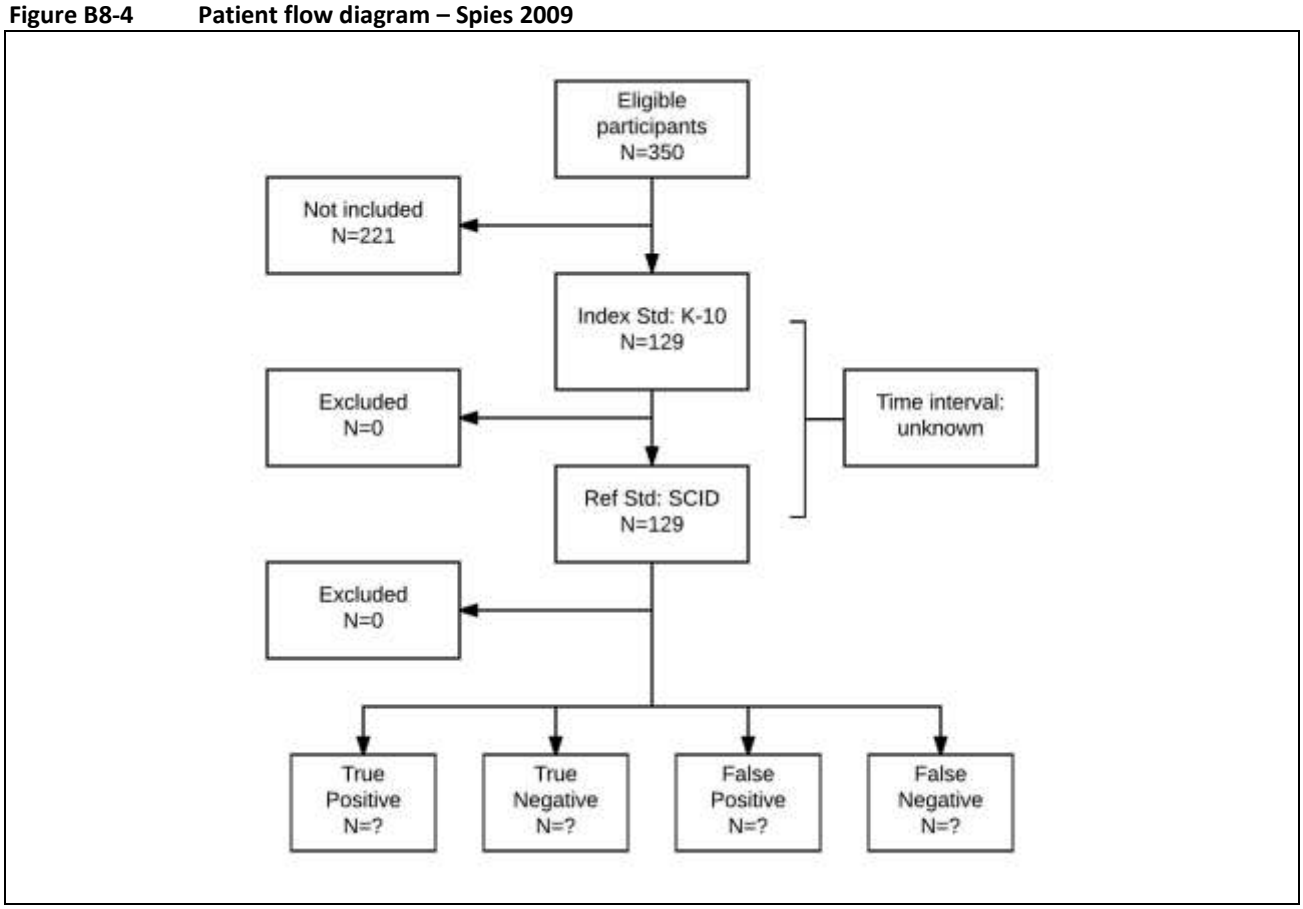
aTable B8-7 Risk of bias and applicability judgments – Simpson 2014

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

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

Abbreviations: EPDS, Edinburgh Postnatal Depression Scale.

Figure B8-4 Patient flow diagram – Spies 2009



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

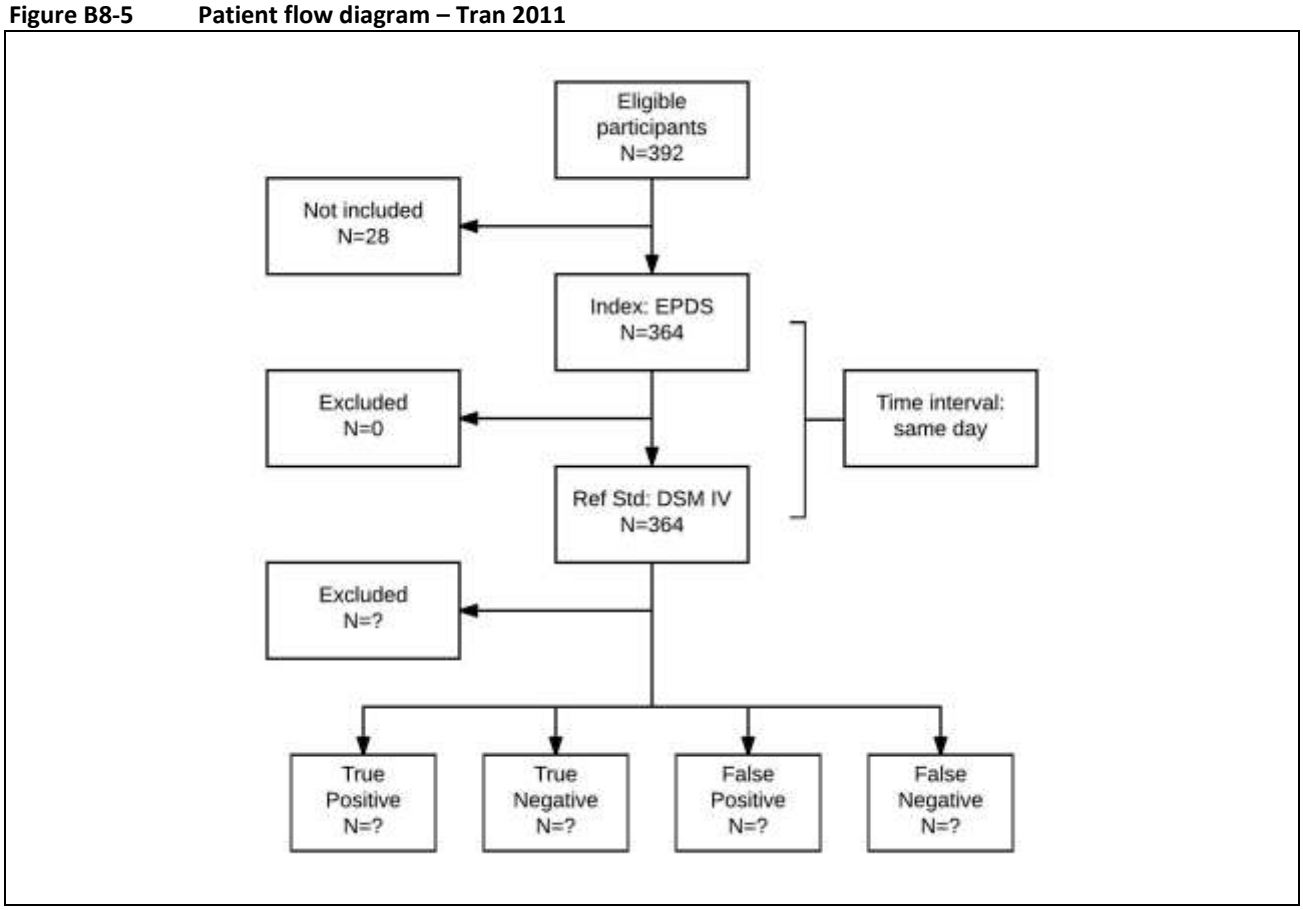
Table B8-8 Risk of bias and applicability judgments – Spies 2009

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

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

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

Figure B8-5 Patient flow diagram – Tran 2011



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

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

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

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

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

Australian Perinatal Mental Health Guideline Evidence Review

Technical Report Part C

Effectiveness of treatment and prevention interventions

Prepared by

hereco
clinical evidence for the real world

June 2017

TABLE OF CONTENTS

C1 INTRODUCTION 1

C2 METHODOLOGY 2

C2.1 Clinical questions 2

C2.2 Criteria for determining study eligibility 3

C2.3 Literature search 5

C2.3.1 Search strategy 5

C2.3.2 Study eligibility 5

C2.4 Assessment of the evidence 6

C2.5 Evidence to recommendations process 6

C2.5.1 Grading of the certainty of the body of evidence 6

C2.5.2 Determining the absolute increase in risk 7

C2.5.3 Drafting of Evidence Statements 7

C3 RESULTS – TREATMENT 8

C3.1 Treatment with psychosocial interventions 8

C3.1.1 Psychoeducation 8

C3.1.2 Psychoeducational booklet 15

C3.1.3 Social/peer support 15

C3.1.4 Home visits 19

C3.1.5 Non-mental-health-focused education/support 22

C3.1.6 Pre-delivery discussion 23

C3.1.7 Post-delivery discussion 24

C3.1.8 Post-miscarriage self-help 24

C3.1.9 Seeing and/or holding stillborn infant 27

C3.1.10 Mother-infant relationship interventions 28

C3.1.11 Co-parenting interventions 36

C3.1.12 Mindfulness 37

C3.2 Treatment with psychological interventions 38

C3.2.1 Structured psychological interventions (CBT or IPT) 38

C3.2.2 Directive counselling 46

C3.2.3 Non-directive counselling 48

C3.2.4 Case management/individual treatment 51

C3.2.5 Self-help and facilitated self-help 52

C3.2.6 Post-traumatic birth counselling 53

C3.2.7 Post-miscarriage counselling 55

C3.3 Treatment with online interventions 57

C3.4 Treatment with pharmacological interventions 58

C3.4.1 Antidepressants 58

C3.4.2 Antipsychotics 69

C3.4.3 Anticonvulsants 69

C3.4.4 Benzodiazepines and z-drugs 69

C3.4.5 Lithium 69

C3.5 Treatment with complementary therapies 70
C3.5.1 Omega-3 fatty acids 70
C3.5.2 St John's wort 71
C3.5.3 Gingko biloba 72

C3.6 Treatment with physical interventions 72
C3.6.1 Exercise 72
C3.6.2 Yoga 74
C3.6.3 Acupuncture 75
C3.6.4 Electroconvulsive therapy 79
C3.6.5 Transcranial magnetic stimulation 79

C4 RESULTS – PREVENTION 80

C4.1 Prevention with psychosocial interventions 80
C4.1.1 Psychoeducation 80
C4.1.2 Psychoeducational booklet 82
C4.1.3 Social/peer support 84
C4.1.4 Home visits 85
C4.1.5 Non-mental-health-focused education/support 89
C4.1.6 Pre-delivery discussion 93
C4.1.7 Post-delivery discussion 94
C4.1.8 Post-miscarriage self-help 95
C4.1.9 Seeing and/or holding stillborn infant 96
C4.1.10 Mother-infant relationship interventions 97
C4.1.11 Co-parenting interventions 100
C4.1.12 Mindfulness 100

C4.2 Prevention with psychological interventions 101
C4.2.1 Structured psychological interventions (CBT or IPT) 101
C4.2.2 Directive counselling 104
C4.2.3 Non-directive counselling 104
C4.2.4 Case management/individual treatment 105
C4.2.5 Self-help and facilitated self-help 106
C4.2.6 Post-traumatic birth counselling 106
C4.2.7 Post-miscarriage counselling 107

C4.3 Prevention with online interventions 107

C4.4 Prevention with pharmacological interventions 109
C4.4.1 Antidepressants 109
C4.4.2 Antipsychotics 112
C4.4.3 Anticonvulsants 112
C4.4.4 Benzodiazepines and z-drugs 112
C4.4.5 Lithium 112

C4.5 Prevention with complementary interventions 113
C4.5.1 Omega-3 fatty acids 113
C4.5.2 St John's wort 116
C4.5.3 Gingko biloba 116

C4.6 Prevention with physical interventions 116
C4.6.1 Exercise 116
C4.6.2 Yoga 117

C4.6.3 Acupuncture 117
C4.6.4 Electroconvulsive therapy 118
C4.6.5 Transcranial magnetic stimulation 118

C4.7 Economic evidence 119

C5 REFERENCES 120

TABLE OF TABLES

Table C2-1 NHMRC Evidence Hierarchy: designation of levels of evidence according to type of research question 3
Table C2-2 PICO criteria used to inform the literature search 4
Table C2-3 Evidence selection criteria 6
Table C3-1 Summary of findings (treatment) – psychologically (CBT/IPT) informed psychoeducation versus treatment as usual or enhanced treatment as usual 10
Table C3-2 Summary of findings (treatment) – IPT-informed psychoeducation versus non-mental health-focused education and support 14
Table C3-3 Summary of findings (treatment) – psychoeducational booklet 15
Table C3-4 Summary of findings (treatment) – social support versus treatment as usual 16
Table C3-5 Summary of findings (treatment) – combined social support and physical exercise versus enhanced treatment as usual 18
Table C3-6 Summary of findings (treatment) – social support versus physical exercise 19
Table C3-7 Summary of findings (treatment) – home visits versus treatment as usual or enhanced treatment as usual 21
Table C3-8 Summary of findings (treatment) – non-mental health-focused education and support versus treatment as usual 23
Table C3-9 Summary of findings (treatment) – pre-delivery discussion 23
Table C3-10 Summary of findings (treatment) – summary of findings: post-delivery discussion 24
Table C3-11 Summary of findings (treatment) – post-miscarriage self-help versus treatment as usual 25
Table C3-12 Summary of findings (treatment) – post-miscarriage facilitated self-help versus treatment as usual 27
Table C3-13 Summary of findings (treatment) – seeing and/or holding stillborn infant 27
Table C3-14 Summary of findings (treatment) – mother-infant relationship interventions versus treatment as usual or enhanced treatment as usual 30
Table C3-15 Summary of findings (treatment) – mother-infant relationship intervention with video feedback versus mother-infant relationship intervention with verbal feedback 35
Table C3-16 Summary of findings (treatment) – mother-infant relationship intervention (and facilitated self-help for eating disorders) versus listening visits (and facilitated self-help for eating disorders) 35
Table C3-17 Summary of findings (treatment) – co-parenting intervention versus enhanced treatment as usual 36
Table C3-18 Summary of findings (treatment) – mindfulness training versus treatment as usual or enhanced treatment as usual 37
Table C3-19 Summary of findings (treatment) – structured psychological interventions (CBT or IPT) versus treatment as usual or enhanced treatment as usual 41
Table C3-20 Summary of findings (treatment) – CBT versus listening visits 45
Table C3-21 Summary of findings (treatment) – IPT versus support group 46
Table C3-22 Summary of findings (treatment) – directive counselling versus treatment as usual 47
Table C3-23 Summary of findings (treatment) – listening visits/non-directive counselling versus treatment as usual 49
Table C3-24 Summary of findings (treatment) – case management/individual treatment 52
Table C3-25 Summary of findings (treatment) – facilitated self-help versus treatment as usual 52
Table C3-26 Summary of findings (treatment) – post-traumatic birth counseling versus treatment as usual 54
Table C3-27 Summary of findings (treatment) – post-miscarriage counselling versus treatment as usual 56
Table C3-28 Summary of findings (treatment) – online interventions 57
Table C3-29 Summary of findings (treatment) – antidepressants versus general supportive care 58
Table C3-30 Summary of findings (treatment) – antidepressants versus listening visits 60
Table C3-31 Summary of findings (treatment) – SSRIs versus placebo 62
Table C3-32 Summary of findings (treatment) – SSRIs + psychological interventions versus placebo + psychological interventions 65
Table C3-33 Summary of findings (treatment) – SSRIs versus TCAs 67
Table C3-34 Summary of findings (treatment) – antipsychotics 69
Table C3-35 Summary of findings (treatment) – anticonvulsants 69
Table C3-36 Summary of findings (treatment) – benzodiazepines and z-drugs 69
Table C3-37 Summary of findings (treatment) – lithium 70
Table C3-38 Summary of findings (treatment) – omega-3 fatty acids versus placebo 70
Table C3-39 Summary of findings (treatment) – St John's wort 72
Table C3-40 Summary of findings (treatment) – gingko biloba 72
Table C3-41 Summary of findings (treatment) – physical activity versus treatment as usual 73
Table C3-42 Summary of findings (treatment) – physical activity versus mutual support 74
Table C3-43 Summary of findings (treatment) – yoga versus control group 75
Table C3-44 Summary of findings (treatment) – acupuncture versus massage 76

Table C3-45 Summary of findings (treatment) – depression-specific acupuncture versus non-depression-specific acupuncture 77
Table C3-46 Summary of findings (treatment) – electroacupuncture versus non-invasive sham acupuncture 78
Table C3-47 Summary of findings (treatment) – electroconvulsive therapy 79
Table C3-48 Summary of findings (treatment) – transcranial magnetic stimulation 79
Table C4-1 Summary of findings (prevention) – psychologically (CBT/IPT) informed psychoeducation versus treatment as usual or enhanced treatment as usual 81
Table C4-2 Summary of findings (prevention) – psychoeducational booklet versus treatment as usual or enhanced treatment as usual 83
Table C4-3 Summary of findings (prevention) – social support versus treatment as usual 85
Table C4-4 Summary of findings (prevention) – home visits versus treatment as usual 87
Table C4-5 Summary of findings (prevention) – non-mental health-focused education and support versus treatment as usual or enhanced treatment as usual 90
Table C4-6 Summary of findings (prevention) – pre-delivery discussion 93
Table C4-7 Summary of findings (prevention) – post-delivery discussion versus enhanced treatment as usual 94
Table C4-8 Summary of findings (prevention) – post-miscarriage self-help versus treatment as usual 95
Table C4-9 Summary of findings (prevention) – seeing and/or holding stillborn infant versus not seeing and/or holding stillborn infant 97
Table C4-10 Summary of findings (prevention) – mother-infant relationship interventions versus treatment as usual 98
Table C4-11 Summary of findings (prevention) – co-parenting interventions 100
Table C4-12 Summary of findings (prevention) – mindfulness 101
Table C4-13 Summary of findings (prevention) – structured psychological interventions versus usual care 103
Table C4-14 Summary of findings (prevention) – directive counselling 104
Table C4-15 Summary of findings (prevention) – non-directive counselling 104
Table C4-16 Summary of findings (prevention) – case-management and individualised treatment versus treatment as usual 105
Table C4-17 Summary of findings (prevention) – self-help 106
Table C4-18 Summary of findings (prevention) – post-traumatic birth counselling 106
Table C4-19 Summary of findings (prevention) – post-miscarriage counselling 107
Table C4-20 Summary of findings (prevention) – online intervention versus offline (face-to-face) intervention 108
Table C4-21 Summary of findings (prevention) – SSRIs (sertraline) versus placebo 109
Table C4-22 Summary of findings (prevention) – TCAs (nortriptyline) versus placebo 110
Table C4-23 Summary of findings (prevention) – antipsychotics 112
Table C4-24 Summary of findings (prevention) – anticonvulsants 112
Table C4-25 Summary of findings (prevention) – benzodiazepines and z-drugs 112
Table C4-26 Summary of findings (prevention) – lithium 113
Table C4-27 Summary of findings (prevention) – omega-3 fatty acids (EPA) versus placebo 114
Table C4-28 Summary of findings (prevention) – omega-3 fatty acid (DHA) versus placebo 115
Table C4-29 Summary of findings (prevention) – St John's wort 116
Table C4-30 Summary of findings (prevention) – gingko biloba 116
Table C4-31 Summary of findings (prevention) – exercise 117
Table C4-32 Summary of findings (prevention) – yoga 117
Table C4-33 Summary of findings (prevention) – acupuncture versus placebo acupuncture 118
Table C4-34 Summary of findings (prevention) – electroconvulsive therapy 118
Table C4-35 Summary of findings (prevention) – transcranial magnetic stimulation 118

ABBREVIATIONS

ASQ:SE	Ages and Stages Questionnaire: Social-Emotional
AQS	Attachment Q Set
BAI	Beck Anxiety Inventory
BDI	Beck Depression Inventory
BSI	Brief Symptom Inventory
CES-D	Center for Epidemiologic Studies Depression Scale
CBCL	Child Behaviour Checklist
CGI-I	Clinical Global Impression – Improvement scale
CGI	Clinical Global Impression scale
CORE-OM	Clinical Outcomes in Routine Evaluation-Outcome Measure
CBT	cognitive behaviour therapy
CIS-R	Computerised version of the Clinical Interview Schedule – Revised
CI	confidence interval
CAGE	cut-annoyed-guilty-eye
DASS	Depression Anxiety Stress Scale
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders, 4th Edition
EPDS	Edinburgh Postnatal Depression Scale
EAS	Emotional Availability Scales
GHQ	General Health Questionnaire
GAD-7	Generalised Anxiety Disorder Assessment
HRSD	Hamilton Rating Scale for Depression
HSCL	Hopkins Symptom Checklist
HADS	Hospital Anxiety and Depression Scale
IES	Impact of Events Scale
IES-R	Impact of Events Scale – Revised
ES-R	Impact of Events Scale – Revised
ICEP	Infant and Caregiver Engagement Phases
ITT	intention-to-treat
IPT	interpersonal psychotherapy
IPT	interpersonal therapy
LOCF	last observation carried forward
LIFE	Longitudinal Interval Follow-up Examination
MDD	major depressive disorder
MMS	Maternal Mood Screener
MD	mean difference
MINI	Mini International Neuropsychiatric Interview
MADRS	Montgomery–Åsberg Depression Rating Scale
NICU	neonatal intensive care unit
NE	not estimable (from data in SR)
NR	not reported
NCAST	Nursing Child Assessment Satellite Training Scale
OCD	obsessive-compulsive disorder
OR	odds ratio
PIRGAS	Parent-Infant Relationship Global Assessment Scale
PSCS	Parenting Sense of Competence Scale
PSS	Perceived Stress Scale
PPQ	Perinatal PTSD Questionnaire
PICS	Pictoral Infant Communication Scales
PND	postnatal depression
PTSD	post-traumatic stress disorder
QIDS	Quick Inventory of Depressive Symptoms
RCT	randomised controlled trial
RR	relative risk
RDC	Research Diagnostic Criteria
SADS	Schedule for Affective Disorders and Schizophrenia
SCAN	Schedules for Clinical Assessment in Neuropsychiatry
SSRI	selective serotonin reuptake inhibitor
SF	Short Form Health Survey
STSI	Short Temperament Scale for Infants

SD	standard deviation
SMD	standardised mean difference
STAI	State-Trait Anxiety Inventory
STAI-S	State-Trait Anxiety Inventory-State
STAI-T	State-Trait Anxiety Inventory-Trait
KID-SCID	Structured Clinical Interview for Childhood Diagnoses
SCID	Structured Clinical Interview for DSM Disorders
SR	systematic review
TAU	treatment as usual
VAS	visual analogue scale
WHO	World Health Organization
WCS	worst case scenario
YBOCS	Yale–Brown Obsessive Compulsive Scale

C1 INTRODUCTION

The aim of this Evidence Review is to assess the evidence relating to the identification and treatment or prevention of mental health problems in women during pregnancy or the postnatal period. The following Technical Reports and associated Appendices are related to this assessment:

- Part B Technical Report and Part B Appendix – Psychosocial Assessment and Screening
- Part C Technical Report and Part C Appendix – Effectiveness of Treatment and Prevention
- Part D Technical Report and Part D Appendix – Harm

This Technical Report and associated Appendix (Part C) present the findings of the assessment of evidence of the effectiveness of *interventions* used for the treatment and prevention of mental health problems in women during the antenatal or postnatal period.

C2 METHODOLOGY

C2.1 CLINICAL QUESTIONS

The two main questions relating to the effectiveness of interventions for the treatment of mental health problems in pregnant or postpartum women, or prevention of mental health problems in pregnant or postpartum women identified as being at risk of developing mental health problems, were each broken down into five sub-questions based on different intervention types. Each sub-question is also broken down further into individual interventions and outcomes; the detailed definitions associated with these interventions and outcomes can be found in **Section C2.2.** All questions were addressed via systematic review.

It should be noted that while the side effects of treatment experienced by the mother were captured in this part of the Technical Report (Part C), the harms to the fetus, infant, child or mother are assessed in **Part D** of the **Technical Report**.

C2.1.1.1 Treatment interventions

Main question:

4. What is the effectiveness of interventions for the treatment of mental health problems in women in the antenatal or postnatal period?

Sub-questions:

4a. What is the effectiveness of psychosocial interventions for the treatment of mental health problems in women in the antenatal or postnatal period?

4b. What is the effectiveness of psychological interventions for the treatment of mental health problems in women in the antenatal or postnatal period?

4c. What is the effectiveness of pharmacological interventions for the treatment of mental health problems in women in the antenatal or postnatal period?

4d. What is the effectiveness of complementary interventions for the treatment of mental health problems in women in the antenatal or postnatal period?

4e. What is the effectiveness of physical interventions for the treatment of mental health problems in women in the antenatal or postnatal period?

C2.1.1.2 Prevention interventions

Main question:

5. What is the effectiveness of interventions for the prevention of mental health problems in women identified as being at risk of developing a mental health problem in the antenatal or postnatal period?

Sub-questions:

5a. What is the effectiveness of psychosocial interventions for the prevention of mental health problems in women identified as being at risk of developing a mental health problem in the antenatal or postnatal period?

5b. What is the effectiveness of psychological interventions for the prevention of mental health problems in women identified as being at risk of developing a mental health problem in the antenatal or postnatal period?

5c. What is the effectiveness of pharmacological interventions for the prevention of mental health problems in women identified as being at risk of developing a mental health problem in the antenatal or postnatal period?

5d. What is the effectiveness of complementary interventions for the prevention of mental health problems in women identified as being at risk of developing a mental health problem in the antenatal or postnatal period?

5e. What is the effectiveness of physical interventions for the prevention of mental health problems in women identified as being at risk of developing a mental health problem in the antenatal or postnatal period?

C2.2 CRITERIA FOR DETERMINING STUDY ELIGIBILITY

To determine the effectiveness of an intervention, a systematic review (SR) of randomised controlled trials (RCTs) provides the highest level of evidence, as shown in the evidence hierarchy for examination of intervention questions (**Table C2-1**).

For each of the intervention-based questions to be assessed by the Evidence Review (effectiveness of treatment and prevention, and harm), the Expert Working Group (EWG) agreed to the appropriate level of evidence for inclusion. For the review of effectiveness of treatment and prevention using psychosocial, psychological and most physical interventions, the EWG agreed that Level I evidence (SR of RCTs) should be used as the basis of the review, with preference given to those SRs that used a GRADE approach for assessment of the evidence. For online, pharmacological, complementary and selected physical interventions (electroconvulsive therapy [ECT] and transcranial magnetic stimulation [TMS]), Level II evidence was to be included if Level I evidence was unavailable or out of date (i.e. pre-2014).

Table C2-1 NHMRC Evidence Hierarchy: designation of levels of evidence according to type of research question[1]

Level	Intervention
I	A systematic review of level II studies
II	A randomised controlled trial
III-1	A pseudorandomised controlled trial (i.e. alternate allocation or some other method)
III-2	A comparative study with concurrent controls: • Non-randomised, experimental trial[2] • Cohort study • Case-control study • Interrupted time series with a control group
III-3	A comparative study without concurrent controls: • Historical control study • Two or more single arm studies[3] • Interrupted time series without a parallel control group
IV	Case series with either post-test or pre-test/post-test outcomes

Table C2-2 summarises the criteria used to determine study eligibility. The populations of interest for interventions that treat and/or prevent mental health problems are pregnant and postpartum women. A comprehensive range of interventions were reviewed under the categories of psychosocial, psychological, online, pharmacological, complementary and physical. Eligible comparators included treatment as usual,

[1] NHRMC (2009) NHMRC additional levels of evidence and grades for recommendations for developers of guidelines. Accessed on 12 May 2017 from https://www.nhmrc.gov.au/_files_nhmrc/file/guidelines/developers/nhmrc_levels_grades_evidence_120423.pdf.

[2] This also includes controlled before-and-after (pre-test/post-test) studies, as well as adjusted indirect comparisons (ie. utilise A vs B and B vs C, to determine A vs C with statistical adjustment for B).

[3] Comparing single arm studies ie. case series from two studies. This would also include unadjusted indirect comparisons (ie. utilise A vs B and B vs C, to determine A vs C but where there is no statistical adjustment for B).

enhanced treatment as usual, no treatment/placebo, waitlist control, or other active interventions. The included outcomes were:

- Maternal mental health outcomes – antenatal or postnatal development of, or change in, a mental health diagnosis or symptomatology using validated instruments.
- Mother-infant interactions – postnatal assessment of attachment problems, positive interactions, and maternal sensitivity, using validated instruments.
- Safety – in terms of side effects to the mother.

Harms due to pharmacological, complementary or physical interventions are assessed in **Part D** of the **Technical Report**.

Table C2-2 PICO criteria used to inform the literature search

Populations	Interventions	Comparators	Outcomes
Pregnant women Post-partum women	Psychosocial interventions Psychoeducation Psychoeducational booklet Social/peer support Home visits Non-mental-health-focused education/support Pre-delivery discussion Post-delivery discussion Post-miscarriage self-help Seeing and/or holding stillborn infant Mother-infant relationship interventions Co-parenting interventions Mindfulness Psychological interventions Structured psychological interventions (CBT and IPT) Directive counselling Non-directive counselling Case management/individualised treatment Self-help or facilitated self-help Post-traumatic birth counselling Post-miscarriage counselling Online interventions Pharmacological interventions Antidepressants Antipsychotics Mood stabilisers (including anticonvulsants,[4] benzodiazepines and z-drugs) Lithium Complementary interventions Omega-3 fatty acids St John's wort Gingko biloba Physical interventions Exercise Yoga Acupuncture Electroconvulsive therapy Transcranial magnetic stimulation	Treatment as usual, enhanced treatment as usual, no treatment/placebo, or waitlist control Other active interventions	Maternal mental health symptomatology or diagnosis Depression/anxiety/PTSD diagnosis Depression/anxiety/PTSD symptomatology Negative thoughts/mood Mother-infant interactions Mother-infant attachment problems Positive mother-infant interaction Maternal sensitivity Safety Side effects

Abbreviations: CBT, cognitive behaviour therapy; IPT, interpersonal psychotherapy; PTSD, post-traumatic stress disorder.

[4] Sodium valproate, carbamazepine and lamotrigine only.

C2.3 LITERATURE SEARCH

C2.3.1 Search strategy

A two-tiered search strategy was undertaken. An initial search was undertaken to identify SRs that assessed various treatments for the main mental health problems seen during the perinatal period; these included depression, anxiety, schizophrenia and bipolar disorder. Full details of the SR search can be found in **Appendix C1.1.1**. It should be noted that this search was conducted to identify studies not only for the assessment of efficacy for treatment and prevention, but also for screening interventions, and the harms associated with pharmacological, complementary and selected physical interventions.

From this search, a list was compiled of SRs that assessed the effectiveness of various interventions for the treatment and prevention of mental health problems during the perinatal period; evidence regarding harms was also identified by this search. The individual studies included in each SR were tabulated and, where possible, a 'foundation review' was selected. The process for identifying the foundation reviews is outlined in **Appendix C2** (treatment) and **Appendix C3** (prevention). The foundation review was defined as the SR that included the most recent and comprehensive set of data for a particular intervention and outcome; SRs that assessed the body of evidence and produced a Summary of Findings table using GRADE methodology were preferentially included where available.

Based on the findings of the SR search, a second series of literature searches were carried out for the online, pharmacological, complementary and selected physical interventions. These 'updated' searches were also aimed at identifying individual studies for the consideration of harms. Full details of the updated searches can be found in **Appendix C1.1.2**.

Searches were conducted in the MEDLINE, Embase and PsychINFO databases (via the OVID and/or Embase.com interfaces), the Cochrane Library, and included examination of the reference lists of included SRs.

C2.3.2 Study eligibility

The aim of the literature search was to identify the highest possible quality evidence for each intervention/outcome. As noted previously, SRs of RCTs provide the highest level of evidence for assessment of the effects of interventions, and for the assessment of treatment and prevention effectiveness, eligibility was limited to this study type for the assessment of psychosocial and psychological interventions, most physical interventions, anticonvulsants and omega-3 fatty acids; individual RCTs were eligible for inclusion for pharmacological, other complementary, and electroconvulsive therapy and transcranial magnetic stimulation.

Citations identified in the literature searches were reviewed and evidence selection criteria were applied hierarchically.

As shown in **Table C2 3**, there was a set of standard inclusion/exclusion criteria that applied to both the SR search for all interventions, and the updated searches for online, pharmacological, complementary and the physical (ECT and TMS) interventions.

Details of the assessment of study eligibility can be found in **Part C Appendix Section C1.2**.

Table C2-3 Evidence selection criteria

Criterion	Description
SR search	
Not a SR	Excludes individual clinical studies, narrative reviews, editorials, animal studies and in vitro studies
Wrong population	Excludes studies that are not conducted in pregnant or postnatal women, or children exposed to intervention antenatally or postnatally
Wrong intervention/exposure	Excludes studies that do not examine one of the exposures included in Table C2-2
Wrong outcome	Excludes studies that do not examine one of the outcomes included in Table C2-2
Not in English	Excludes SRs not available in English.
Updated searches – all	
Not a clinical study	Excludes narrative reviews, editorials, animal studies and in vitro studies
Not a SR	Excludes reviews described as systematic that are not, or that limit identification of evidence to MEDLINE/PubMed only.
Wrong population	Excludes studies that are not conducted in pregnant or postnatal women (or children exposed to intervention antenatally or postnatally for the assessment of infant harms)
Wrong intervention/exposure	Excludes studies that do not examine one of the interventions included in Table C2-2
Wrong/no comparator	Excludes studies that do not compare the intervention with no intervention or a relevant active intervention
Wrong outcome	Excludes studies that do not examine one of the outcomes included in Table C2-2
Wrong study type	Excludes RCTs where SR evidence is available and all observational studies (assessment of effectiveness using individual studies limited to RCT evidence only)
Protocol only	Excludes publications describing a study protocol only
Duplicate data	Excludes studies that include data that has already been included from another publication
Not in English	Excludes studies not available in English
Abstract only	Excludes studies available as a conference abstract only. Where identified, an additional search is conducted to see if the study was subsequently published.

Abbreviations: SR, systematic review.

C2.4 Assessment of the evidence

Grading of Recommendations, Assessment, Development and Evaluations (GRADE) methodology has been used (either within the identified foundation SR or performed specifically for this Evidence Review) to grade the quality of the evidence for each intervention and outcome and translate this into recommendations and practice points. For further details about GRADE see http://www.gradeworkinggroup.org/.

According to the GRADE process, the body of evidence is summarised in either an Evidence Profile (EP) table or Summary of Findings (SoF) table. For the purpose of the assessment of effectiveness of treatment and prevention, the evidence is presented in SoF tables.

C2.5 Evidence to recommendations process

C2.5.1 Grading of the certainty of the body of evidence

Assessing the certainty of a body of evidence using GRADE involves consideration of the following five domains: risk of bias, inconsistency, indirectness, imprecision and publication bias. For an evidence base drawn from RCTs, the grading of the certainty of the body of evidence starts at 'high' (⊕⊕⊕⊕) and can be downgraded for each of the five domains depending on whether the limitation is considered serious (downgrade one level) or very serious (downgrade two levels).

For this Evidence Review, the aim of the assessment of effectiveness was to identify existing SRs that had assessed and presented the body of evidence using GRADE methodology. Where this was available, the assessment performed by the SR was accepted, and the SoF table was taken directly from the SR and presented in this Technical Report. Where this was not available, relevant SRs or original studies were identified and the body of evidence was assessed using GRADE methodology and presented in SoF or EP tables.

C2.5.2 Determining the absolute increase in risk

The absolute increase in risk can be calculated for dichotomous outcomes that are reported as risk ratios (RRs). The baseline risk generally comes from the control group of the SR of RCTs, and is multiplied by the RR and 95% confidence interval (CI) to determine the increase or decrease in risk associated with an intervention.

Where an existing SoF table was available, the absolute increase in risk reported in the table was used. Where an SoF table was developed de novo, the absolute increase in risk was calculated.

C2.5.3 Drafting of Evidence Statements

Evidence Statements have been derived from the data presented in the SoF tables. Although Evidence Statements are not a requirement for GRADE, it was agreed that describing the data in words is a useful bridge from the SoF tables to the Recommendations and Practice Points. All of the summary findings presented below are based on data from RCTs. Consequently, the following general 'phrasing rules' have been applied to the Evidence Statements:

- Where there is a statistically significant effect, and the quality of the evidence has been rated 'high' or 'moderate', the phrasing "improves [outcome]" has been used.
- Where there is a statistically significant effect, and the quality of the evidence has been rated 'low' or 'very low', the phrasing "may improve [outcome]" has been used.
- Where no statistically significant effect is observed, and the quality of the evidence has been rated 'high' or 'moderate', the phrasing "has no effect on [outcome]" has been used.
- Where no statistically significant effect is observed, and the quality of the evidence has been rated 'low' or 'very low', the phrasing "appears to have no effect on [outcome]" has been used.

In some circumstances, where there is a large but not statistically significant effect, and the quality of the evidence has been rated 'high' or 'moderate', the phrasing "may have an effect on [outcome]" is used. Likewise, in some circumstances, where there is a large effect that is close to being statistically significant (P=0.05-1.0), irrespective of the quality of the evidence, the phrasing "may have an effect on [outcome]" is used.

C3 RESULTS – TREATMENT

For the majority of psychosocial, psychological, physical and pharmacological interventions, the National Institute for Health and Care Excellence Clinical Guideline Number 192 (NICE 2015) was chosen as the foundation review, primarily due to its currency, comprehensiveness, and high quality. The EWG agreed that the current Review would reproduce the Summary of Findings (SoF) tables from NICE 2015, and that replication of data extraction tables, risk-of-rias assessment, or Evidence Profile (EP) tables was not required. Readers are referred to the NICE guideline CG192 and appendices for this information.

NICE 2015 did not specifically assess evidence relating to online (web-based or computer-based) interventions. As such, the current Evidence Review relied on other published SRs, together with a literature search update to identify recent RCTs not included in the published reviews. For all relevant RCTs relating to online interventions, a full assessment of the evidence was required, including data extraction, risk-of-bias assessment, and EP tables; these are available in the **Part C Appendix**. Evidence from additional SRs was used for yoga (treatment) and antidepressants (treatment); the quality of the SR was assessed, the SR evidence was transcribed directly into an SoF table, and the certainty of the body of evidence was informed by the SR assessment of risk of bias of the individual studies.

C3.1 TREATMENT WITH PSYCHOSOCIAL INTERVENTIONS

C3.1.1 Psychoeducation

Only one SR (NICE 2015) was identified for psychoeducation in the literature search. NICE 2015 included a total of 17 individual RCTs, two of which were from Australia. Sixteen RCTs (N=2,955) compared psychologically informed psychoeducation with treatment as usual or enhanced treatment as usual[5]. Psychoeducation was cognitive behaviour therapy (CBT)-informed in 10 RCTs, IPT-informed in five RCTs, and CBT- and IPT-informed in one RCT. In one of these RCTs the intervention was aimed at women with subthreshold symptoms of OCD, and in the remaining 15 RCTs the intervention was intended for women with symptoms (or subthreshold symptoms) of depression. The timing of the intervention was antenatal in four RCTs, postnatal in four RCTs, and both antenatal and postnatal in eight RCTs. In all RCTs the intervention was delivered face-to-face but one study also involved one postnatal telephone follow-up.

NICE 2015 also included one RCT from the United States (N=38) that compared face-to-face IPT-informed high-intensity group psychoeducation with an active intervention (non-mental health-focused education and support group) in pregnant women with a diagnosis of MDD.

C3.1.1.1 Psychologically (CBT/IPT) informed psychoeducation versus treatment as usual or enhanced treatment as usual

NICE 2015 did not consider the timing of the intervention, format or mode of delivery in their analyses.

[5] Enhanced treatment as usual was a psychoeducational booklet (two RCTs), non-mental-health-focused education and support group (two RCTs), and a psychoeducation group without the CBT component (one RCT). In some cases, these comparators could be considered to be active interventions.

Evidence from up to eight studies showed moderate effects of psychoeducation on depression diagnosis at endpoint using an ITT approach; however, the confidence in these effect estimates is very low due to imprecision (**Table C3-1**). There was also high-certainty evidence from five studies for small-to-moderate, statistically significant benefits associated with psychoeducation on depression symptomatology. However, treatment effects on mean depression scores at endpoint, while statistically significant in many cases, failed to reach the threshold for clinically significant benefits, either at endpoint or at short- or long-term follow up. There was also no evidence for any statistically or clinically significant treatment effects for any outcome measures at intermediate follow up or for depression diagnosis at long-term follow up (using an ITT approach).

There was no evidence for statistically or clinically significant benefits of psychologically informed psychoeducation for anxiety diagnosis at endpoint or at long-term follow-up in women with symptoms (or subthreshold symptoms) of depression. Furthermore, there was no evidence for statistically significant benefits associated with psychoeducation for any of the PTSD outcome measures; the very low certainty of evidence due to risk-of-bias concerns, very serious imprecision and selective outcome reporting prohibits any clear conclusions being drawn from the evidence.

There was very low quality single-study evidence for delayed but statistically significant moderate-to-large effects of psychoeducation on mean OCD symptoms at intermediate and long-term followup, with statistically and clinically non-significant effects at endpoint.

Thirteen studies (N=2,375) found no evidence for clinically or statistically significant effects of psychologically (CBT/IPT) informed psychoeducational interventions relative to treatment as usual or enhanced treatment as usual on attrition (as a proxy measure for retention in services and treatment acceptability).

Table C3-1 Summary of findings (treatment) – psychologically (CBT/IPT) informed psychoeducation versus treatment as usual or enhanced treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Quality of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression diagnosis					
Post-treatment – ITT analysis MINI, Schedule for Affective Disorders and Schizophrenia (SADS), Maternal Mood Screener (MMS), SCID or Longitudinal Interval Follow-up Examination (LIFE) (4-52 weeks)	*Study population*		RR 0.67 (0.41, 1.08)	985 (8 studies)	●○○○ Very low (a,b,c)
	163 per 1000	109 per 1000 (67, 176)			
	Moderate				
	239 per 1000	160 per 1000 (98, 258)			
Post-treatment – available case analysis SADS, MMS or SCID or LIFE (4-52 weeks)	*Study population*		RR 0.50 (0.22, 1.14)	464 (6 studies)	●○○○ Very low (a,b,c,d)
	170 per 1000	71 per 1000 (-31, 180)			
	Moderate				
	219 per 1000	92 per 1000 (-39, 232)			
Intermediate follow-up (17-24 weeks post-intervention) – ITT analysis MINI, SADS or MMS (6-36 weeks)	*Study population*		RR 1.1 (0.75, 1.6)	734 (4 studies)	●○○○ Very low (a,b,c,f)
	113 per 1000	125 per 1000 (85, 181)			
	Moderate				
	86 per 1000	95 per 1000 (65, 138)			
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis SADS or MMS (26-36 weeks)	*Study population*		RR 1.1 (0.58, 2.09)	233 (2 studies)	●○○○ Very low (a,b,c)
	128 per 1000	141 per 1000 (74, 268)			
	Moderate				
	77 per 1000	85 per 1000 (45, 161)			
Long Follow-up (25-103 weeks post-intervention) – ITT analysis MINI, SADS, MMS or SCID (32-75 weeks)	*Study population*		RR 0.8 (0.56, 1.13)	812 (5 studies)	●○○○ Very low (a,b,c)
	217 per 1000	173 per 1000 (121, 245)			
	Moderate				
	250 per 1000	200 per 1000 (140, 282)			
Long Follow-up (25-103 weeks post-intervention) – available case analysis SADS, MMS or SCID (32-75 weeks)	*Study population*		RR 0.6 (0.36, 1.03)	266 (3 studies)	●○○○ Very low (a,b,c)
	227 per 1000	136 per 1000 (82, 233)			
	Moderate				
	250 per 1000	150 per 1000 (90, 257)			
Depression symptomatology					
Post-treatment – ITT analysis HSCL: Sum/20 >0.75 depression, EPDS≥13, Leverton Questionnaire (Elliott et al., 2000) ≥12, QIDS ≥11 or BDI: Treatment non-response (4-26 weeks)	*Study population*		**RR 0.74 (0.62, 0.88)**	1,518 (5 studies)	●●●● High
	351 per 1000	260 per 1000 (218, 309)			
	Moderate				
	480 per 1000	355 per 1000 (298, 422)			
Post-treatment – available case analysis HSCL: Sum/20 >0.75 depression, QIDS≥11 or BDI: Treatment non-response (4-26 weeks)	*Study population*		**RR 0.82 (0.68, 0.98)**	997 (3 studies)	●●●○ Moderate (a)
	320 per 1000	262 per 1000 (218, 314)			
	Moderate				
	458 per 1000	376 per 1000 (311, 449)			
Depression mean scores					
Post-treatment – ITT analysis EPDS or CES-D (4-31 weeks)	SMD -0.25 (-0.58, 0.08)		-	436 (4 studies)	●●●○ Moderate (d)
Post-treatment – available case analysis BDI-II, BDI, EPDS or CES-D (4-31 weeks)	**SMD -0.26 (-0.48, -0.05)**		-	351 (7 studies)	●●●○ Moderate (e)

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Quality of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Short Follow-up (9-16 weeks post-intervention) – ITT analysis EPDS (13-27 weeks)	**SMD -0.37 (-0.63, -0.11)**		-	235 (2 studies)	●●●○ Moderate (e)
Short Follow-up (9-16 weeks post-intervention) – available case analysis EPDS or BDI-II (19-27 weeks)	**SMD -0.42 (-0.82, -0.02)**		-	100 (2 studies)	●○○○ Very low (c,e)
Intermediate follow-up (17-24 weeks post-intervention) – ITT analysis EPDS (26-36 weeks)	SMD -0.07 (-0.35, 0.21)		-	197 (2 studies)	●●○○ Low (e)
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis EPDS (mean 36 weeks)	SMD -0.28 (-0.89, 0.34)		-	41 (1 study)	●○○○ Very low (b,c,e,f)
Long Follow-up (25-103 weeks post-intervention) – ITT analysis EPDS (57-75 weeks)	SMD -0.43 (-0.86, 0)		-	86 (2 studies)	●●○○ Low (e)
Long Follow-up (25-103 weeks post-intervention) – available case analysis EPDS or BDI-II (32-75 weeks)	**SMD -0.44 (-0.75, -0.12)**		-	161 (3 studies)	●○○○ Very low (c,e)
Anxiety diagnosis					
Post-treatment – ITT analysis MINI or SADS (9-52 weeks)	*Study population*		RR 0.97 (0.61, 1.54)	476 (2 studies)	●○○○ Very low (a,b,c,f)
	136 per 1000	132 per 1000 (83, 209)			
	Moderate				
	138 per 1000	134 per 1000 (84, 213)			
Post-treatment – available case analysis SADS (mean 9 weeks)	*Study population*		RR 0.78 (0.32, 1.88)	199 (1 study)	●○○○ Very low (a,b,c)
	102 per 1000	80 per 1000 (33, 192)			
	Moderate				
	102 per 1000	80 per 1000 (33, 192)			
Long Follow-up (25-103 weeks post-intervention) – ITT analysis MINI	*Study population*		RR 1 (0.56, 1.78)	277 (1 study)	●○○○ Very low (a,b,c,f)
	163 per 1000	163 per 1000 (91, 290)			
	Moderate				
	163 per 1000	163 per 1000 (91, 290)			
PTSD diagnosis					
Post-treatment – ITT analysis LIFE (mean 13 weeks)	*Study population*		RR 0.74 (0.22, 2.47)	54 (1 study)	●○○○ Very low (a,b,c,g)
	192 per 1000	142 per 1000 (42, 475)			
	Moderate				
	192 per 1000	142 per 1000 (42, 474)			
Post-treatment – available case analysis LIFE (mean 13 weeks)	*Study population*		RR 2.54 (0.11, 59.23)	46 (1 study)	●○○○ Very low (a,b,c,g)
	0 per 1000	0 per 1000 (0, 0)			
	Moderate				

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Quality of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
	0 per 1000	0 per 1000 (0, 0)			
PTSD mean scores					
Post-treatment – available case analysis Davidson Trauma Scale or LIFE: Psychiatric Status Ratings mean PTSD score (6-13 weeks)	SMD -0.4 (-0.81, 0)		-	96 (2 studies)	●○○○ Very low (c,e)
OCD mean scores					
Post-treatment – available case analysis Yale–Brown Obsessive Compulsive Scale (YBOCS) (mean 4 weeks)	SMD -0.41 (-0.94, 0.11)		-	58 (1 study)	●○○○ Very low (b,c,e)
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis YBOCS (mean 19 weeks)	**SMD -0.71 (-1.29, -0.12)**		-	50 (1 study)	●○○○ Very low (c,e)
Long follow-up (25-103 weeks post-intervention) – available case analysis YBOCS (mean 32 weeks)	**SMD -0.76 (-1.35, -0.17)**		-	49 (1 study)	●○○○ Very low (c,e)

Evidence Statements:

Psychologically (CBT/IPT) informed psychoeducation has inconsistent effects on depression diagnosis at endpoint or first measurement (very low certainty evidence), at intermediate follow-up (17-24 weeks post-intervention) (very low certainty evidence), and at long follow-up (25-103 weeks post-intervention) (very low certainty evidence) compared with treatment as usual or enhanced treatment as usual in women who have symptoms (or subthreshold symptoms) of depression in the perinatal period.

Psychologically (CBT/IPT) informed psychoeducation improves depression symptomatology (high certainty evidence) at endpoint or first measurement compared with treatment as usual or enhanced treatment as usual in women who have symptoms (or subthreshold symptoms) of depression in the perinatal period.

Psychologically (CBT/IPT) informed psychoeducation has inconsistent effects on depression mean scores at endpoint or first measurement (moderate certainty evidence), at short follow-up (9-16 weeks post-intervention) (moderate certainty evidence), at intermediate follow-up (17-24 weeks post-intervention) (low certainty evidence), and at long follow-up (25-103 weeks post-intervention) (low certainty evidence) compared with treatment as usual or enhanced treatment as usual in women who have symptoms (or subthreshold symptoms) of depression in the perinatal period; however, the magnitude of any benefit may not be clinically significant.

Psychologically (CBT/IPT) informed psychoeducation appears to have no effect on anxiety diagnosis at endpoint or first measurement (very low certainty evidence) or at long follow-up (25-103 weeks post-intervention) (very low certainty evidence) compared with treatment as usual or enhanced treatment as usual in women who have symptoms (or subthreshold symptoms) of depression in the perinatal period.

Psychologically (CBT/IPT) informed psychoeducation appears to have no effect on PTSD diagnosis at endpoint or first measurement (very low certainty evidence) compared with enhanced treatment as usual in women who have experienced intimate partner violence and have subthreshold symptoms of depression in the perinatal period.

Psychologically (CBT/IPT) informed psychoeducation appears to have no effect on PTSD mean scores at endpoint or first measurement (very low certainty evidence) compared with treatment as usual or enhanced treatment as usual in women who have experienced intimate partner violence or have infants in the neonatal intensive care unit, and have subthreshold symptoms of depression in the perinatal period.

Psychologically (CBT/IPT) informed psychoeducation may improve OCD mean scores at endpoint or first measurement at post-treatment (very low certainty evidence), at intermediate follow-up (17-24 weeks post-intervention) (very low certainty evidence) and at long follow-up (25-103 weeks post-intervention) (very low certainty evidence) compared with enhanced treatment as usual in pregnant women who have subthreshold symptoms of OCD; however, the magnitude of the benefits may not be clinically significant.

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Quality of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. Total number of events is less than 300 (a threshold rule-of-thumb) b. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25) c. Papers omit data d. There was evidence of substantial heterogeneity between effect sizes e. Total population size is less than 400 (a threshold rule-of-thumb) f. Risk of bias due to statistically significant group differences at baseline g. Risk of bias due to unclear blinding of outcome assessment					

Source: NICE 2015, Table 141, Table 164, Table 172, Table 174, Table 183

Abbreviations: BDI, Beck Depression Inventory; CBT, cognitive behaviour therapy; CES-D, Center for Epidemiologic Studies Depression Scale; CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; HSCL, Hopkins Symptom Checklist; IPT, interpersonal psychotherapy; ITT, intention-to-treat; LIFE, Longitudinal Interval Follow-up Examination; MINI, Mini International Neuropsychiatric Interview; MMS, Maternal Mood Screener; OCD, obsessive-compulsive disorder; PSCS, Parenting Sense of Competence Scale; PTSD, post-traumatic stress disorder; QIDS, Quick Inventory of Depressive Symptoms; RR, relative risk; SADS, Schedule for Affective Disorders and Schizophrenia; SCID, Structured Clinical Interview for DSM Disorders; SMD, standardised mean difference; VAS, Visual Analogue Scale; YBOCS, Yale–Brown Obsessive Compulsive Scale.

Note: Statistically significant differences are shown in bold.

C3.1.1.2 IPT-informed psychoeducation versus non-mental-health-focused education and support

There was no evidence that IPT-informed psychoeducation was more effective than non-mental-health-focused education and support for treating depression symptomatology (**Table C3-2**).

Table C3-2 Summary of findings (treatment) – IPT-informed psychoeducation versus non-mental health-focused education and support

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression symptomatology					
Post-treatment – ITT Analysis	*Study population*		RR 0.76 (0.53, 1.07)	38 (1 study)	●●○○ Low (a,b)
EPDS	882 per 1000	671 per 1000 (468, 944)			
(mean 16 weeks)	*Moderate*				
	882 per 1000	670 per 1000 (467, 944)			
Evidence Statement: *IPT-informed group psychoeducation appears to have no effect on depressive symptomatology at endpoint or first measurement (low certainty evidence) compared with non-mental-health-focused education and support in pregnant women with a diagnosis of MDD.*					

Footnotes:

* The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).

a. Total number of events is less than 300 (a threshold rule-of-thumb).

b. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25).

Source: NICE 2015, Table 142

Abbreviations: CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; ITT, intention-to-treat; MDD, major depressive disorder; RR, relative risk; SMD, standardised mean difference.

C3.1.2 Psychoeducational booklet

The literature search identified no SRs that relate to this intervention.

Table C3-3 Summary of findings (treatment) – psychoeducational booklet

Evidence Statement: *There is no RCT evidence for psychoeducational booklet in women who have mental health problems in the perinatal period.*

C3.1.3 Social/peer support

Of the three SRs identified in the literature search, NICE 2015 was chosen as the foundation review due to its high quality and comprehensiveness. NICE 2015 included four RCTs that compared social support (peer-mediated support or support group) with treatment as usual (or waitlist) in postnatal women with symptoms of depression. One RCT was from Taiwan and the other three were from Canada. The Taiwanese RCT assessed face-to-face group support, whereas the Canadian RCTs assessed individual telephone support. In one RCT, peer-mediated support was provided at home and/or via the telephone and the intervention included mother-infant relationship content.

NICE 2015 also included one Australian RCT that compared a combined psychosocial (informal support group) and physical exercise (pram walking) intervention with enhanced treatment as usual (telephone support), and another Australian RCT that compared social support group with physical exercise (a pram walking exercise program). In both RCTs, the interventions were aimed at postnatal women with symptoms of depression.

C3.1.3.1 Social support versus treatment as usual

NICE 2015 did not consider the format or mode of delivery of the intervention in their analyses.

There were mixed results for treatment effects on depression outcomes associated with peer-mediated support or support groups (mutual support). There was low-to-moderate-certainty evidence from three studies (N=807) for moderate benefits of social support on depression symptomatology at endpoint using an ITT approach (**Table C3-4**). However, these effects appeared to be transient as no clinically or statistically significant benefits were observed on depression symptomatology at short-term follow-up. Moreover, there was no evidence for clinically or statistically significant benefits of social support on depression diagnosis at endpoint using ITT analysis or for mean depression symptoms at endpoint or short-term follow-up.

There was no evidence for clinically or statistically significant benefits of social support on anxiety symptomatology or anxiety mean symptoms, and no clinically or statistically significant benefits of social support for positive mother-infant feeding or teaching interactions.

Three studies (N=807) found evidence for a moderate effect of social support relative to treatment as usual on attrition with higher drop-out associated with peer-mediated support or a support group. However, this effect was not statistically significant due to very serious imprecision.

Table C3-4 Summary of findings (treatment) – social support versus treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression diagnosis					
Post-treatment – ITT analysis SCID (mean 12 weeks)	*Study population*		RR 1.11 (0.81, 1.52)	701 (1 study)	●○○○ Very low (a,b,c)
	170 per 1000	189 per 1000 (138, 259)			
	Moderate				
	171 per 1000	190 per 1000 (139, 260)			
Post-treatment – available case analysis SCID (mean 12 weeks)	*Study population*		RR 0.65 (0.34, 1.23)	612 (1 study)	●○○○ Very low (a,b,c)
	73 per 1000	47 per 1000 (13, 83)			
	Moderate				
	73 per 1000	47 per 1000 (13, 83)			
Depression symptomatology					
Post-treatment – ITT analysis BDI≥10 or EPDS≥12 (8-14 weeks)	*Study population*		RR 0.69 (0.47, 1.01)	807 (3 studies)	●●○○ Low (a,b)
	359 per 1000	248 per 1000 (169, 363)			
	Moderate				
	546 per 1000	377 per 1000 (257, 551)			
Post-treatment – available case analysis BDI≥10 or EPDS≥12 (8-14 weeks)	*Study population*		**RR 0.52 (0.39, 0.7)**	713 (3 studies)	●●●○ Moderate (a)
	292 per 1000	152 per 1000 (114, 205)			
	Moderate				
	524 per 1000	272 per 1000 (204, 367)			
Short Follow-up (9-16 weeks post-intervention) – ITT analysis EPDS≥12 (mean 24 weeks)	*Study population*		RR 1.12 (0.87, 1.44)	701 (1 study)	●●○○ Low (a,b)
	239 per 1000	267 per 1000 (208, 344)			
	Moderate				
	239 per 1000	268 per 1000 (208, 344)			
Short Follow-up (9-16 weeks post-intervention) – available case analysis EPDS≥12 (mean 24 weeks)	*Study population*		RR 0.83 (0.54, 1.26)	600 (1 study)	●●○○ Low (a,b)
	138 per 1000	115 per 1000 (75, 174)			
	Moderate				
	138 per 1000	115 per 1000 (75, 174)			
Depression mean scores					
Post-treatment – available case analysis BDI or EPDS (12-14 weeks)	SMD -0.12 (-0.68, 0.45)		-	723 (3 studies)	●○○○ Very low (b,d)
Short Follow-up (9-16 weeks post-intervention) – available case analysis EPDS (mean 24 weeks)	SMD -0.13 (-0.29, 0.03)		-	600 (1 study)	●●●● High
Anxiety symptomatology					
Post-treatment – ITT analysis STAI-S >44 (mean 12 weeks)	*Study population*		RR 0.93 (0.75, 1.14)	701 (1 study)	●●○○ Low (a,c)
	349 per 1000	325 per 1000 (262, 398)			
	Moderate				
	349 per 1000	325 per 1000 (262, 398)			
Post-treatment – available case analysis STAI-S>44 (mean 12 weeks)	*Study population*		RR 0.75 (0.56, 1)	612 (1 study)	●○○○ Very low (a,c)
	273 per 1000	205 per 1000 (153, 273)			
	Moderate				
	273 per 1000	205 per 1000 (153, 273)			

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression diagnosis					
Anxiety mean scores					
Post-treatment – available case analysis STAI-S (mean 12 weeks)	SMD -0.14 (-0.3, 0.02)		-	612 (1 study)	●●●○ Moderate (c)
Short follow-up (9-16 weeks post-intervention) – available case analysis STAI-S (mean 24 weeks)	SMD -0.07 (-0.23, 0.09)		-	600 (1 study)	●●●○ Moderate (c)
Mother-infant interaction					
Post-treatment – available case analysis NCAST: Feeding (mean 12 weeks)	SMD -0.18 (-0.79, 0.42)		-	43 (1 study)	●●○○ Low (b,e)
Post-treatment – available case analysis NCAST: Teaching (mean 12 weeks)	SMD -0.45 (-1.04, 0.13)		-	46 (1 study)	●●○○ Low (b,e)

Evidence Statement:

Social support (peer-mediated support or support group) may have an effect[6] on depression symptomatology at endpoint or first measurement (low certainty evidence) compared with treatment as usual in women who have symptoms of depression in the postnatal period; however, the effect is not maintained at short-term follow-up (9-16 weeks post-intervention) (low certainty evidence).

Social support (peer-mediated support or support group) appears to have no effect on depression mean scores at endpoint or first measurement (very low certainty evidence) compared with treatment as usual in women who have symptoms of depression in the postnatal period.

Telephone peer-mediated support appears to have no effect on depression diagnosis at endpoint or first measurement (very low certainty evidence) compared with treatment as usual in women who have symptoms of depression in the postnatal period.

Telephone peer-mediated support has no effect on anxiety mean scores (moderate certainty evidence) and appears to have no effect or on anxiety symptomatology (low certainty evidence) at endpoint or first measurement compared with treatment as usual in women who have symptoms of depression in the postnatal period.

Peer-mediated support (with mother-infant relationship intervention content) appears to have no effect on mother-infant feeding interactions at endpoint or first measurement (low certainty evidence) compared with treatment as usual in women who have symptoms of depression in the postnatal period.

Footnotes:

* The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).

a. Total number of events is less than 300 (a threshold rule-of-thumb).

b. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25).

c. Papers omit data.

d. There was evidence of considerable heterogeneity between effect sizes.

e. Total population size is less than 400 (a threshold rule-of-thumb).

Source: NICE 2015, Table 140, Table 163, Table 190

Abbreviations: BDI, Beck Depression Inventory; CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; ITT, intention-to-treat; NCAST, Nursing Child Assessment Satellite Training Scale; RR, relative risk; SCID, Structured Clinical Interview for DSM Disorders; SMD, standardised mean difference; STAI-S, State-Trait Anxiety Inventory-State.

Note: Statistically significant differences are shown in bold.

[6] RR 0.69 (95% CI 0.47, 1.01); P=0.05

C3.1.3.2 Combined social support and physical exercise versus enhanced treatment as usual

There was single-study evidence (N=20) for large benefits of a combined informal social support group and pram walking exercise program on depression symptomatology and mean depression symptoms; however, confidence in these effect estimates is low due to the extremely low event rate and very small sample size (**Table C3-5**).

Table C3-5 Summary of findings (treatment) – combined social support and physical exercise versus enhanced treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression symptomatology					
Post-treatment – ITT analysis EPDS ≥12 (mean 12 weeks)	*Study population*		RR 0.07 (0, 1.03)	20 (1 study)	●●○○ Low (a,b)
	700 per 1000	49 per 1000 (0, 721)			
	Moderate				
	700 per 1000	49 per 1000 (0, 721)			
Post-treatment – available case analysis EPDS ≥12 (mean 12 weeks)	*Study population*		RR 0.07 (0, 1.03)	20 (1 study)	●●○○ Low (a,b)
	700 per 1000	49 per 1000 (0, 721)			
	Moderate				
	700 per 1000	49 per 1000 (0, 721)			
Depression mean symptoms					
Post-treatment – ITT analysis EPDS (mean 12 weeks)	**SMD -1.64 (-2.68, -0.59)**		-	20 (1 study)	●●○○ Low (c)
Post-treatment – available case analysis EPDS (mean 12 weeks)	**SMD -1.64 (-2.68, -0.59)**		-	20 (1 study)	●●○○ Low (c)
Evidence Statement: *Social support group combined with physical exercise (a pram walking exercise program) may improve depression mean symptoms (low certainty evidence) and may have an effect[7] on depression symptomatology (low certainty evidence) at endpoint or first measurement compared with enhanced treatment as usual (telephone support) in women who have symptoms of depression in the postnatal period.*					
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. Total number of events is less than 300 (a threshold rule-of-thumb). b. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25). c. Total population size is less than 400 (a threshold rule-of-thumb).					

Source: NICE 2015, Table 152

Abbreviations: CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; ITT, intention-to-treat; RR, relative risk; SMD, standardised mean difference.

Note: Statistically significant differences are shown in bold.

[7] RR 0.07 (95% CI 0, 1.03)

C3.1.3.3 Social support versus physical exercise

In order to tease apart the combined intervention discussed above and assess the individual contribution of psychosocial and physical interventions, the same researchers compared social support and physical exercise in a head-to-head trial and provided single-study evidence (N=20) for a large effect of social support (social support group) relative to physical exercise (pram walking exercise program) on depression mean symptoms. However, confidence in this effect estimate was low due to imprecision as a result of the very small sample size.

Table C3-6 Summary of findings (treatment) – social support versus physical exercise

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression mean symptoms					
Post-Treatment – available case analysis EPDS (mean 12 weeks)	**SMD -1.09 (-2.07, -0.11)**		-	19 (1 study)	●●○○ Low (a)
Evidence Statement: *Social support group may improve depression mean symptoms at endpoint or first measurement (low certainty evidence) compared with physical exercise (a pram walking exercise program) in women who have symptoms of depression in the postnatal period.*					
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. Total population size is less than 400 (a threshold rule-of-thumb)					

Source: NICE 2015, Table 153

Abbreviations: CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; SMD, standardised mean difference.

Note: Statistically significant differences are shown in bold.

C3.1.4 Home visits

Only one SR (NICE 2015) was identified for home visits in the literature search. NICE 2015 included five RCTs, one of which was from Australia. The five RCTs compared face-to-face, home-based listening visits (non-directive counselling) with treatment as usual or enhanced treatment as usual[8] in women with a diagnosis of depression or symptoms (or subthreshold symptoms) of depression. The intervention was postnatal in two RCTs and antenatal and postnatal in the remaining three RCTs.

The intervention in the Australian study (Armstrong 1999; N=181) was a structured program of weekly child health nurse visits for vulnerable, or 'high risk', families. 'High risk' was defined at two levels: a) at least one of the following four: sole parenthood; ambivalence to the pregnancy (sought termination, no antenatal care); physical forms of domestic violence; childhood abuse of either parent; and b) three or more of the following: maternal age<18 years old; unstable housing; financial stress; maternal education <10 years; low family income; social isolation; history of mental health disorder (either parent); alcohol or drug

[8] Enhanced treatment as usual involved Medicaid enhanced prenatal/postnatal services.

abuse; domestic violence other than physical abuse. At baseline, 23% of study participants had an Edinburgh Postnatal Depression Scale (EPDS) score of >12 (mean baseline EPDS 8.7 [SD 3.5]). The intervention aimed to enhance parenting self-esteem and confidence, provide anticipatory guidance for normal child development problems, promote preventive child health care, and facilitate access to appropriate community services.

A French RCT (N=364) assessed a manualised multifocal perinatal home-visiting program that was specifically tailored to empower mothers in terms of developing parenting skills, using the health and social care system, and making the most of their personal networks and local community services. Study participants were first-time mothers aged less than 26 years old, who had at least one of the following: less than 12 years' education; were planning to bring up their child without the child's father; or had low income. At baseline, 45% of study participants had EPDS >11. The home visitations were undertaken by a team of psychologists specifically trained to promote mental health and attachment quality, provide social and emotional support within a solid working alliance, and address depression should it occur.

An RCT from the United States (N=440) assessed a voluntary intensive, long-term home visiting program (the Healthy Families Alaska Program). Families were eligible if they scored ≥25 on the Kempe's Family Stress Checklist. Trained home visitors provided information, made referrals to community resources, helped parents prepare for developmental milestones, screened and referred for developmental delay, and promoted child environmental safety. They also supported positive parent–child interaction via role modelling and reinforcement of positive interactions and parental empathy. Visits were offered weekly for the first 6–9 months and less frequently as the family functioning improved. At baseline, 57% of women had a CES-D >15.

Another RCT from the United States (N=613) assessed a Nurse-Community Health Worker (Nurse-CHW) home visiting team intervention in the context of Medicaid enhanced pre/postnatal services. Nurses guided the CHW care, led a multidisciplinary team assessment, provided crisis intervention and case management, assessed and managed health problems (including screening for depression), and had periodic office visits with prenatal providers. CHWs provided relationship-based support through phone and face-to-face contacts. Although a study exclusion criterion was diagnosis or treatment for a pre-existing mental health condition within the last two years, 56% of study participants had CES-D ≥16.

A Japanese RCT (N=18) assessed home visits provided by mental health nurses to women with a diagnosis of minor or major depression using the Structural Clinical Interview for DSM Disorders, modified for postnatal depression (SCID-PND) and EPDS >9. The intervention involved active listening, providing support and acceptance, psychoeducation on depressive symptoms, and advice on coping strategies for problematic life issues, including parenting and increasing access to social or family support. The intervention was provided flexibly according to the individual's needs and the nurse's assessment of each woman's self-care level and mental state.

C3.1.4.1 Home visits versus treatment as usual or enhanced treatment as usual

There was single-study evidence (N=18) for moderate benefits of postnatal home visits on depression diagnosis in Japanese women with a diagnosis of minor or major depression, using an ITT approach (**Table C3-7**). However, confidence in these effect estimates is very low due to the 95% CI including estimates of both no effect and clinically meaningful treatment benefits. The intervention was provided by mental health nurses and involved active listening, providing support and acceptance of the woman, psychoeducation on depressive symptoms, and advice on coping strategies for problematic life issues, including parenting and increasing access to social or family support.

There was no evidence from three studies of clinically or statistically significant treatment effects on depression symptomology, or clinically significant treatment effects on mean depression symptoms.

There was no evidence of clinically or statistically significant benefits of home visits on mother-infant attachment problems.

Table C3-7 Summary of findings (treatment) – home visits versus treatment as usual or enhanced treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression diagnosis					
Post-treatment – ITT analysis SCID (mean 6 weeks)	*Study population*		RR 0.67 (0.28, 1.58)	18 (1 study)	●○○○ Very low (a,b,c)
	667 per 1000	447 per 1000 (187, 1000)			
	Moderate				
	667 per 1000	447 per 1000 (187, 1000)			
Post-treatment – available case analysis SCID (mean 6 weeks)	*Study population*		RR 0.43 (0.12, 1.51)	16 (1 study)	●○○○ Very low (a,b,c)
	667 per 1000	287 per 1000 (-173, 740)			
	Moderate				
	667 per 1000	287 per 1000 (-173, 740)			
Depression symptomatology					
Post-treatment – ITT analysis EPDS ≥10/12 or CES-D ≥24 (22-104 weeks)	*Study population*		RR 0.92 (0.8, 1.06)	985 (3 studies)	●●●○ Moderate (d)
	451 per 1000	415 per 1000 (361, 479)			
	Moderate				
	477 per 1000	439 per 1000 (382, 506)			
Post-treatment – available case analysis EPDS ≥10/12 or CES-D ≥24 (22-104 weeks)	*Study population*		RR 0.87 (0.69, 1.1)	754 (3 studies)	●○○○ Very low (b,c,d)
	279 per 1000	243 per 1000 (193, 307)			
	Moderate				
	220 per 1000	191 per 1000 (152, 242)			
Depression mean scores					
Post-treatment – available case analysis EPDS or CES-D (22-52 weeks)	**SMD -0.17 (-0.3, -0.05)**		-	960 (3 studies)	●●●● High
Mother-infant attachment problems					
Post-treatment – ITT analysis NCAST ≤35 (mean 104 weeks)	*Study population*		RR 0.87 (0.69, 1.09)	364 (1 study)	●○○○ Very low (b,c,e)
	476 per 1000	414 per 1000 (328, 518)			
	Moderate				
	476 per 1000	414 per 1000 (328, 519)			
Post-treatment – available case analysis NCAST ≤35 (mean 104 weeks)	*Study population*		RR 0.79 (0.47, 1.32)	249 (1 study)	●○○○ Very low (b,c,e)
	211 per 1000	167 per 1000 (99, 279)			
	Moderate				
	211 per 1000	167 per 1000 (99, 279)			

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression diagnosis					
Evidence Statements: *Home visits improve depression mean scores at endpoint or first measurement (high certainty evidence) compared with treatment as usual in women who have symptoms of depression in the perinatal period; however, the magnitude of the benefit may not be clinically significant.* *Home visits have no effect on depression symptomatology (moderate certainty evidence) and appear to have no effect on depression diagnosis (very low certainty evidence) at endpoint or first measurement compared with treatment as usual in women who have a diagnosis of depression in the postnatal period.* *A long-term home visiting program to prevent child abuse appears to have no effect on mother-infant attachment problems (very low certainty evidence) at endpoint or first measurement compared with treatment as usual in families that screen positive for family stress in the perinatal period.*					

Footnotes:

* The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).

a. Risk of bias due to unclear blinding of outcome assessment.

b. Total number of events is less than 300 (a threshold rule-of-thumb).

c. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25).

d. Papers omit data.

e. Risk of bias due to statistically significant group differences at baseline.

Source: NICE 2015, Table 144, Table 184, Table 192

Abbreviations: CAGE, cut-annoyed-guilty-eye; CBCL, Child Behaviour Checklist; CES-D, Center for Epidemiologic Studies Depression Scale; CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; ITT, intention-to-treat; NCAST, Nursing Child Assessment Satellite Training Scale; RR, relative risk; SCID, Structured Clinical Interview for DSM Disorders; SMD, standardised mean difference.

Note: Statistically significant differences are shown in bold.

C3.1.5 Non-mental-health-focused education/support

Only one SR (NICE 2015) was identified for non-mental-health-focused education/support in the literature search. NICE 2015 included one RCT from Tanzania (N=331) that compared a face-to-face group counselling intervention for HIV-positive women (antenatal and postnatal) with treatment as usual. Approximately 73% of the study sample had symptoms of depression (Hopkins Symptom Checklist [HSCL]-25>1.06).

C3.1.5.1 Non-mental-health-focused education and support versus treatment as usual

Based on one RCT, there was no evidence for clinically or statistically significant benefits associated with non-mental health-focused education and support for depression symptomatology in HIV-positive women (**Table C3-8**).

Table C3-8 Summary of findings (treatment) – non-mental health-focused education and support versus treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression symptomatology					
Post-treatment – ITT analysis HSCL-25 >1.06 (mean 12 weeks)	*Study population*		RR 0.91 (0.82, 1.01)	331 (1 study)	●●●○ Moderate (a)
	847 per 1000	770 per 1000 (694, 855)			
	Moderate				
	847 per 1000	771 per 1000 (695, 855)			
Post-treatment – available case analysis HSCL-25 >1.06 (mean 12 weeks)	*Study population*		RR 0.82 (0.67, 1.01)	188 (1 study)	●●○○ Low (a,b)
	725 per 1000	595 per 1000 (486, 733)			
	Moderate				
	725 per 1000	595 per 1000 (486, 732)			
Evidence Statement: *Non-mental-health-focused education and support during the perinatal period has no effect on depression symptomatology at endpoint or first measurement compared with treatment (moderate certainty evidence) as usual in HIV-positive women.*					
Footnotes: * The 'assumed risk' for the study population is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The moderate risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. Total number of events is less than 300 (a threshold rule-of-thumb). b. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25).					

Source: NICE 2015, Table 143

Abbreviations: CI, confidence interval; HSCL, Hopkins Symptom Checklist; ITT, intention-to-treat; RR, relative risk; SMD, standardised mean difference.

C3.1.6 Pre-delivery discussion

The literature search identified one SR (NICE 2015) relating to the assessment of pre-delivery discussion/psychoeducation for fear of childbirth (symptoms of tokophobia). However, the outcomes reported in the included RCTs are not relevant to the current Review.

Table C3-9 Summary of findings (treatment) – pre-delivery discussion

Evidence Statement: *There is no RCT evidence for pre-delivery discussion in pregnant women who have mental health problems.*

C3.1.7 Post-delivery discussion

The literature search identified no SRs that relate to this intervention.

Table C3-10 Summary of findings (treatment) – summary of findings: post-delivery discussion

Evidence Statement: *There is no RCT evidence for post-delivery discussion in women who have mental health problems in the perinatal period.*

C3.1.8 Post-miscarriage self-help

Only one SR (NICE 2015) was identified for post-miscarriage self-help in the literature search. NICE 2015 included one RCT from Germany that compared online post-miscarriage self-help (internet-based cognitive behaviour therapy (CBT) involving 10 written assignments with therapist feedback on each) with treatment as usual in self-referring women with complicated grief (mean baseline IES was 33.7 [SD 10.1]), and one multi-armed RCT from the United States that compared post-miscarriage self-help (three 18-minute video accompanied by his and her workbooks) with post-miscarriage facilitated self-help (video and workbook delivery and face-to-face nurse-led counselling) with treatment as usual in women with symptoms of depression (mean baseline CES-D was 21).

C3.1.8.1 Post-miscarriage self-help versus treatment as usual

There was low quality, single-study evidence for moderate benefits of post-miscarriage self-help on depression symptomatology compared with treatment as usual in women with complicated grief (**Table C3-11**); however, the measure for depression symptomatology was treatment non-response (based on reverse scale rating of reliable change index) on the BSI Depression subscale rather than a depression-specific validated checklist. No clinically or statistically significant benefits on mean depression symptoms were observed.

Based on one, low quality study, there was no evidence for statistically or clinically significant benefits of post-miscarriage self-help on anxiety symptomatology or mean symptoms in women with complicated grief.

There was low quality, single-study evidence for moderate effects of post-miscarriage self-help on PTSD symptomatology (analysed using an ITT approach) and large effects on mean PTSD symptoms.

Table C3-11 Summary of findings (treatment) – post-miscarriage self-help versus treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression symptomatology					
Post-treatment – ITT analysis BSI: Depression (Treatment nonresponse: reliable change index) (mean 5 weeks)	*Study population*		**RR 0.65 (0.45, 0.92)**	78 (1 study)	⨁⨁◯◯ Low (a)
	758 per 1000	492 per 1000 (341, 697)			
	Moderate				
	758 per 1000	493 per 1000 (341, 697)			
Post-treatment – available case analysis BSI: Depression (Treatment nonresponse: reliable change index) (mean 5 weeks)	*Study population*		**RR 0.44 (0.25, 0.78)**	59 (1 study)	⨁⨁◯◯ Low (a)
	692 per 1000	305 per 1000 (173, 540)			
	Moderate				
	692 per 1000	304 per 1000 (173, 540)			
Depression mean scores					
Post-treatment – ITT analysis BSI: Depression or CES-D (5-12 weeks)	SMD -0.3 (-1.19, 0.6)		-	250 (2 studies)	⨁◯◯◯ Very low (b,c)
Long Follow-up (25-103 weeks post-intervention) – ITT analysis CES-D (mean 46 weeks)	SMD -0.15 (-0.45, 0.15)		-	172 (1 study)	⨁⨁◯◯ Low (c)
Anxiety symptomatology					
Post-treatment – ITT analysis BSI: Anxiety (Treatment nonresponse: reliable change index) (mean 5 weeks)	*Study population*		RR 0.95 (0.71, 1.26)	78 (1 study)	⨁⨁◯◯ Low (a,d)
	727 per 1000	691 per 1000 (516, 916)			
	Moderate				
	727 per 1000	691 per 1000 (516, 916)			
Post-treatment – available case analysis BSI: Anxiety (Treatment nonresponse: reliable change index) (mean 5 weeks)	*Study population*		RR 0.83 (0.56, 1.23)	59 (1 study)	⨁⨁◯◯ Low (a,d)
	692 per 1000	575 per 1000 (388, 852)			
	Moderate				
	692 per 1000	574 per 1000 (388, 851)			
Anxiety mean scores					
Post-treatment – ITT analysis BSI: Anxiety (mean 5 weeks)	SMD -0.23 (-0.68, 0.23)		-	78 (1 study)	⨁⨁◯◯ Low (c,d)
PTSD symptomatology					
Post-treatment – ITT analysis IES: Treatment nonresponse (reliable change index) (mean 5 weeks)	*Study population*		**RR 0.59 (0.38, 0.94)**	78 (1 study)	⨁⨁◯◯ Low (a)
	636 per 1000	375 per 1000 (242, 598)			
	Moderate				
	636 per 1000	375 per 1000 (242, 598)			
Post-treatment – available case analysis IES: Treatment nonresponse (reliable change index) (mean 5 weeks)	*Study population*		**RR 0.32 (0.14, 0.7)**	59 (1 study)	⨁⨁◯◯ Low (a)
	577 per 1000	185 per 1000 (81, 404)			
	Moderate				
	577 per 1000	185 per 1000 (81, 404)			
PTSD mean scores					
Post-treatment – ITT analysis IES: Traumatic stress (mean 5 weeks)	**SMD -0.84 (-1.31, -0.37)**		-	78 (1 study)	⨁⨁◯◯ Low (c)

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Evidence Statements: ***Women with symptoms of depression*** *Post-miscarriage self-help appears to have no effect on depression mean scores at long follow-up (25-103 weeks post-intervention) (low certainty evidence) compared with treatment as usual in women with symptoms of depression.* ***Women with subthreshold symptoms of PTSD*** *Post-miscarriage self-help may improve depression symptomatology (low certainty evidence) but appears to have no effect on depression mean scores (very low certainty evidence) at endpoint or first measurement compared with treatment as usual in women with subthreshold symptoms of PTSD.* *Post-miscarriage self-help appears to have no effect on anxiety symptomatology (low certainty evidence) or on anxiety mean scores (low certainty evidence) at endpoint or first measurement compared with treatment as usual in women with subthreshold symptoms of PTSD.* *Post-miscarriage self-help may improve PTSD symptomatology (low certainty evidence) and PTSD mean scores (low certainty evidence) at endpoint or first measurement compared with treatment as usual in women with subthreshold symptoms of PTSD.*					

Footnotes:

* The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).

a. Total number of events is less than 300 (a threshold rule-of-thumb)

b. There was evidence of considerable heterogeneity between effect sizes

c. Total population size is less than 400 (a threshold rule-of-thumb)

d. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25)

Source: NICE 2015, Table 134, Table 158, Table 170, Table 179

Abbreviations: BSI, Brief Symptom Inventory; CES-D, Center for Epidemiologic Studies Depression Scale; CI, confidence interval; IES, Impact of Events Scale; ITT, intention-to-treat; PTSD, post-traumatic stress disorder; RR, relative risk; SMD, standardised mean difference.

Note: Statistically significant differences are shown in bold.

C3.1.8.2 Post-miscarriage facilitated self-help versus treatment as usual

There was no evidence for clinically or statistically significant benefits of post-miscarriage facilitated self-help on mean depression symptoms (**Table C3-12**).

Table C3-12 Summary of findings (treatment) – post-miscarriage facilitated self-help versus treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression mean scores					
Post-treatment – ITT analysis CES-D (mean 12 weeks)	SMD 0.13 (-0.17, 0.43)		-	171 (1 study)	●●○○ Low (a)
Long Follow-up (25-103 weeks post-intervention) – ITT analysis CES-D (mean 46 weeks)	SMD -0.1 (-0.4, 0.2)		-	171 (1 study)	●●○○ Low (a)
Evidence Statement: *Post-miscarriage facilitated self-help (video and workbook delivery and face-to-face support) appears to have no effect on depression mean scores at endpoint or first measurement (low certainty evidence), or at long follow-up (25-103 weeks post-intervention) (low certainty evidence), compared with treatment as usual in women with symptoms of depression.*					
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. Total population size is less than 400 (a threshold rule-of-thumb).					

Source: NICE 2015, Table 135

Abbreviations: CES-D, Center for Epidemiologic Studies Depression Scale; CI, confidence interval; ITT, intention-to-treat; SMD, standardised mean difference.

C3.1.9 Seeing and/or holding stillborn infant

The literature search identified no SRs that relate to this intervention.

Table C3-13 Summary of findings (treatment) – seeing and/or holding stillborn infant

Evidence Statement: *There is no RCT evidence for seeing and/or holding a stillborn infant in women who have mental health problems in the perinatal period.*

C3.1.10 Mother-infant relationship interventions

Of the three SRs identified for mother-infant relationship interventions in the literature search, NICE 2015 was chosen as the foundation review due to its high quality and comprehensiveness. NICE 2015 included six RCTs that compared mother-infant relationship interventions with treatment as usual. The intervention was individual in five RCTs and involved a group intervention in the sixth. Participants had a diagnosis of MDD in two RCTs, symptoms of depression in three RCTs, and subthreshold symptoms of depression in one RCT. However, none of the interventions were specifically designed to treat maternal depression; treatment was primarily directed at improving the quality of the mother-infant interaction. Across the studies, several of the outcome measures for maternal-infant interactions were observed rather than based solely on maternal report.

Of the two RCTs that recruited women with a diagnosis of MDD, one assessed a relationship/attachment-based (CBT-informed) therapy that was directed at problems identified by the mother in the management of her infant (concerning, for example, feeding or sleeping), as well as at observed problems in the quality of the mother-infant interaction. The mother was provided with advice about managing particular infant problems, was helped to solve such problems in a systematic way, was encouraged to examine her patterns of thinking about her infant and herself as a mother, and was helped through modelling and reinforcement to alter aspects of her interactional style. In the second RCT, 95% of study participants met DSM-IV criteria for a major depressive episode or dysthymia (mean baseline BDI was 23.6 [SD 8.6]). The intervention involved a home visitor (qualified prevention specialist) who monitored and videotaped the mother and child during everyday activities, and subsequently discussed the interactions with the mother.

NICE 2015 also included two RCTs with active comparators. One Australian RCT (N=51) assessed an in-hospital mother-infant relationship intervention with video feedback compared with a mother-infant relationship intervention with verbal feedback in women with a diagnosis of MDD. NICE 2015 also included an RCT from the United Kingdom (N=80) that assessed a home-based mother-infant relationship intervention compared with listening visits in women with a diagnosed eating disorder (participants in both study arms also received facilitated self-help aimed at their eating disorder).

One additional RCT from the United States that was published after the NICE 2015 literature search compared home-based, nurse-delivered perinatal dyadic psychotherapy with usual care plus depression monitoring by phone in depressed first-time mothers. The pilot RCT (N=42) found no effect of the intervention on depression remission (RR 0.95; 95% CI 0.84, 1.09) or depression symptoms (mean difference in change from baseline -53 (95% CI: -3.74, 0.68). The SR that identified the RCT did not include any mother-infant relationship data from this RCT.

C3.1.10.1 Mother-infant relationship interventions versus treatment as usual or enhanced treatment as usual

NICE 2015 did not consider the format of the intervention in their analyses.

There was mixed evidence for the effects of mother-infant relationship interventions on mother-infant attachment outcomes (**Table C3-14**). Very low certainty evidence from two studies (one using mother-infant psychotherapy and the other using CBT-informed relationship/attachment-based therapy) showed a statistically significant, moderate benefit of mother-infant relationship interventions on reducing attachment problems at endpoint in women with a diagnosis of MDD or symptoms of depression (EPDS = 12), but the positive effect was not maintained in the one study with long follow-up, which showed mild risk of harm (not statistically significant) of the CBT-informed therapy. The same RCT (in women with a diagnosis of MDD) showed a statistically significant, moderate benefit of mother-infant relationship interventions on mother-infant behavior management problems at endpoint.

Evidence from up to four studies found no evidence for statistically or clinically significant effects on continuous measures of mother-infant attachment or positive interactions at endpoint or at intermediate or long-term follow-up. There was single study evidence, in women with a diagnosis of a major depressive episode or dysthymia, for a large harm of a mother–baby intervention on mother-infant positive interaction mean scores at very long follow-up with effects favouring enhanced treatment as usual (telephone support). The 15-week intervention from the Netherlands involved a home visitor videotaping mother and child everyday activities and subsequently discussing the observed interactions with the mother.

There was single-study evidence for moderate benefits of mother-infant relationship interventions on maternal sensitivity treatment response; however, confidence in the effect estimate was very low due to risk-of-bias concerns (statistically significant differences in infant age at baseline and selective reporting bias) and very serious imprecision. Evidence from four studies found no evidence for statistically or clinically significant effects on continuous measures of maternal sensitivity at endpoint or intermediate follow-up; however, there was low quality single-study evidence for moderate benefits of mother-infant relationship interventions on maternal sensitivity at long follow-up.

Evidence for treatment effects of mother-infant relationship interventions on depression outcome measures was very inconsistent (**Table C3-14**). There was single-study evidence (N=95) in women that met DSM-III-R criteria for MDD for moderate benefits of a relationship/attachment-based (CBT-informed) therapy on depression diagnosis at endpoint, but evidence suggestive of harms at long and very long follow-up. However, the certainty of the evidence was low and there was serious imprecision. Low certainty evidence from meta-analyses of up to six studies provided no evidence of clinically or statistically significant benefits of mother-infant relationship interventions on depression symptomatology at endpoint, or depression mean symptoms at endpoint or long-term follow-up. Evidence from one study showed moderate harms of mother-infant relationship interventions on depression symptomatology at intermediate follow-up; however, there was very serious imprecision associated with this effect.

Based on one RCT, there was no evidence for clinically or statistically significant benefits of a mother-infant relationship intervention on anxiety symptomatology (using an ITT approach) or anxiety mean scores; however, there was very serious imprecision.

One RCT showed no evidence for clinically or statistically significant benefits or harms associated with mother-infant relationship interventions for PTSD symptomatology at endpoint or at intermediate follow-up using an ITT analysis approach, and no clinically or statistically significant effects on PTSD mean symptoms.

Five studies (N=576) found no evidence for clinically or statistically significant effects of mother-infant relationship interventions relative to treatment as usual or enhanced treatment as usual on attrition (as a proxy measure for retention in services and treatment acceptability).

Table C3-14 Summary of findings (treatment) – mother-infant relationship interventions versus treatment as usual or enhanced treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression diagnosis					
Post-treatment – ITT analysis SCID (mean 20 weeks)	*Study population*		RR 0.72 (0.48, 1.07)	95 (1 study)	⨁⨁◯◯ Low (a,b)
	615 per 1000	443 per 1000 (295, 658)			
	Moderate				
	615 per 1000	443 per 1000 (295, 658)			
Post-treatment – available case analysis SCID (mean 20 weeks)	*Study population*		RR 0.71 (0.47, 1.08)	92 (1 study)	⨁⨁◯◯ Low (a,b)
	600 per 1000	426 per 1000 (228, 630)			
	Moderate				
	600 per 1000	426 per 1000 (228, 630)			
Intermediate follow-up (17-24 weeks post-intervention) – ITT analysis SCID (mean 39 weeks)	*Study population*		RR 0.83 (0.46, 1.48)	95 (1 study)	⨁⨁◯◯ Low (a,b)
	365 per 1000	303 per 1000 (168, 541)			
	Moderate				
	365 per 1000	303 per 1000 (168, 540)			
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis SCID (mean 39 weeks)	*Study population*		RR 0.8 (0.4, 1.58)	88 (1 study)	⨁⨁◯◯ Low (a,b)
	312 per 1000	250 per 1000 (125, 494)			
	Moderate				
	313 per 1000	250 per 1000 (125, 495)			
Long Follow-up (25-103 weeks post-intervention) – ITT analysis SCID (mean 78 weeks)	*Study population*		RR 1.21 (0.63, 2.33)	95 (1 study)	⨁⨁◯◯ Low (a,b)
	250 per 1000	302 per 1000 (157, 582)			
	Moderate				
	250 per 1000	302 per 1000 (157, 582)			
Long Follow-up (25-103 weeks post-intervention) – available case analysis SCID (mean 78 weeks)	*Study population*		RR 1.52 (0.71, 3.25)	90 (1 study)	⨁⨁◯◯ Low (a,b)
	188 per 1000	285 per 1000 (133, 609)			
	Moderate				
	188 per 1000	286 per 1000 (133, 611)			
Very long Follow-up (≥104 weeks post-intervention) – ITT analysis SCID (mean 260 weeks)	*Study population*		RR 1.21 (0.63, 2.33)	95 (1 study)	⨁⨁◯◯ Low (a,b)
	250 per 1000	302 per 1000 (157, 582)			
	Moderate				
	250 per 1000	302 per 1000 (157, 582)			
Very long Follow-up (≥104 weeks post-intervention) – available case analysis SCID (mean 260 weeks)	*Study population*		RR 0.69 (0.27, 1.73)	73 (1 study)	⨁⨁◯◯ Low (a,b)
	243 per 1000	168 per 1000 (66, 421)			
	Moderate				
	243 per 1000	168 per 1000 (66, 420)			
Depression symptomatology					
Post-treatment - ITT analysis EPDS: Treatment nonresponse (reliable change index-no improvement)/EPDS ≥12 or CES-D ≥16 (5-26 weeks)	*Study population*		RR 0.87 (0.69, 1.1)	396 (3 studies)	⨁⨁◯◯ Low (a,b)
	565 per 1000	492 per 1000 (390, 622)			
	Moderate				
	717 per 1000	624 per 1000 (495, 789)			
Post-treatment – available case analysis EPDS: Treatment nonresponse (reliable change index-no improvement)/EPDS ≥12 or CES-D ≥16 (5-26 weeks)	*Study population*		RR 0.85 (0.58, 1.25)	288 (3 studies)	⨁⨁◯◯ Low (a,b)
	379 per 1000	322 per 1000 (220, 473)			
	Moderate				
	472 per 1000	401 per 1000 (274, 590)			

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Intermediate follow-up (17-24 weeks post-intervention) – ITT analysis EPDS≥12 (mean 25 weeks)	*Study population*		RR 1.27 (0.73, 2.21)	121 (1 study)	●●○○ Low (a,b)
	262 per 1000	333 per 1000 (191, 580)			
	Moderate				
	262 per 1000	333 per 1000 (191, 579)			
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis EPDS ≥12 (mean 25 weeks)	*Study population*		RR 1.63 (0.49, 5.41)	96 (1 study)	●●○○ Low (a,b)
	80 per 1000	130 per 1000 (39, 433)			
	Moderate				
	80 per 1000	130 per 1000 (39, 433)			
Depression mean scores					
Post-treatment – available case EPDS, BDI, BDI-II or CES-D (5-28 weeks)	SMD 0.02 (-0.38, 0.41)		-	566 (6 studies)	●●○○ Low (c)
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis EPDS (mean 39 weeks)	SMD -0.11 (-0.53, 0.31)		-	88 (1 study)	●●○○ Low (b,d)
Long Follow-up (25-103 weeks post-intervention) – available case analysis EPDS or BDI (57-78 weeks)	SMD 0.08 (-0.23, 0.39)		-	161 (2 studies)	●●○○ Low (d)
Very long Follow-up (≥104 weeks post-intervention) – available case analysis EPDS (mean 260 weeks)	SMD -0.17 (-0.66, 0.32)		-	65 (1 study)	●●○○ Low (b,d)
Anxiety symptomatology					
Post-treatment – ITT analysis STAI-S >40 (mean 7 weeks)	*Study population*		RR 0.94 (0.47, 1.89)	121 (1 study)	●●○○ Low (a,b)
	213 per 1000	200 per 1000 (100, 403)			
	Moderate				
	213 per 1000	200 per 1000 (100, 403)			
Post-treatment – available case analysis STAI-S >40 (mean 7 weeks)	*Study population*		RR 0.21 (0.01, 4.23)	98 (1 study)	●●○○ Low (a,b)
	40 per 1000	8 per 1000 (0, 169)			
	Moderate				
	40 per 1000	8 per 1000 (0, 169)			
Anxiety mean scores					
Post-treatment – available case analysis STAI-S (mean 7 weeks)	SMD -0.16 (-0.55, 0.24)		-	98 (1 study)	●●○○ Low (b,d)
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis STAI-S (mean 25 weeks)	SMD -0.3 (-0.7, 0.11)		-	96 (1 study)	●●○○ Low (b,d)
PTSD symptomatology					
Post-treatment – ITT analysis PPQ: Scores in clinical range (no further detail) (mean 7 weeks)	*Study population*		RR 1.18 (0.71, 1.94)	121 (1 study)	●●○○ Low (a,b)
	311 per 1000	368 per 1000 (221, 604)			
	Moderate				
	312 per 1000	368 per 1000 (222, 605)			
Post-treatment – available case analysis PPQ: Scores in clinical range (no further detail) (mean 7 weeks)	*Study population*		RR 1.3 (0.56, 3.02)	98 (1 study)	●●○○ Low (a,b)
	160 per 1000	208 per 1000 (90, 483)			
	Moderate				
	160 per 1000	208 per 1000 (90, 483)			

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Intermediate follow-up (17-24 weeks post-intervention) – ITT analysis PPQ: Scores in clinical range (no further detail) (mean 25 weeks)	*Study population*		RR 1.02 (0.63, 1.63)	121 (1 study)	●●○○ Low (a,b)
	361 per 1000	368 per 1000 (227, 588)			
	Moderate				
	361 per 1000	368 per 1000 (227, 588)			
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis PPQ: Scores in clinical range (no further detail) (mean 25 weeks)	*Study population*		RR 0.79 (0.35, 1.79)	96 (1 study)	●●○○ Low (a,b)
	220 per 1000	174 per 1000 (77, 394)			
	Moderate				
	220 per 1000	174 per 1000 (77, 394)			
PTSD mean scores					
Post-treatment – available case analysis PPQ (mean 7 weeks)	SMD -0.1 (-0.5, 0.29)		-	98 (1 study)	●●○○ Low (b,d)
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis PPQ (mean 25 weeks)	SMD -0.25 (-0.66, 0.15)		-	96 (1 study)	●●○○ Low (b,d)
Mother-infant attachment problems					
Post-treatment – ITT analysis Maternal report: Mother-infant relationship problems or Parent-Infant Relationship Global Assessment Scale (PIRGAS): Treatment nonresponse (no improvement-reliable change index) (20-26 weeks)	*Study population*		**RR 0.55 (0.42, 0.72)**	175 (2 studies)	●○○○ Very low (a,f)
	793 per 1000	436 per 1000 (333, 571)			
	Moderate				
	789 per 1000	434 per 1000 (331, 568)			
Post-treatment – available case analysis Maternal report: Mother-infant relationship problems or PIR-GAS: Treatment non-response (no improvement-reliable change index) (20-26 weeks)	*Study population*		**RR 0.55 (0.41, 0.74)**	151 (2 studies)	●○○○ Very low (a,f)
	736 per 1000	405 per 1000 (302, 545)			
	Moderate				
	736 per 1000	405 per 1000 (302, 545)			
Long follow-up (25-103 weeks post-intervention) – ITT analysis Maternal report: Mother-infant relationship problems (mean 78 weeks)	*Study population*		RR 1.16 (0.79, 1.71)	95 (1 study)	●●○○ Low (a,b)
	481 per 1000	558 per 1000 (380, 822)			
	Moderate				
	481 per 1000	558 per 1000 (380, 823)			
Long follow-up (25-103 weeks post-intervention) – available case Maternal report: Mother-infant relationship problems (mean 78 weeks)	*Study population*		RR 1.26 (0.81, 1.95)	88 (1 study)	●●○○ Low (a,b)
	426 per 1000	536 per 1000 (345, 830)			
	Moderate				
	426 per 1000	537 per 1000 (345, 831)			
	371 per 1000	316 per 1000 (171, 591)			
	Moderate				
	371 per 1000	315 per 1000 (171, 590)			
Mother-infant positive interaction mean scores					
Post-treatment – available case analysis Dyadic Mutuality Code, PIR-GAS, Behavioural observation: Positive mother-infant interaction or Global Rating Scales of Mother-infant Interaction: Overall mother-infant interaction (5-26 weeks)	SMD 0.15 (-0.26, 0.56)		-	378 (4 studies)	●○○○ Very low (b,d,g)
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis Global Rating Scales of Mother-infant Interaction: Overall mother-infant interaction (mean 25 weeks)	SMD 0 (-0.4, 0.4)		-	96 (1 study)	●●○○ Low (d)
Very long follow-up (>104 weeks post-intervention) – available case analysis Behavioural observation: Positive mother-infant interaction (mean 271 weeks)	**SMD -1.82 (-2.44, -1.2)**		-	58 (1 study)	●●○○ Low (d)

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Maternal sensitivity treatment response					
Post-treatment – ITT analysis Emotional Availability Scales (EAS): Maternal sensitivity: Treatment response (improvement-reliable change index) (mean 26 weeks)	*Study population*		RR 1.67 (0.43, 6.51)	80 (1 study)	●○○○ Very low (a,b,e,h)
	75 per 1000	125 per 1000 (32, 488)			
	Moderate				
	75 per 1000	125 per 1000 (32, 488)			
Post-treatment – available case analysis EAS: Maternal sensitivity: Treatment response (improvement-reliable change index) (mean 26 weeks)	*Study population*		RR 1.62 (0.42, 6.31)	75 (1 study)	●○○○ Very low (a,b,e,h)
	81 per 1000	131 per 1000 (34, 512)			
	Moderate				
	81 per 1000	131 per 1000 (34, 511)			
Maternal sensitivity mean scores					
Post-treatment – available case analysis EAS: Maternal sensitivity or Behavioural observation: Maternal sensitivity or Global Rating Scales of Mother-infant Interaction: Maternal sensitive behaviour (5-28 weeks)	SMD 0.23 (-0.08, 0.53)		-	332 (4 studies)	●○○○ Very low (b,d,i)
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis Global Rating Scales of Mother-infant Interaction: Maternal sensitive behavior (mean 25 weeks)	SMD 0.15 (-0.25, 0.55)		-	96 (1 study)	●●○○ Low (b,d)
Long follow-up (25-103 weeks post-intervention)- Available case analysis EAS: Maternal sensitivity (mean 57 weeks)	**SMD 0.81 (0.33, 1.3)**		-	71 (1 study)	●●○○ Low (d)

Evidence Statements:

Individual mother-infant relationship interventions

Mother-infant relationship interventions (individual) may improve mother-infant attachment problems (very low certainty evidence) at endpoint or first measurement compared with treatment as usual in women with a diagnosis of MDD or symptoms of depression.

Mother-infant relationship interventions (individual) appear to have no effect on (or may be harmful to) mother-infant attachment problems at long follow-up (25-103 weeks post-intervention) (low certainty evidence) compared with treatment as usual in women with a diagnosis of MDD.

Mother-infant relationship interventions (individual) appear to have no effect on mother-infant positive interaction mean scores at intermediate follow-up (17-24 weeks post-intervention) (low certainty evidence) compared with enhanced treatment as usual (a booklet about infant care) in women with symptoms of depression.

Mother-infant relationship interventions (individual) appear to be harmful to mother-infant positive interaction mean scores at very long follow-up (>104 weeks post-intervention) (low certainty evidence) compared with enhanced treatment as usual (telephone support) in women with a diagnosis of a major depressive episode or dysthymia.

Mother-infant relationship interventions (individual mother-infant psychotherapy) appear to have no effect on maternal sensitivity treatment response at endpoint or first measurement (very low certainty evidence) compared with treatment as usual in women with symptoms of depression.

Mother-infant relationship interventions (individual) may have an effect[9] on depression diagnosis at endpoint or first measurement (low certainty evidence), but appear to have no effect on depression diagnosis at intermediate follow-up (17-24 weeks post-intervention) (low certainty evidence), at long follow-up (25-103 weeks post-intervention) (low certainty evidence), or at very long follow up (>103 weeks post-intervention) (low certainty evidence) compared with treatment as usual in women with a diagnosis of MDD.

Mother-infant relationship interventions (individual) appear to have no effect on depression mean scores (low certainty evidence) at intermediate (17-24 weeks post-intervention), long (25-103 weeks post-intervention), or very long (>103 weeks post-treatment) follow-up compared with treatment as usual or enhanced treatment as usual in women with a diagnosis of depression.

Mother-infant relationship interventions (individual) appear to have no effect on depression symptomatology (low certainty evidence) at intermediate follow-up (17-24 weeks post-intervention) than enhanced treatment as usual (a booklet about infant care) in women with symptoms of depression.

[9] RR 0.72 (95% CI 0.48, 1.07); P=0.10

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Mother-infant relationship interventions (individual) appear to have no effect on anxiety symptomatology at endpoint or first measurement (low certainty evidence), and appear to have no effect on anxiety mean scores at endpoint or first measurement (low certainty evidence) or at intermediate follow-up (17-24 weeks post-intervention) (low certainty evidence), compared with enhanced treatment as usual (a booklet about infant care) in women with symptoms of depression. *Mother-infant relationship interventions (individual) appear to have no effect on PTSD symptomatology (low certainty evidence) or on PTSD mean symptoms (low certainty evidence) at endpoint or first measurement, or at intermediate follow-up (17-24 weeks post-intervention), compared with enhanced treatment as usual (a booklet about infant care) in women with symptoms of depression.* ***Individual or group mother-infant relationship interventions*** *Mother-infant relationship interventions (individual or group) appear to have no effect on mother-infant positive interaction mean scores at endpoint or first measurement (very low certainty evidence) compared with treatment as usual or enhanced treatment as usual in women with a diagnosis of MDD or symptoms (or subthreshold symptoms) of depression.* *Mother-infant relationship interventions (individual or group) appear to have no effect on maternal sensitivity mean scores at endpoint or first measurement (very low certainty evidence) compared with treatment as usual or enhanced treatment as usual in women with symptoms (or subthreshold symptoms) of depression or a diagnosis of a major depressive episode or dysthymia, but may improve maternal sensitivity mean scores at long follow-up (low certainty evidence).* *Mother-infant relationship interventions (individual or group) appear to have no effect on depression symptomatology (low certainty evidence) or on depression mean scores (low certainty evidence) at endpoint or first measurement compared with treatment as usual or enhanced treatment as usual in women with a diagnosis of depression or symptoms (or subthreshold symptoms) of depression.*					

Footnotes:

* The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).

a. Total number of events is less than 300 (a threshold rule-of-thumb).
b. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25).
c. There was evidence of considerable heterogeneity between effect sizes.
d. Total population size is less than 400 (a threshold rule-of-thumb).
e. Risk of bias due to statistically significant group differences at baseline.
f. Risk of bias due to statistically significant group differences at baseline and non-blind outcome assessment.
g. There is evidence of substantial heterogeneity of study effect sizes.
h. Paper omits data.
i. There is evidence of moderate heterogeneity of study effect sizes.

Source: NICE 2015, Table 145, Table 165, Table 173, Table 193

Abbreviations: AQS, Attachment Q Set; ASQ:SE, Ages and Stages Questionnaire: Social-Emotional; BDI, Beck Depression Inventory; CBCL, Child Behaviour Checklist; CES-D, Center for Epidemiologic Studies Depression Scale; CI, confidence interval; EAS, Emotional Availability Scales; EPDS, Edinburgh Postnatal Depression Scale; ITT, intention-to-treat; MDD, major depressive disorder; PIRGAS, Parent-Infant Relationship Global Assessment Scale; PPQ, Perinatal PTSD Questionnaire; PTSD, post-traumatic stress disorder; RR, relative risk; SCID, Structured Clinical Interview for DSM Disorders; SMD, standardised mean difference; STAI-S, State-Trait Anxiety Inventory-State; TAU, treatment as usual.

Note: Statistically significant differences are shown in bold.

C3.1.10.2 Mother-infant relationship intervention with video feedback versus mother-infant relationship intervention with verbal feedback

A single study (N=51) found no advantage of video feedback compared with verbal feedback for effects of mother-infant relationship interventions on mean depression symptoms (**Table C3-15**). There was no clinically or statistically significant difference between study arms on attrition.

The study outcomes relating to mother-infant interactions (maternal confidence/competence mean scores; maternal perceptions of infant behavior mean scores) are not captured in **Table C3-15** as they are not considered to be key outcomes for the current Review.

Table C3-15 Summary of findings (treatment) – mother-infant relationship intervention with video feedback versus mother-infant relationship intervention with verbal feedback

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression mean scores					
Post-treatment – available case analysis EPDS (mean 3 weeks)	SMD 0.29 (-0.36, 0.94)		-	37 (1 study)	●●○○ Low (a,b)
Evidence Statement: *Mother-infant relationship intervention (individual) with video feedback appears to have no effect on depression mean scores at endpoint or first measurement (low certainty evidence) compared with a mother-infant relationship intervention (individual) with verbal feedback in women with a diagnosis of MDD.*					
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. Total population size is less than 400 (a threshold rule-of-thumb) b. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25)					

Source: NICE 2015, Table 146

Abbreviations: CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; MDD, major depressive disorder; RR, relative risk; SMD, standardised mean difference.

C3.1.10.3 Mother-infant relationship intervention (and facilitated self-help for eating disorders) versus listening visits (and facilitated self-help for eating disorders

There was very low quality single-study evidence (N=80) for moderate-to-large benefits of a mother-infant relationship intervention relative to listening visits for women with eating disorders for reducing mealtime conflict, maternal inappropriate verbal responses, and infant autonomy, but not for maternal intrusions. However, none of these outcomes are considered to be key outcomes for the current Review.

There was higher dropout observed in the mother-infant relationship intervention group; however, this effect was not statistically significant due to very serious imprecision.

Table C3-16 Summary of findings (treatment) – mother-infant relationship intervention (and facilitated self-help for eating disorders) versus listening visits (and facilitated self-help for eating disorders)

Evidence Statement: *There is no RCT evidence for any pre-defined important outcomes for mother-infant relationship interventions relative to listening visits for women with eating disorders.*

C3.1.11 Co-parenting interventions

Only one SR (NICE 2015) was identified for co-parenting interventions in the literature search. NICE 2015 identified one RCT from Canada (N=29; Misri 2000), which compared a face-to-face co-parenting intervention with enhanced treatment as usual (monitoring) in postpartum women with a diagnosis of MDD.

C3.1.11.1 Co-parenting intervention versus enhanced treatment as usual

There was single-study evidence for a moderate effect of a co-parenting intervention on depression diagnosis; however, confidence in this effect estimate was very low due to very serious imprecision (**Table C3-17**). In addition, the same study showed no evidence for statistically or clinically significant benefits of a co-parenting intervention on mean depression symptoms. There were no drop-outs in either arm.

Table C3-17 Summary of findings (treatment) – co-parenting intervention versus enhanced treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression diagnosis					
Post-treatment – ITT analysis MINI (mean 6 weeks)	*Study population*		RR 0.51 (0.22, 1.18)	29 (1 study)	●○○○ Very low (a,b,c)
	615 per 1000	314 per 1000 (135, 726)			
	Moderate				
	615 per 1000	314 per 1000 (135, 726)			
Post-treatment – available case analysis MINI (mean 6 weeks)	*Study population*		RR 0.51 (0.22, 1.18)	29 (1 study)	●○○○ Very low (a,b,c)
	615 per 1000	314 per 1000 (-37, 665)			
	Moderate				
	615 per 1000	314 per 1000 (-37, 664)			
Depression mean scores					
Post-treatment – available case analysis EPDS (mean 6 weeks)	SMD -0.47 (-1.22, 0.29)		-	28 (1 study)	●○○○ Very low (a,c,d)
Evidence Statement: *Co-parenting interventions appear to have no effect on depression diagnosis (very low certainty evidence) or depression mean scores (very low certainty evidence) at endpoint or first measurement compared with enhanced treatment as usual (monitoring) in postpartum women with a diagnosis of MDD.*					

Footnotes:

* The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).

a. Risk of bias as blinding of outcome assessment was unclear.

b. Total number of events is less than 300 (a threshold rule-of-thumb).

c. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25).

d. Total population size is less than 400 (a threshold rule-of-thumb).

Source: NICE 2015, Table 147

Abbreviations: CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; ITT, intention-to-treat; MDD, major depressive disorder; MINI, Mini International Neuropsychiatric Interview; RR, relative risk; SMD, standardised mean difference.

C3.1.12 Mindfulness

Of the four SRs identified in the literature search, NICE 2015 was chosen as the foundation review due to its high quality and comprehensiveness. NICE 2015 included two RCTs from the United States that compared antenatal, face-to-face, group mindfulness training with enhanced treatment as usual[10] or waitlist. The intervention in one RCT (N=47) was aimed at women with elevated levels of perceived stress or pregnancy-specific anxiety, and the other RCT (N=34) was aimed at women with mood concerns (31% of all participants exceeded a score of 16 on the CES-D, and the mean baseline CES-D score in the intervention group was 20.4, which is above the clinical cut-off of 16).

C3.1.12.1 Mindfulness training versus treatment as usual or enhanced treatment as usual

There was no evidence for statistically or clinically significant benefits associated with mindfulness training on mean depression symptoms or negative affect mean scores, or on mean anxiety symptoms (**Table C3-18**).

There was evidence for a moderate effect of mindfulness training relative to enhanced treatment as usual on attrition, with higher drop-out in the mindfulness training group; however, this effect was not statistically significant due to very serious imprecision.

Table C3-18 Summary of findings (treatment) – mindfulness training versus treatment as usual or enhanced treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression mean scores					
Post-treatment – available case analysis CES-D (mean 10 weeks)	SMD -0.13 (-0.85, 0.58)		-	31 (1 study)	●○○○ Very low (a,b,c)
Anxiety mean scores					
Post-treatment – ITT analysis STAI-S (mean 6 weeks)	SMD 0.23 (-0.35, 0.8)		-	47 (1 study)	●●○○ Low (a,b)
Post-treatment – available case analysis STAI-S (mean 10 weeks)	SMD -0.02 (-0.74, 0.69)		-	31 (1 study)	●○○○ Very low (a,b,c)
Evidence Statement: *Group mindfulness training appears to have no effect on depression mean scores at endpoint or first measurement (very low certainty evidence) compared with waitlist in pregnant women with mood concerns.* *Group mindfulness training appears to have no effect on anxiety mean scores at endpoint or first measurement (low certainty evidence) compared with enhanced treatment as usual (non-mental health-focused education and support booklet) in pregnant women with elevated levels of perceived stress or pregnancy-specific anxiety.*					

[10] Enhanced treatment as usual involved non-mental health-focused education and support (book).

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. Total population size is less than 400 (a threshold rule-of-thumb) b. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25) c. Paper omits data					

Source: NICE 2015, Table 151, Table 168

Abbreviations: CES-D, Center for Epidemiologic Studies Depression Scale; CI, confidence interval; ITT, intention-to-treat; RR, relative risk; SMD, standardised mean difference; STAI-S, State-Trait Anxiety Inventory-State.

C3.2 TREATMENT WITH PSYCHOLOGICAL INTERVENTIONS

C3.2.1 Structured psychological interventions (CBT or IPT)

Of the 10 SRs identified for structured psychological interventions in the literature search, NICE 2015 was chosen as the foundation review due to its currency, high quality and comprehensiveness. While the NICE 2015 SR separated structured psychological interventions (CBT or IPT) from psychologically (CBT/IPT) informed psychoeducation (which is covered in **Section C3.1.1**), several of the other identified SRs combined these interventions.

NICE 2015 included 14 RCTs (N=2,099) that compared face-to-face structured psychological interventions (CBT or IPT) with treatment as usual or enhanced treatment as usual in women with a diagnosis of depression (MDD, major depressive episode, minor depression, depressive disorder) or symptoms of depression. The intervention was IPT in four RCTs and CBT in the remaining 10 RCTs. Four of the included RCTs were from Australia (three CBT studies and one IPT study). Across the 14 RCTs, the timing and format of the intervention varied considerably. In nine RCTs (including the four Australian studies) the intervention was postnatal, two RCTs assessed antenatal interventions, and in three RCTs the intervention was both antenatal and postnatal. The format was individual in 12 RCTs, group in one RCT, and both individual and group in one RCT.

The comparator also varied across the 14 included RCTs. One RCT compared CBT plus home visits with home visits only, one RCT compared IPT with waitlist, six RCTs compared CBT or IPT with treatment as usual, and six RCTs compared CBT or IPT with enhanced treatment as usual[11].

NICE 2015 also included four RCTs that compared CBT or IPT with active interventions. One RCT from the United States (N=34[12]) compared face-to-face individual CBT with listening visits in pregnant women with a diagnosis of MDD. One RCT from the United Kingdom (N=3,449) compared face-to-face individual CBT with listening visits in postpartum women with symptoms of depression. One RCT from Brazil (N=60) compared face-to-face individual CBT with relational constructivist

[11] Enhanced treatment as usual varied across the six RCTs: single session psychoeducation; GP training; single session post-delivery discussion; non-specific emotional support and mothercraft advice; and psychoeducation booklet, monitoring and improved access to support.

[12] In Hayden 2012 the number of randomised women is unclear but there were 34 participants in the completer analysis.

therapy in postpartum women with symptoms of depression. One RCT from the United States (N=48) compared face-to-face group IPT with a support group in pregnant women with a diagnosis of MDD or dysthymia.

Although NICE 2015 included only two RCTs of group CBT (one of which also had an individual component), one of the identified SRs (Scope 2013) specifically assessed the effectiveness of group CBT on PND. Scope 2013 used a broad definition of CBT in their review to include studies of group programs that included elements derived from cognitive behaviour principles, including psychoeducational activities. Three RCTs[13], two non-RCTs, and two observational studies met the inclusion criteria. No RCTs directly compared individual CBT with group CBT. Meta-analyses showed that group CBT appeared to be clinically effective when compared to routine primary care, usual care or a waiting list group, although the reduction in depression scores was not consistent across time. The authors note that the results should be interpreted with caution due to the limited number and quality of the studies. In addition, some of the studies included concurrent therapy, the effects of which are difficult to separate from group treatment. There was also uncertainty as to how accurately some of the described group treatments reflect CBT.

C3.2.1.1 Structured psychological interventions versus treatment as usual or enhanced treatment as usual

NICE 2015 did not separately consider CBT and IPT interventions in their analyses. Furthermore, NICE 2015 did not consider the timing of the intervention, format, setting or mode of delivery in their analyses.

Very low-to-high certainty evidence from up to ten studies showed that structured psychological interventions (CBT or IPT) were more effective than treatment as usual or enhanced treatment as usual in reducing depression diagnosis, depression symptomatology, and depression mean scores at post-treatment, with large to moderate effects observed for all outcomes and some low certainty evidence for maintained moderate-to-large effects at short-term follow-up (**Table C3-19**). At intermediate follow-up periods, there was evidence for moderate benefits associated with structured psychological interventions; however, confidence that these were true measures of effect was low to very low due to wide confidence intervals. At longer-term follow-ups (>24 weeks post-intervention), the evidence for structured psychological interventions is very inconsistent with point estimates of effect in favour of CBT or IPT for depression symptomatology, but in favour of treatment as usual or enhanced treatment as usual for depression diagnosis.

There was low quality, single-study evidence for a large effect of a structured psychological intervention on mean state anxiety symptoms (using an ITT analysis approach; however, an available case analysis approach (two studies) revealed no evidence for clinically significant benefits (although differences were statistically significant) associated with mean state anxiety symptoms, and the small benefit for trait anxiety symptoms found in a single-study analysis also failed to reach the threshold for appreciable benefit despite meeting statistical-significance criteria.

There was low-to-very low certainty evidence from up to two studies for moderate-to-large benefits of structured psychological interventions (CBT or IPT) on general mental health outcomes at endpoint, and at short-term and intermediate follow-ups. There was also evidence for a statistically significant, but not clinically significant, effect of CBT on reducing the risk of self-harm.

[13] One of the RCTs was included in NICE 2015, one was specifically excluded, and one was classified by NICE 2015 as a psychoeducational intervention.

There was high to very low certainty evidence from up to two studies for moderate-to-large benefits of structured psychological interventions (CBT or IPT) in reducing mother-infant attachment problems at endpoint and at long-term follow-up, mother-infant attachment mean scores and mother-infant play frequency. There was, however, no evidence for clinically or statistically significant benefits on mother-infant attachment mean scores at short-term follow-up.

Twelve studies (N=1,983) found no evidence for clinically or statistically significant effects of structured psychological interventions (CBT or IPT) relative to treatment as usual or enhanced treatment as usual on attrition (as a proxy measure for retention in services and treatment acceptability).

Table C3-19 Summary of findings (treatment) – structured psychological interventions (CBT or IPT) versus treatment as usual or enhanced treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression diagnosis					
Post-treatment – ITT SCID or CIS-R (12-44 weeks)	*Study population*		**RR 0.48 (0.39, 0.6)**	1307 (6 studies)	●●●● High
	652 per 1000	313 per 1000 (254, 391)			
	Moderate				
	687 per 1000	330 per 1000 (268, 412)			
Post-treatment – available case analysis SCID or CIS-R (12-44 weeks)	*Study population*		**RR 0.38 (0.24, 0.58)**	1,066 (5 studies)	●●○○ Low (a)
	602 per 1000	229 per 1000 (145, 349)			
	Moderate				
	615 per 1000	234 per 1000 (148, 357)			
Short Follow-up (9-16 weeks post-intervention) – ITT SCID (mean 28 weeks)	*Study population*		**RR 0.39 (0.19, 0.8)**	93 (1 study)	●●○○ Low (e)
	435 per 1000	170 per 1000 (83, 348)			
	Moderate				
	435 per 1000	170 per 1000 (83, 348)			
Intermediate follow-up (17-24 weeks post-intervention) – ITT CIS-R or SCID (mean 33 weeks)	*Study population*		RR 0.59 (0.24, 1.41)	138 (2 studies)	●○○○ Very low (a,e,f)
	471 per 1000	278 per 1000 (113, 665)			
	Moderate				
	572 per 1000	337 per 1000 (137, 807)			
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis CIS-R or SCID (mean 33 weeks)	*Study population*		RR 0.5 (0.23, 1.08)	118 (2 studies)	●●○○ Low (e,f)
	373 per 1000	186 per 1000 (86, 403)			
	Moderate				
	474 per 1000	237 per 1000 (109, 512)			
Long Follow-up (25-103 weeks post-intervention) – ITT SCID (mean 78 weeks)	*Study population*		RR 1.68 (0.95, 2.98)	102 (1 study)	●●○○ Low (e,f)
	250 per 1000	420 per 1000 (237, 745)			
	Moderate				
	250 per 1000	420 per 1000 (237, 745)			
Long Follow-up (25-103 weeks post-intervention) – available case analysis SCID (mean 78 weeks)	*Study population*		RR 1.56 (0.73, 3.33)	89 (1 study)	●●○○ Low (e,f)
	188 per 1000	292 per 1000 (137, 624)			
	Moderate				
	188 per 1000	293 per 1000 (137, 626)			
Very long Follow-up (>104 weeks post-intervention) – ITT SCID (mean 260 weeks)	*Study population*		**RR 1.92 (1.11, 3.33)**	102 (1 study)	●●○○ Low (e)
	250 per 1000	480 per 1000 (278, 832)			
	Moderate				
	250 per 1000	480 per 1000 (278, 832)			
Very long Follow-up (>104 weeks post-intervention) – available case analysis SCID (mean 260 weeks)	*Study population*		RR 0.87 (0.37, 2.08)	70 (1 study)	●●○○ Low (e,f)
	243 per 1000	212 per 1000 (90, 506)			
	Moderate				
	243 per 1000	211 per 1000 (90, 505)			

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression symptomatology					
Post-treatment – ITT EPDS ≥10/EPDS ≥12/Treatment nonresponse (baseline to endpoint decrease <4 points and EPDS >13)/Treatment nonresponse (<50% improvement) or BDI ≥16 or BDI-II ≥14 (6-44 weeks)	*Study population*		**RR 0.69 (0.56, 0.85)**	969 (10 studies)	⨁⨁◯◯ Low (b,c)
	643 per 1000	444 per 1000 (360, 547)			
	Moderate				
	626 per 1000	432 per 1000 (351, 532)			
Post-treatment – available case analysis EPDS ≥10/EPDS ≥12/Treatment nonresponse (baseline to endpoint decrease <4 points and EPDS >13) or BDI ≥ 16 or BDI-II ≥ 14 (6-16 weeks)	*Study population*		**RR 0.62 (0.53, 0.73)**	702 (9 studies)	⨁⨁⨁⨁ High
	559 per 1000	347 per 1000 (296, 408)			
	Moderate				
	588 per 1000	365 per 1000 (312, 429)			
Short Follow-up (9-16 weeks post-intervention) – ITT BDI-II ≥14 (mean 29 weeks)	*Study population*		RR 0.89 (0.54, 1.47)	55 (1 study)	⨁⨁◯◯ Low (e,f)
	560 per 1000	498 per 1000 (302, 823)			
	Moderate				
	560 per 1000	498 per 1000 (302, 823)			
Short Follow-up (9-16 weeks post-intervention) – available case analysis BDI-II ≥14 (mean 29 weeks)	*Study population*		RR 0.57 (0.31, 1.07)	42 (1 study)	⨁⨁◯◯ Low (e)
	667 per 1000	380 per 1000 (207, 713)			
	Moderate				
	667 per 1000	380 per 1000 (207, 713)			
Long Follow-up (25-103 weeks post-intervention) – ITT EPDS ≥10 (mean 32 weeks)	*Study population*		RR 0.71 (0.2, 2.53)	37 (1 study)	⨁◯◯◯ Very low (e,f,g)
	250 per 1000	178 per 1000 (50, 632)			
	Moderate				
	250 per 1000	178 per 1000 (50, 632)			
Long Follow-up (25-103 weeks post-intervention) – available case analysis EPDS ≥10 (mean 32 weeks)	*Study population*		RR 0.4 (0.05, 3.46)	33 (1 study)	⨁◯◯◯ Very low (e,f,g)
	167 per 1000	67 per 1000 (8, 577)			
	Moderate				
	167 per 1000	67 per 1000 (8, 578)			
Depression mean scores					
Post-treatment – ITT EPDS or BDI-II (6-44 weeks)	**SMD -1.31 (-2.36, -0.26)**		-	306 (5 studies)	⨁⨁⨁◯ Moderate (a,d)
Post-treatment – available case analysis EPDS, BDI, BDI-II or HRSD (6-16 weeks)	**SMD -0.6 (-0.8, -0.4)**		-	1,508 (10 studies)	⨁⨁⨁◯ Moderate (b)
Short Follow-up (9-16 weeks post-intervention) – ITT EPDS or BDI-II (28-29 weeks)	SMD -1.84 (-4.31, 0.64)		-	148 (2 studies)	⨁◯◯◯ Very low (a,d,f)
Short Follow-up (9-16 weeks post-intervention) – available case analysis EPDS or BDI-II (21-29 weeks)	**SMD -0.66 (-1.14, -0.18)**		-	89 (2 studies)	⨁⨁◯◯ Low (d)
Intermediate Follow-up (17-24 weeks post-intervention) – available case analysis EPDS (mean 33 weeks)	SMD -0.51 (-1.72, 0.7)		-	118 (2 studies)	⨁◯◯◯ Very low (a,d,f)
Long Follow-up (25-103 weeks post-intervention) – available case analysis EPDS or BDI (32-78 weeks)	SMD -0.28 (-0.8, 0.23)		-	142 (3 studies)	⨁⨁◯◯ Low (d,f)
Very long Follow-up (>104 weeks post-intervention) – available case analysis EPDS (mean 260 weeks)	SMD -0.17 (-0.67, 0.33)		-	62 (1 study)	⨁⨁◯◯ Low (d,f)

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Negative thoughts/mood mean scores					
Available case analysis Automatic Thought Questionnaire (mean 4 weeks)	**SMD -0.94 (-1.83, -0.04)**		-	22 (1 study)	●○○○ Very low (d,g)
Anxiety mean scores					
Post-treatment – ITT analysis Beck Anxiety Inventory (BAI) (mean 44 weeks)	**SMD -1.34 (-1.94, -0.74)**		-	53 (1 study)	●●○○ Low (d)
Post-treatment – available case analysis BAI or STAI-S (12-26 weeks)	**SMD -0.35 (-0.58, -0.13)**		-	315 (2 studies)	●●○○ Low (c,d)
Mother-infant attachment problems					
Post-treatment – ITT analysis Maternal report: Mother-infant relationship problems (mean 20 weeks)	*Study population*		**RR 0.65 (0.49, 0.87)**	102 (1 study)	●●○○ Low (e)
	827 per 1000	537 per 1000 (405, 719)			
	Moderate				
	827 per 1000	538 per 1000 (405, 719)			
Post-treatment – available case analysis Maternal report: Mother-infant relationship problems (mean 20 weeks)	*Study population*		**RR 0.63 (0.43, 0.91)**	78 (1 study)	●●○○ Low (e)
	743 per 1000	468 per 1000 (319, 676)			
	Moderate				
	743 per 1000	468 per 1000 (319, 676)			
Long follow-up (25-103 weeks post-intervention) – ITT analysis Maternal report: Mother-infant relationship problems (mean 78 weeks)	*Study population*		RR 1.29 (0.9, 1.84)	102 (1 study)	●●○○ Low (e,f)
	481 per 1000	620 per 1000 (433, 885)			
	Moderate				
	481 per 1000	620 per 1000 (433, 885)			
Long follow-up (25-103 weeks post-intervention) – available case analysis Maternal report: Mother-infant relationship problems (mean 78 weeks)	*Study population*		RR 1.23 (0.79, 1.92)	87 (1 study)	●●○○ Low (e,f)
	426 per 1000	523 per 1000 (336, 817)			
	Moderate				
	426 per 1000	524 per 1000 (337, 818)			
Mother-infant attachment mean scores					
Post-treatment – available case analysis Prenatal Attachment Inventory or Maternal Attachment Inventory (8-15 weeks)	SMD 2.28 (-1.17, 5.73)		-	76 (2 studies)	●○○○ Very low (d,f,h)
Short follow-up (9-16 weeks post-intervention) – available case analysis Maternal Attachment Inventory (mean 21 weeks)	SMD 0.32 (-0.27, 0.91)		-	45 (1 study)	●●○○ Low (d,f)

Evidence Statements:

CBT or IPT

Structured psychological interventions (individual CBT or IPT) improve depression diagnosis at endpoint or first measurement (high certainty evidence) compared with treatment as usual or enhanced treatment as usual in pregnant or postpartum women with a diagnosis of depression.

Structured psychological interventions (individual CBT or IPT) appear to have no effect on depression diagnosis at intermediate follow-up (17-24 weeks post-intervention) (low certainty evidence) compared with treatment as usual in pregnant or postpartum women with a diagnosis of MDD or depression.

Structured psychological interventions (individual or group CBT or IPT) may improve depression symptomatology at endpoint or first measurement (low certainty evidence) compared with treatment as usual or enhanced treatment as usual in pregnant or postpartum women with a diagnosis of depression or symptoms of depression.

Structured psychological interventions (individual CBT or IPT) improve depression mean scores at endpoint or first measurement (moderate certainty evidence) compared with treatment as usual or enhanced treatment as usual in pregnant and postpartum women with a diagnosis of depression or symptoms of depression.

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Structured psychological interventions (individual CBT or IPT) appear to have no effect on depression mean scores at intermediate follow-up (17-24 weeks post-intervention) (very low certainty evidence) compared with treatment as usual in pregnant or postpartum women with a diagnosis of MDD or depression. *Structured psychological interventions (individual or group CBT or IPT) appear to have no effect on depression mean scores at long follow-up (>24 weeks post-intervention) (low certainty evidence) compared with treatment as usual or enhanced treatment as usual in postpartum women with a diagnosis of MDD or depression.* *Structured psychological interventions (individual or group CBT or IPT) appear to have no effect on mother-infant attachment mean scores at endpoint or first measurement (very low certainty evidence) compared with treatment as usual in pregnant or postpartum women with a diagnosis of depression or MDD.* ***CBT*** *Structured psychological interventions (individual CBT and home visits) may improve depression diagnosis at short follow-up (9-16 weeks post-intervention) (low certainty evidence) compared with home visits alone in postpartum women with a diagnosis of MDD.* *Structured psychological interventions (individual CBT) appear to have no effect on depression symptomatology at short follow-up (9-16 weeks post-intervention) (low certainty evidence) compared with treatment as usual in pregnant or postpartum women with a diagnosis of MDD.* *Structured psychological interventions (individual CBT) appear to have no effect on depression symptomatology at long follow-up (>24 weeks post-intervention) (very low certainty evidence) compared with enhanced treatment as usual non-specific emotional support and mothercraft advice) in postpartum women with a diagnosis of MDD.* *Structured psychological interventions (individual CBT with or without home visits) appear to have no effect on depression mean scores at short follow-up (9-16 weeks post-intervention) (very low certainty evidence) compared with treatment as usual or home visits alone in pregnant or postpartum women with a diagnosis of MDD.* *Structured psychological interventions (individual CBT) may improve negative thoughts/mood mean score at endpoint or first measurement (very low certainty evidence) compared with enhanced treatment as usual (single session psychoeducation) in pregnant women with a diagnosis of depressive disorder.* ***IPT*** *Structured psychological interventions (individual IPT) may improve anxiety mean scores at endpoint or first measurement (low certainty evidence) compared with enhanced treatment as usual (psychoeducation booklet, monitoring and improved access to support) in pregnant or postpartum women with a diagnosis of depression; however, the magnitude of the benefit may not be clinically significant.* *Structured psychological interventions (individual and group IPT) appear to have no effect on mother-infant attachment mean scores at short follow-up (9-16 weeks post-intervention) (low certainty evidence) compared with treatment as usual in postpartum women with a diagnosis of MDD.* ***IPT – psychodynamic therapy*** *Structured psychological interventions (individual IPT [psychodynamic therapy]) appear to be less effective at improving depression diagnosis at long follow-up (>24 weeks post-intervention) (low certainty evidence) and at very long follow-up (>104 weeks post-intervention) (low certainty evidence) than treatment as usual in postpartum women with a diagnosis of MDD.* *Structured psychological interventions (individual IPT [psychodynamic therapy]) appear to have no effect on depression mean scores at very long follow-up (>104 weeks post-intervention) (low certainty evidence) compared with treatment as usual in postpartum women with a diagnosis of MDD.* *Structured psychological interventions (individual IPT [psychodynamic therapy]) may improve mother-infant attachment problems at endpoint or first measurement (low certainty evidence) compared with treatment as usual in postpartum women with a diagnosis of MDD.* *Structured psychological interventions (individual IPT [psychodynamic therapy]) appear to have no effect on (and may be harmful to) mother-infant attachment problems at long follow-up (>24 weeks) (low certainty evidence) compared with treatment as usual in postpartum women with a diagnosis of MDD.*					

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. There was evidence of substantial heterogeneity between effect sizes. b. There was evidence of moderate-to-substantial heterogeneity between effect sizes. c. Papers omit data. d. Total population size is less than 400 (a threshold rule-of-thumb). e. Total number of events is less than 300 (a threshold rule-of-thumb). f. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25). g. Risk of bias due to statistically significant group differences at baseline. h. There is evidence of considerable heterogeneity of study effect sizes. i. Risk of bias due to unclear blinding of outcome assessment.					

Source: NICE 2015, Table 129, Table 154, Table 177, Table 187

Abbreviations: CBT, cognitive behaviour therapy; BAI, Beck Anxiety Inventory; BDI, Beck Depression Inventory; BSI, Brief Symptom Inventory; CI, confidence interval; CIS-R, Computerised version of the Clinical Interview Schedule – Revised; CORE-OM, Clinical Outcomes in Routine Evaluation-Outcome Measure; ES-R, Impact of Events Scale – Revised; IPT, interpersonal psychotherapy; ITT, intention-to-treat; MDD, major depressive disorder; PTSD, post-traumatic stress disorder; RR, relative risk; SCID, Structured Clinical Interview for DSM Disorders; SF, Short Form Health Survey; SMD, standardised mean difference; STAI-S, State-Trait Anxiety Inventory-State; STAI-T, State-Trait Anxiety Inventory-Trait; WHO, World Health Organization.

Note: Statistically significant differences are shown in bold.

C3.2.1.2 CBT versus listening visits

There was no evidence for benefits associated with CBT relative to listening visits on mean depression symptoms at endpoint or first measurement (**Table C3-20**).

Table C3-20 Summary of findings (treatment) – CBT versus listening visits

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression mean scores					
Post-treatment – available case analysis BDI or EPDS (mean 26 weeks)	SMD -0.06 (-0.33, 0.22)		-	301 (2 studies)	●●○○ Low (a)
Evidence Statement: *Individual CBT appears to have no effect on depression means scores at endpoint or first measurement (low certainty evidence) compared with listening visits in pregnant or postpartum women with a diagnosis of MDD or symptoms of depression.*					
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. Papers omit data					

Source: NICE 2015, Table 130

Abbreviations: BDI, Beck Depression Inventory; CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; MDD, major depressive disorder; SMD, standardised mean difference.

C3.2.1.3 IPT versus support group

There was no evidence for clinically or statistically significant effects of IPT relative to a support group on mean depression or anxiety symptoms (**Table C3-21**).

Table C3-21 Summary of findings (treatment) – IPT versus support group

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression mean scores					
Post-treatment – available case analysis CES-D (mean 12 weeks)	SMD -0.49 (-1.09, 0.11)		-	44 (1 study)	●○○○ Very low (a,b,c)
Anxiety mean scores					
Post-treatment – available case STAI-S (mean 12 weeks)	SMD -0.48 (-1.09, 0.12)		-	44 (1 study)	●○○○ Very low (a,b,c)
Evidence Statement: *Group IPT appears to have no effect on depression mean scores (very low certainty evidence) or on anxiety mean scores (very low certainty evidence) at endpoint or first measurement compared with a support group in pregnant women with a diagnosis of MDD or dysthymia.*					

Footnotes:
* The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).
a. Risk of bias due to statistically significant group differences at baseline.
b. Total population size is less than 400 (a threshold rule-of-thumb).
c. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25).

Source: NICE 2015, Table 132, Table 156

Abbreviations: CES-D, Center for Epidemiologic Studies Depression Scale; CI, confidence interval; MDD, major depressive disorder; SMD, standardised mean difference; STAI-S, State-Trait Anxiety Inventory-State.

C3.2.2 Directive counselling

Only one SR (NICE 2015) was identified for directive counselling in the literature search. NICE 2015 included one Australian RCT that compared face-to-face directive counselling (individual and group) with treatment as usual in postpartum women with a diagnosis of minor depression or MDD.

C3.2.2.1 Directive counselling versus treatment as usual

There was low quality, single-study evidence that directive counselling was more effective than treatment as usual for depression symptomatology with moderate effects observed on dichotomous measures at endpoint and a large effect observed on a continuous measure at long-term follow-up; however, the effects on mean depression symptoms at endpoint were not statistically or clinically significant.

There was low quality single-study evidence for moderate effects of directive counselling on mean anxiety symptoms using an available case analysis approach.

There was no evidence for clinically or statistically significant effects of directive counselling relative to treatment as usual on attrition (as a proxy measure for retention in services and treatment acceptability).

Table C3-22 Summary of findings (treatment) – directive counselling versus treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression symptomatology					
Post-treatment - ITT analysis BDI≥16 (mean 12 weeks)	*Study population*		**RR 0.72 (0.59, 0.88)**	146 (1 study)	⬤⬤◯◯ Low (a)
	848 per 1000	611 per 1000 (501, 747)			
	Moderate				
	849 per 1000	611 per 1000 (501, 747)			
Post-treatment - available case analysis BDI≥16 (mean 12 weeks)	*Study population*		**RR 0.54 (0.36, 0.81)**	90 (1 study)	⬤⬤◯◯ Low (a)
	722 per 1000	390 per 1000 (260, 585)			
	Moderate				
	722 per 1000	390 per 1000 (260, 585)			
Depression mean scores					
Post-treatment – available case analysis BDI (mean 12 weeks)	SMD -0.42 (-0.95, 0.1)		-	90 (1 study)	⬤⬤◯◯ Low (b,c)
Long Follow-up (25-103 weeks post-intervention) – available case analysis BDI (mean 52 weeks)	**SMD -1.46 (-2.29, -0.63)**		-	45 (1 study)	⬤⬤◯◯ Low (b)
Anxiety mean scores					
Post-treatment – available case analysis BAI (mean 12 weeks)	**SMD -0.56 (-1.09, -0.04)**		-	90 (1 study)	⬤⬤◯◯ Low (b)

Evidence Statements:

Directive counselling may improve depression symptomatology (low certainty evidence) at endpoint or first measurement compared with treatment as usual in postpartum women with a diagnosis of minor depression or MDD.

Directive counselling appears to have no effect on depression mean scores at endpoint or first measurement (low certainty evidence) but may improve depression mean scores at long follow-up (25-103 weeks post-intervention) (low certainty evidence) compared with treatment as usual in postpartum women with a diagnosis of minor depression or MDD.

Directive counselling may improve anxiety mean scores at endpoint or first measurement (low certainty evidence) compared with treatment as usual in postpartum women with a diagnosis of minor depression or MDD.

Footnotes:

* The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).

a. Total number of events is less than 300 (a threshold rule-of-thumb).

b. Total population size is less than 400 (a threshold rule-of-thumb).

c. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25).

Source: NICE 2015, Table 137, Table 160

Abbreviations: BAI, Beck Anxiety Inventory; BDI, Beck Depression Inventory; CI, confidence interval; ITT, intention-to-treat; MDD, major depressive disorder; RR, relative risk; SMD, standardised mean difference.

Note: Statistically significant differences are shown in bold.

C3.2.3 Non-directive counselling

Of the three SRs of non-directive counseling identified in the literature search, NICE 2015 was chosen as the foundation review due to its high quality and comprehensiveness. NICE 2015 included five RCTs that compared face-to-face non-directive counselling (listening visits) in the home with treatment as usual in postpartum women with symptoms (or subthreshold symptoms) of depression or a diagnosis of depression or MDD.

One additional RCT from the United States (N=66) published after the NICE 2015 literature search assessed listening visits provided by point-of-care providers (e.g. home visitor or physician's assistant) and delivered to low-income, ethnic minority, depressed pregnant women or mothers of young children. Listening visits provided clinically and statistically significant benefits on depression mean scores compared with waitlist controls (women receiving standard social/health services).

C3.2.3.1 Listening visits/non-directive counselling versus treatment as usual

There was no evidence for statistically significant benefits of listening visits on depression diagnosis at endpoint, using an ITT approach (**Table C3-23**). At follow-up, there was some evidence that listening visits may be less effective than treatment as usual, with point estimates suggestive of clinically significant harms on depression diagnosis at long and very long follow-up. There was no statistically or clinically significant effect of listening visits on depression symptomatology at endpoint or at follow-up and no clinically significant effect of listening visits on depression mean scores at endpoint or at follow-up.

There was low quality single-study evidence for statistically significant effects of listening visits on mean state anxiety symptoms; however, the effect was small and failed to reach a threshold indicative of clinically significant treatment benefits. In addition, the confidence in the effect estimate was low due to small sample size and selective outcome reporting.

There was low quality, single-study evidence for moderate benefits of listening visits on reducing mother-infant attachment problems and behavior management problems. However, the effect on behavior management problems was not statistically significant and effects on mother-infant attachment problems were not maintained at long-term follow-up.

Three studies (N=1,211) found no evidence for clinically or statistically significant effects of listening visits relative to treatment as usual on attrition (as a proxy measure for retention in services and treatment acceptability).

Table C3-23 Summary of findings (treatment) – listening visits/non-directive counselling versus treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression diagnosis					
Post-treatment – ITT analysis SCID (mean 20 weeks)	*Study population*		RR 0.74 (0.51, 1.08)	100 (1 study)	●●○○ Low (a,b)
	615 per 1000	455 per 1000 (314, 665)			
	Moderate				
	615 per 1000	455 per 1000 (314, 664)			
Post-treatment – available case analysis SCID or Goldberg's standardised psychiatric interview: Research diagnostic criteria or psychiatric interview using Montgomery–Åsberg Depression Rating Scale (MADRS) (7-20 weeks)	*Study population*		**RR 0.54 (0.31, 0.93)**	179 (3 studies)	●○○○ Very low (a,c,d)
	633 per 1000	317 per 1000 (82, 551)			
	Moderate				
	625 per 1000	312 per 1000 (81, 544)			
Intermediate follow-up (17-24 weeks post-intervention) – ITT analysis SCID (mean 20 weeks)	*Study population*		RR 0.97 (0.57, 1.64)	100 (1 study)	●●○○ Low (a,b)
	365 per 1000	354 per 1000 (208, 599)			
	Moderate				
	365 per 1000	354 per 1000 (208, 599)			
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis SCID (mean 20 weeks)	*Study population*		RR 1.09 (0.61, 1.94)	95 (1 study)	●●○○ Low (a,b)
	312 per 1000	341 per 1000 (191, 606)			
	Moderate				
	313 per 1000	341 per 1000 (191, 607)			
Long Follow-up (25-103 weeks post-intervention) – ITT analysis SCID (mean 20 weeks)	*Study population*		RR 1.42 (0.77, 2.6)	100 (1 study)	●●○○ Low (a,b)
	250 per 1000	355 per 1000 (192, 650)			
	Moderate				
	250 per 1000	355 per 1000 (192, 650)			
Long Follow-up (25-103 weeks post-intervention) – available case analysis SCID (mean 20 weeks)	*Study population*		RR 1.66 (0.8, 3.45)	93 (1 study)	●●○○ Low (a,b)
	188 per 1000	311 per 1000 (150, 647)			
	Moderate				
	188 per 1000	312 per 1000 (150, 649)			
Very long Follow-up (>104 weeks post-intervention) – ITT analysis SCID (mean 260 weeks)	*Study population*		RR 1.83 (1.04, 3.22)	100 (1 study)	●●○○ Low (a)
	250 per 1000	458 per 1000 (260, 805)			
	Moderate				
	250 per 1000	458 per 1000 (260, 805)			
Very long Follow-up (>104 weeks post-intervention) – available case analysis SCID (mean 260 weeks)	*Study population*		RR 0.87 (0.37, 2.08)	70 (1 study)	●●○○ Low (a,b)
	243 per 1000	212 per 1000 (90, 506)			
	Moderate				
	243 per 1000	211 per 1000 (90, 505)			
Depression symptomatology					
Post-treatment – ITT analysis EPDS≥12 (26-52 weeks)	*Study population*		RR 0.96 (0.84, 1.09)	1,111 (2 studies)	●●●○ Moderate (d)
	452 per 1000	434 per 1000 (380, 493)			
	Moderate				
	494 per 1000	474 per 1000 (415, 538)			

Outcomes (follow-up)	**Illustrative comparative risks***		Relative effect (95% CI)	**No. of participants (studies)**	**Certainty of the evidence (GRADE)**
	Assumed risk Control	**Corresponding risk Intervention**			
Post-treatment – available case analysis EPDS≥12 (26-52 weeks)	*Study population*		RR 0.82 (0.66, 1.01)	885 (2 studies)	●●○○ Low (a,b,d)
	331 per 1000	271 per 1000 (218, 334)			
	Moderate				
	373 per 1000	306 per 1000 (246, 377)			
Long Follow-up (25-103 weeks post-intervention) – ITT analysis GHQ≥12 (mean 78 weeks)	*Study population*		RR 0.98 (0.87, 1.11)	731 (1 study)	●●●○ Moderate (d)
	651 per 1000	638 per 1000 (567, 723)			
	Moderate				
	652 per 1000	639 per 1000 (567, 724)			
Long Follow-up (25-103 weeks post-intervention) – available case analysis GHQ≥12 (mean 78 weeks)	*Study population*		RR 0.96 (0.79, 1.15)	549 (1 study)	●●○○ Low (a,d)
	538 per 1000	516 per 1000 (425, 618)			
	Moderate				
	538 per 1000	516 per 1000 (425, 619)			
Depression mean scores					
Post-treatment – available case analysis EPDS (20-26 weeks)	**SMD -0.34 (-0.55, -0.14)**		-	375 (2 studies)	●●●○ Moderate (d)
Intermediate follow-up (17-24 weeks post-intervention) – by intervention EPDS or CES-D (4-12 weeks)	SMD -0.07 (-0.35, 0.21)		-	197 (2 studies)	●●●○ Moderate (e)
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis EPDS (mean 20 weeks)	SMD 0.07 (-0.33, 0.48)		-	94 (1 study)	●●○○ Low (e)
Long Follow-up (25-103 weeks post-intervention) – available case analysis EPDS (mean 78 weeks)	SMD 0.14 (-0.26, 0.55)		-	92 (1 study)	●●○○ Low (b,e)
Very long Follow-up (>104 weeks post-intervention) – available case analysis EPDS (mean 260 weeks)	SMD -0.19 (-0.67, 0.29)		-	67 (1 study)	●●○○ Low (b,e)
Anxiety mean scores					
Post-treatment – available case analysis STAI-S (mean 26 weeks)	**SMD -0.29 (-0.53, -0.04)**			260 (1 study)	●●○○ Low (d,e)
Mother-infant attachment problems					
Post-treatment – ITT analysis Maternal report: Mother-infant relationship problems (mean 20 weeks)	*Study population*		**RR 0.71 (0.54, 0.92)**	100 (1 study)	●●○○ Low (a)
	827 per 1000	587 per 1000 (447, 761)			
	Moderate				
	827 per 1000	587 per 1000 (447, 761)			
Post-treatment – available case analysis Maternal report: Mother-infant relationship problems (mean 20 weeks)	*Study population*		RR 0.72 (0.51, 1.01)	78 (1 study)	●●○○ Low (a,b)
	743 per 1000	535 per 1000 (379, 750)			
	Moderate				
	743 per 1000	535 per 1000 (379, 750)			
Long follow-up (25-103 weeks post-intervention) – ITT analysis Maternal report: Mother-infant relationship problems (mean 78 weeks)	*Study population*		RR 1.08 (0.73, 1.6)	100 (1 study)	●●○○ Low (a,b)
	481 per 1000	519 per 1000 (351, 769)			
	Moderate				
	481 per 1000	519 per 1000 (351, 770)			

Outcomes (follow-up)	**Illustrative comparative risks***		Relative effect (95% CI)	**No. of participants (studies)**	**Certainty of the evidence (GRADE)**
	Assumed risk Control	**Corresponding risk Intervention**			
Long follow-up (25-103 weeks post-intervention) – available case analysis Maternal report: Mother-infant relationship problems (mean 78 weeks)	*Study population*		RR 0.96 (0.58, 1.59)	86 (1 study)	⨁⨁◯◯ Low (a,b)
	426 per 1000	409 per 1000 (247, 677)			
	Moderate				
	426 per 1000	409 per 1000 (247, 677)			

Evidence Statements:

Non-directive counselling in the home appears to have no effect on depression diagnosis at endpoint or first measurement (low certainty evidence) or at intermediate follow-up (17-24 weeks post-intervention) (low certainty evidence) or at long follow-up (25-103 weeks post-intervention) (low certainty evidence), and may be less effective on depression diagnosis at very long follow-up (>104 weeks post-intervention) (low certainty evidence) than treatment as usual in postpartum women with a diagnosis of MDD.

Listening visits in the home have no effect on depression symptomatology at endpoint or first measurement (moderate certainty evidence), or at long follow-up (25-103 weeks post-intervention) (moderate certainty evidence), compared with treatment as usual in postpartum women with symptoms (or subthreshold symptoms) of depression.

Non-directive counselling/listening visits in the home improve depression mean scores at endpoint or first measurement (moderate certainty evidence) compared with treatment as usual in postpartum women with a diagnosis of depression or symptoms of depression; however, the magnitude of the benefit is not clinically significant.

Non-directive counselling in the home appears to have no effect on depression mean scores at intermediate follow-up (17-24 weeks post-intervention) (low certainty evidence), long follow-up (>24 weeks post-intervention) (low certainty evidence) and very long follow-up (>104 weeks post-intervention) (low certainty evidence) compared with treatment as usual in postpartum women with a diagnosis of MDD.

Non-directive counselling in the home may improve state anxiety mean scores (low certainty evidence) at endpoint or first measurement compared with treatment as usual in postpartum women with symptoms of depression; however, the magnitude of the benefits may not be clinically significant.

Non-directive counselling in the home may improve mother-infant attachment problems at endpoint or first measurement (low certainty evidence), but appears to have no effect on mother-infant attachment problems at long follow-up (25-103 weeks post-intervention) (low certainty evidence) compared with treatment as usual in postpartum women with a diagnosis of MDD.

Footnotes:

* The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).

a. Total number of events is less than 300 (a threshold rule-of-thumb)

b. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25)

c. There was evidence of moderate-to-substantial heterogeneity between effect sizes

d. Papers omit data

e. Total population size is less than 400 (a threshold rule-of-thumb)

Source: NICE 2015, Table 136, Table 159, Table 180, Table 189

Abbreviations: CI, confidence interval; CIS-R, Computerised version of the Clinical Interview Schedule – Revised; EPDS, Edinburgh Postnatal Depression Scale; GHQ, General Health Questionnaire; ITT, intention-to-treat; MADRS, Montgomery–Åsberg Depression Rating Scale; MDD, major depressive disorder; PTSD, post-traumatic stress disorder; RR, relative risk; SCID, Structured Clinical Interview for DSM Disorders; SMD, standardised mean difference; STAI-S, State-Trait Anxiety Inventory-State.

Note: Statistically significant differences are shown in bold.

C3.2.4 Case management/individual treatment

The literature search identified no SRs that relate to this intervention.

Table C3-24 Summary of findings (treatment) – case management/individual treatment

Evidence Statement:
There is no RCT evidence for case management or individualised treatment in women who have mental health problems in the perinatal period.

C3.2.5 Self-help and facilitated self-help

Only one SR on self-help or facilitated self-help was identified in the literature search; NICE 2015 included three RCTs that compared facilitated self-help with treatment as usual. One Australian study assessed facilitated self-help (workbook delivery and telephone support) in pregnant women with subthreshold symptoms of depression. The other two RCTs from the United Kingdom assessed internet-delivered self-help, one of which also offered online (chat room) support for postpartum women with symptoms of depression, while the other offered telephone support to postpartum women with a diagnosis of MDD.

C3.2.5.1 Facilitated self-help versus treatment as usual

There was very low-to-high quality data from up to three studies for moderate benefits of facilitated self-help relative to treatment as usual for depression symptomatology and mean depression symptoms (**Table C3-25**). There was very low quality, single-study evidence for moderate benefits of facilitated self-help relative to treatment as usual for treating anxiety symptomatology and for mean anxiety symptoms.

Table C3-25 Summary of findings (treatment) – facilitated self-help versus treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression symptomatology					
Post-treatment – ITT Analysis BDI-II≥14 or EPDS>12 (15-20 weeks)	*Study population*		**RR 0.73 (0.53, 0.99)**	1,136 (3 studies)	●○○○ Very low (a,b)
	817 per 1000	596 per 1000 (433, 809)			
	Moderate				
	762 per 1000	556 per 1000 (404, 754)			
Post-treatment – available case analysis BDI-II≥14 or EPDS>12 (15-20 weeks)	*Study population*		**RR 0.58 (0.44, 0.77)**	503 (3 studies)	●●○○ Low (b,c)
	567 per 1000	329 per 1000 (250, 437)			
	Moderate				
	586 per 1000	340 per 1000 (258, 451)			
Depression mean scores					
Post-treatment – available case analysis EPDS (15-17 weeks)	**SMD -0.56 (-0.76, -0.37)**		-	414 (2 studies)	●●●● High
Anxiety symptomatology					
Post-treatment – ITT analysis Depression Anxiety Stress Scale (DASS): Anxiety ≥8 (mean 20 weeks)	*Study population*		**RR 0.67 (0.47, 0.96)**	143 (1 study)	●○○○ Very low (b,c)
	569 per 1000	382 per 1000 (268, 547)			
	Moderate				
	569 per 1000	381 per 1000 (267, 546)			

Outcomes (follow-up)	**Illustrative comparative risks***		**Relative effect (95% CI)**	**No. of participants (studies)**	**Certainty of the evidence (GRADE)**
	Assumed risk Control	**Corresponding risk Intervention**			
Post-treatment – available case analysis Depression Anxiety Stress Scale (DASS): Anxiety ≥8 (mean 20 weeks)	*Study population*		**RR 0.24 (0.07, 0.81)**	89 (1 study)	●○○○ Very low (b,c)
	262 per 1000	63 per 1000 (18, 212)			
	Moderate				
	262 per 1000	63 per 1000 (18, 212)			
Anxiety mean scores					
Post-treatment – available case analysis Generalised Anxiety Disorder Assessment (GAD-7) (mean 17 weeks)	SMD -0.5 (-1.02, 0.02)		-	59 (1 study)	●○○○ Very low (b,d,e)
Evidence Statements: *Facilitated self-help (internet delivery with online or telephone support) improves depression mean scores at endpoint or first measurement (high certainty evidence) compared with treatment as usual in postpartum women with a diagnosis of MDD or symptoms of depression.* *Facilitated self-help (workbook or internet delivery with online or telephone support) may improve depression symptomatology at endpoint or first measurement (very low certainty evidence) compared with treatment as usual in pregnant or postpartum women with a diagnosis of MDD or symptoms (or subthreshold symptoms) of depression.* *Facilitated self-help (workbook delivery with telephone support) may improve anxiety symptomatology at endpoint or first measurement (very low certainty evidence) compared with treatment as usual in pregnant women with subthreshold symptoms of depression.* *Facilitated self-help (internet delivery with telephone support) appears to have no effect on anxiety mean scores at endpoint or first measurement (very low certainty evidence) compared with treatment as usual in postpartum women with a diagnosis of MDD.*					
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. There was evidence of considerable heterogeneity between effect sizes. b. Papers omit data. c. Total number of events is less than 300 (a threshold rule-of-thumb). d. Total population size is less than 400 (a threshold rule-of-thumb). e. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25).					

Source: NICE 2015, Table 133, Table 157

Abbreviations: BDI, Beck Depression Inventory; CI, confidence interval; DASS, Depression Anxiety Stress Scale; EPDS, Edinburgh Postnatal Depression Scale; GAD-7, Generalised Anxiety Disorder Assessment: ITT, intention-to-treat; MDD, major depressive disorder; RR, relative risk; SMD, standardised mean difference.

Note: Statistically significant differences are shown in bold.

C3.2.6 Post-traumatic birth counselling

Only one SR (NICE 2015) was identified for post-traumatic birth counselling in the literature search. NICE 2015 included one Australian RCT (N=103) that compared individual post-traumatic birth counselling (face-to-face and telephone) with treatment as usual in women with a diagnosis of PTSD.

C3.2.6.1 Post-traumatic birth counselling versus treatment as usual

There was low quality, single-study evidence for large effects of post-traumatic birth counselling on depression symptomatology (**Table C3-26**). The same study showed a large effect of post-traumatic birth counselling on anxiety symptomatology; however, confidence that this is a true measure of the effect is low due to the low number of events and imprecision. There was no evidence for statistically or clinically significant benefits of post-traumatic birth counselling on PTSD diagnosis and no evidence for a clinically significant effect (despite meeting statistical significance criteria) on mean PTSD symptoms. The study reported no drop-outs from either study arm.

Table C3-26 Summary of findings (treatment) – post-traumatic birth counseling versus treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression symptomatology					
Post-treatment – ITT analysis EPDS≥12 (mean 13 weeks)	*Study population*		**RR 0.25 (0.09, 0.69)**	103 (1 study)	●●○○ Low (a)
	321 per 1000	80 per 1000 (29, 221)			
	Moderate				
	321 per 1000	80 per 1000 (29, 221)			
Post-treatment – available case analysis EPDS≥12 (mean 13 weeks)	*Study population*		**RR 0.25 (0.09, 0.69)**	103 (1 study)	●●○○ Low (a)
	321 per 1000	80 per 1000 (29, 221)			
	Moderate				
	321 per 1000	80 per 1000 (29, 221)			
Anxiety symptomatology					
Post-treatment – ITT analysis Depression Anxiety Stress Scale (DASS): Anxiety >9 (mean 13 weeks)	*Study population*		RR 0.18 (0.02, 1.42)	103 (1 study)	●●○○ Low (a,b)
	113 per 1000	20 per 1000 (2, 161)			
	Moderate				
	113 per 1000	20 per 1000 (2, 160)			
Post-treatment – available case analysis Depression Anxiety Stress Scale (DASS): Anxiety >9 (mean 13 weeks)	*Study population*		RR 0.18 (0.02, 1.42)	103 (1 study)	●●○○ Low (a,b)
	113 per 1000	20 per 1000 (2, 161)			
	Moderate				
	113 per 1000	20 per 1000 (2, 160)			
PTSD diagnosis					
Post-treatment – ITT analysis Mini- PTSD Diagnosis Interview (mean 13 weeks)	*Study population*		RR 0.35 (0.1, 1.23)	103 (1 study)	●●○○ Low (a,b)
	170 per 1000	59 per 1000 (17, 209)			
	Moderate				
	170 per 1000	59 per 1000 (17, 209)			
Post-treatment – available case analysis Mini- PTSD Diagnosis Interview (mean 13 weeks)	*Study population*		RR 0.35 (0.1, 1.23)	103 (1 study)	●●○○ Low (a,b)
	170 per 1000	59 per 1000 (17, 209)			
	Moderate				
	170 per 1000	59 per 1000 (17, 209)			

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
PTSD mean scores					
Post-treatment – ITT analysis Mini- PTSD Diagnosis Interview: 'Trauma symptoms', rating scale unclear (mean 13 weeks)	**SMD -0.41 (-0.81, -0.02)**		-	103 (1 study)	⨁⨁◯◯ Low (c)
Post-treatment – available case analysis Mini- PTSD Diagnosis Interview: 'Trauma symptoms', rating scale unclear (mean 13 weeks)	**SMD -0.41 (-0.81, -0.02)**		-	103 (1 study)	⨁⨁◯◯ Low (c)
Evidence Statements: *Individual post-traumatic birth counselling may improve depression symptomatology at endpoint or first measurement (low certainty evidence) compared with treatment as usual in postpartum women with a diagnosis of PTSD.* *Individual post-traumatic birth counselling appears to have no effect on anxiety symptomatology at endpoint or first measurement (low certainty evidence) compared with treatment as usual in postpartum women with a diagnosis of PTSD.* *Individual post-traumatic birth counselling may improve PTSD mean scores (low certainty evidence), but appears to have no effect on PTSD diagnosis (low certainty evidence) at endpoint or first measurement compared with treatment as usual in postpartum women with a diagnosis of PTSD.*					
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. Total number of events is less than 300 (a threshold rule-of-thumb). b. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25). c. Total population size is less than 400 (a threshold rule-of-thumb).					

Source: NICE 2015, Table 139, Table 162, Table 171

Abbreviations: CI, confidence interval; DASS, Depression Anxiety Stress Scale; EPDS, Edinburgh Postnatal Depression Scale; ITT, intention-to-treat; PTSD, post-traumatic stress disorder; RR, relative risk; SMD, standardised mean difference.

Note: Statistically significant differences are shown in bold.

C3.2.7 Post-miscarriage counselling

Only one SR (NICE 2015) was identified for post-miscarriage counselling in the literature search. NICE 2015 included three RCTs that compared post-miscarriage counselling with treatment as usual or enhanced treatment as usual[14]. One RCT from the United States assessed face-to-face nurse-led counselling in postpartum women with symptoms of depression. Another RCT from the United States assessed interpersonal counselling via telephone in postpartum women with symptoms of depression. One RCT from the United Kingdom assessed face-to-face psychological counselling (with medical investigations into causes of miscarriage) in postpartum women with symptoms of anxiety.

[14] Enhanced treatment as usual involved medical investigations into causes of miscarriage without counselling.

C3.2.7.1 Post-miscarriage counselling versus treatment as usual

NICE 2015 did not consider the setting or mode of delivery of the intervention in their analyses.

There was no evidence for clinically or statistically significant benefits associated with post-miscarriage counselling on mean depression symptoms or anxiety mean scores at endpoint or at follow-up (**Table C3-27**).

There was no evidence for clinically or statistically significant effects of post-miscarriage counselling relative to treatment as usual or enhanced treatment as usual on attrition (as a proxy measure for retention in services and treatment acceptability).

Table C3-27 Summary of findings (treatment) – post-miscarriage counselling versus treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression mean scores					
Post-treatment – ITT analysis CES-D or HRSD (7-12 weeks)	SMD 0.17 (-0.12, 0.46)		-	189 (2 studies)	⨁⨁◯◯ Low (a)
Post-treatment – available case analysis HRSD or HADS – Depression (2-7 weeks)	SMD 0.14 (-0.29, 0.58)		-	81 (2 studies)	⨁⨁◯◯ Low (a,b)
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis HADS – Depression (mean 17 weeks)	SMD -0.23 (-0.71, 0.26)		-	66 (1 study)	⨁⨁◯◯ Low (a,b)
Long Follow-up (25-103 weeks post-intervention) – ITT analysis CES-D (mean 46 weeks)	SMD -0.08 (-0.38, 0.22)		-	170 (1 study)	⨁⨁◯◯ Low (a)
Anxiety mean scores					
Post-treatment – available case analysis HADS – Anxiety (mean 2 weeks)	SMD 0.11 (-0.38, 0.59)		-	66 (1 study)	⨁⨁◯◯ Low (a,b)
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis HADS – Anxiety (mean 17 weeks)	SMD -0.31 (-0.8, 0.17)		-	66 (1 study)	⨁⨁◯◯ Low (a,b)
Evidence Statements: *There is that individual post-miscarriage counselling (telephone or face-to-face at home) appears to have no effect on depression mean scores at endpoint or first measurement (low certainty evidence), or on depression mean scores at long follow-up (25-103 weeks post-intervention) (low certainty evidence) compared with treatment as usual in women with symptoms of depression.* *Individual post-miscarriage counselling (face-to-face clinic-based psychological counselling plus medical investigations into causes of miscarriage) appears to have no effect on depression mean scores at intermediate follow-up (17-24 weeks post-intervention) (low certainty evidence) or on anxiety mean scores at intermediate follow-up (17-24 weeks post-intervention) (low certainty evidence) compared with enhanced treatment as usual (medical investigations into causes of miscarriage without counselling) in women with symptoms of anxiety.*					
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. Total population size is less than 400 (a threshold rule-of-thumb) b. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25)					

Source: NICE 2015, Table 138, Table 161

Abbreviations: CES-D, Center for Epidemiologic Studies Depression Scale; CI, confidence interval; HADS, Hospital Anxiety and Depression Scale; HRSD, Hamilton Rating Scale for Depression; ITT, intention-to-treat; RR, relative risk; SMD, standardised mean difference.

C3.3 TREATMENT WITH ONLINE INTERVENTIONS

Two SRs (Ashford 2016; Lee 2016) were identified in the literature search. Ashford 2016 included five RCTs (plus two studies with a quasi-experimental pretest/post-test design) of web- or computer-based interventions for the treatment of mental health problems. With the exception of the RCT of the 'Maternal Depression Online' intervention that was published in a thesis, all other RCTs included in Ashford 2016 were included in NICE 2015 (as self-help or post-miscarriage self-help interventions) and have been included in other sections of the current report (**Section C3.2.5, Section C3.1.8, Section C4.1.8**).[15]

The majority of interventions were developed for the treatment of depression in pregnant women or postpartum women; however, one intervention (with two publications) targeted complicated grief and mental health in women and their partners after pregnancy loss. The therapeutic approach used in the web-based interventions was CBT (three RCTs) or behavioural activation (two RCTs). The majority of interventions also included therapist contact (either on the phone, via email, or in real-time online), which occurred mostly on a weekly basis. In the studies that targeted women and their partners after pregnancy loss, support was provided in the form of written feedback for writing assignments, with assignments personalised by a therapist. The duration of the interventions ranged from 5 weeks to 17 weeks across the studies. Sample attrition between pre- and post-intervention time points was up to 62%.

Across the five RCTs included in Ashford 2016, the comparator was either waitlist control or treatment as usual. *No studies compared an online intervention with an offline version of the same intervention for the treatment of mental health problems.* The authors concluded that computer- or web-based mental health interventions, particularly those targeting depression or complicated grief, may be a promising approach to the treatment and reduction of maternal mental health issues during the perinatal period; however, there are significant gaps in the current evidence base so further research is needed.

A literature search was conducted to identify RCTs of online interventions published after the literature search date of the Ashford 2016 SR. Only those studies that compared an online intervention with an offline version of the same intervention were considered eligible. *No additional studies, published in full, were identified in the literature search update.*

Table C3-28 Summary of findings (treatment) – online interventions

Evidence Statement: *There is no RCT evidence for online interventions compared with offline versions of the same intervention in women who have mental health problems in the perinatal period.*

[15] Of note, Ashford 2016 classified an RCT of a cognitive behaviour web-based intervention (Kersting 2013) as a treatment study (see **Section C3.3**), whereas NICE 2015 classified this study as a preventive intervention (see **Section C4.1.8** on post-miscarriage self-help interventions).

C3.4 TREATMENT WITH PHARMACOLOGICAL INTERVENTIONS

C3.4.1 Antidepressants

Of the eight SRs of antidepressants identified in the scoping and updated searches, NICE 2015 and Molyneaux 2014 were chosen as the foundation reviews due to their currency, high quality and comprehensiveness. It should be noted that both SRs reporting on treatment with antidepressants (NICE 2015; Molyneaux 2014) included the same set of six RCTs, which included one comparing antidepressants as a group with general supportive care and a psychological therapy (listening visits), two comparing SSRIs with placebo, two comparing SSRIs plus psychological therapy with placebo plus psychological therapy,[16] and one comparing SSRIs with TCAs. Rather than choosing a single foundation SR for the assessment of treatment, both reviews are included because they analysed the available data in slightly different ways.

C3.4.1.1 Any antidepressants

Two comparisons were available for the assessment of the efficacy and side effects of treatment with antidepressants as a group – one against general supportive care and one against listening visits – in one identified study (Sharp 2010).

Table C3-29 summarises the evidence and provides Evidence Statements relating to the comparison between antidepressants as a group and general supportive care. The type of antidepressant used was at the discretion of the general practitioner (GP), although use of SSRIs as first-line treatment was encouraged and the majority of women received citalopram, fluoxetine or sertraline. General supportive care was described as women seeing their GP or practice health visitor (PHV) as often as they liked, with no antidepressant prescription from the GP and no listening visit with the PHV, unless the severity of the depression required a protocol deviation. The results of this comparison, based on very low certainty evidence, show that antidepressants may be significantly beneficial compared with general supportive care in terms of increasing remission at 4 weeks post-treatment, reducing depression symptomatology and decreasing depression mean score.

Table C3-30 summarises the evidence and provides Evidence Statements relating to the comparison between antidepressants as a group and listening visits. Listening visits were described as "a psychotherapeutic intervention that uses a form of non-directive counselling, often referred to as 'active listening'." The results of this comparison, based on very low certainty evidence, show that there does not appear to be a difference in remission rate at 4 weeks post-treatment between antidepressants and listening visits.

Table C3-29 Summary of findings (treatment) – antidepressants versus general supportive care

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Remission					
Remission rate at post-treatment	*Study population*		**RR 2.11**	254	●○○○
Post-treatment	176 per 1000	371 per 1000 (239, 577	**(1.36, 3.28)**	(1 RCT)[17]	Very low (a,b)

[16] While listed as included in the Molyneaux 2014 review, one of these (Appleby 1997) did not contribute to any of the presented analyses.
[17] Molyneaux 2014 (Sharp 2010).

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
< 13 EPDS	*Moderate*				
(4 weeks)	NR	NR			
Depression symptomatology					
Depression symptomatology	*Study population*		**RR 0.76**	254	●○○○
Post-treatment – ITT analysis	824 per 1000	626 per 1000 (536, 733)	**(0.65, 0.89)**	(1 RCT)[18]	Very low (c,d)
EPDS > 13	*Moderate*				
(4 weeks)	NR	NR			
Depression symptomatology	*Study population*		**RR 0.68**	218	●○○○
Post-treatment – available case analysis	804 per 1000	546 per 1000 (450, 667)	**(0.56, 0.83)**	(1 RCT)[19]	Very low (c,d)
EPDS > 13	*Moderate*				
(mean 4 weeks)	NR	NR			
Depression mean scores					
Depression mean scores	**SMD -0.48**		-	218	●○○○
Post-treatment – available case analysis	**(-0.75, -0.21)**			(1 RCT)	Very low (c,d)
EPDS					
(4 weeks)					

Evidence Statements:

Treatment with antidepressants may improve remission rate at 4 weeks post-treatment compared with general supportive care, in women with postnatal depression, from a rate of 18% to 37% (very low certainty evidence).

Treatment with antidepressants may improve depression symptomatology at 4 weeks post-treatment compared with general supportive care, in women with postnatal depression, from a rate of 82% to 55% (very low certainty evidence).

Treatment with antidepressants may improve depression mean score at 4 weeks post-treatment compared with general supportive care, in women with postnatal depression (very low certainty evidence).

Footnotes:

* The 'assumed risk' for the study population is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The moderate risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).

a. Downgraded twice due to high risk of bias in two domains (lack of blinding of outcome assessors and low adherence).

b. Downgraded due to imprecision (only one study available for this comparison).

c. High risk of performance bias and only 56% reporting taking antidepressants in intervention group.

d. Total population size is less than 400 (a threshold rule-of-thumb).

Source: Molyneaux 2014 – Analysis 2.1; NICE 2015 – Table 297.

Abbreviations: CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; ITT, intention-to-treat; NR, not reported; PND, postnatal depression; RCT, randomised controlled trial; RR, relative risk; SMD, standardised mean difference.

[18] NICE 2015 (Sharp 2010).

[19] NICE 2015 (Sharp 2010).

Table C3-30 Summary of findings (treatment) – antidepressants versus listening visits

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Remission					
Remission rate at post-treatment	*Study population*		RR 1.04	254	●○○○
Post-treatment	448 per 1000	466 per 1000 (354, 609)	(0.79, 1.36)	(1 RCT)[20]	Very low (a,b)
As defined in individual studies	*Moderate*				
(4 weeks)	NR	NR			
Evidence Statements:					
Treatment with antidepressants appears to have no effect on remission rate at 4 weeks post-treatment compared with treatment with listening visits, in women with postnatal depression (very low certainty evidence).					

Footnotes:

* The 'assumed risk' for the study population is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The moderate risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).

a. Downgraded twice due to high risk of bias in two domains (lack of blinding of outcome assessors and low adherence).

b. Downgraded due to imprecision (only one study available for this comparison).

Source: Molyneaux 2014 – Analysis 3.1.

Abbreviations: CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; ITT, intention-to-treat; NR, not reported; PND, postnatal depression; RCT, randomised controlled trial; RR, relative risk; SMD, standardised mean difference.

C3.4.1.2 SSRIs

Two comparisons were available for the use of SSRIs in postnatal depression – one against placebo and one against TCAs – in two SRs (NICE 2015, Molyneaux 2014), which each included four studies (Hantsoo 2014, Bloch 2012, Yonkers 2008 and Wisner 2006). It should be noted that the study by Bloch 2012 compared SSRIs plus psychological therapy with placebo plus psychological therapy. In the Molyneaux 2014 SR, the findings from the Bloch 2012 study have been included in the 'versus placebo' comparison, while in the NICE 2015 SR they have been included in a separate comparison 'versus placebo plus psychological therapy'. Both comparisons are presented in the SoF tables below

Table C3-31 summarises the evidence and provides Evidence Statements relating to the comparison between SSRIs as a group and placebo. Two of the three RCTs included in the SRs examined sertraline (Hantsoo 2014, Bloch 2012) while the remaining RCT examined paroxetine (Yonkers 2008). The analyses of three RCTs conducted by Molyneaux 2014 provides very low certainty evidence that use of SSRIs may improve response and remission at 6-8 weeks post-treatment compared with placebo. There was also low certainty evidence that SSRIs may significantly reduce mean global severity and improvement scores compared with placebo. There was no difference between SSRIs and placebo in terms of mean depression scores and adverse events; these findings were based on very low certainty evidence and were subject to imprecision. As shown in **Table C3-32**, when the comparison between SSRIs plus psychological therapy and placebo plus psychological therapy was considered separately, based on very low certainty evidence there was no difference in response, remission, global improvement mean

[20] Molyneaux 2014 (Sharp 2010).

score and mean distress score; however, there was low-to-moderate certainty evidence that SSRIs plus psychological therapy may result in greater reduction in mean depression and mean global severity scores.

Table C3-33 summarises the evidence and provides Evidence Statements relating to the comparison between SSRIs and TCAs; however, the single included study compared only one of each class – sertraline and nortriptyline (Wisner 2006). Based on the analyses of data from this study, there was low-to-very low certainty evidence of no difference between SSRIs and TCAs for any of the outcomes assessed, including response score, remission score, global severity and improvement, depression mean score, global assessment of functioning mean score, and social problems.

Table C3-31 Summary of findings (treatment) – SSRIs versus placebo

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Response					
Response rate at post-treatment	*Study population*		**RR 1.43**	146	●○○○
Post-treatment	365 per 1000	522 per 1000 (369, 741)	**(1.01, 2.03)**	(3 RCTs)[21]	Very low (a,b,c)
As defined in individual studies	*Moderate*				
(6-8 weeks)	NR	NR			
Response rate at post-treatment	*Study population*		RR 1.62	106	●○○○
Post-treatment	296 per 1000	480 per 1000 (290, 790)	(0.98, 2.67)	(2 RCTs)[22]	Very low (b,c)
As defined in individual studies	*Moderate*				
(6-8 weeks)	NR	NR			
Non-response to postnatal treatment	*Study population*		RR 0.74	106	●○○○
Post-treatment – ITT[23]	704 per 1000	521 per 1000 (366, 746)	(0.52, 1.06)	(2 RCTs)[24]	Very low (e)
> 10 HRSD, > 50% decrease, improvement on CGI or CGI -I=1 or 2	*Moderate*				
(6 weeks)	711 per 1000	526 per 1000 (370, 754)			
Non-response to postnatal treatment	*Study population*		RR 0.46	33	●●○○
Post-treatment – available case analysis[25]	722 per 1000	332 per 1000 (152, 722)	(0.21, 1.00)	(1 RCT)[26]	Low (d)
> 10 HRSD, > 50% decrease, improvement on CGI	*Moderate*				
(6 weeks)	722 per 1000	332 per 1000 (152, 722)			
Remission					
Remission rate at post-treatment	*Study population*		**RR 1.79**	146	●○○○
Post-treatment	257 per 1000	460 per 1000 (278, 766)	**(1.08, 2.98)**	(3 RCTs)[27]	Very low (a,b,c)
As defined in individual studies	*Moderate*				
(6-8 weeks)	NR	NR			
Remission rate at post-treatment	*Study population*		**RR 2.56**	106	●○○○
Post-treatment	167 per 1000	428 per 1000 (219, 835)	**(1.31, 5.00)**	(2 RCTs)[28]	Very low (b,c)
As defined in individual studies	*Moderate*				
(6-8 weeks)	NR	NR			
Non-remission	*Study population*		**RR 0.70**	106	●○○○
Post-treatment – ITT[23]	883 per 1000	583 per 1000 (450, 758)	**(0.54, 0.91)**	(2 RCTs)[29]	Very low (d,e)
HRSD > 7 or HRSD > 8	*Moderate*				
(6 weeks)	823 per 1000	576 per 1000 (444, 749)			
Non-remission	*Study population*		RR 0.51	33	●●○○
Post-treatment – available case analysis[25]	778 per 1000	397 per 1000 (202, 778)	(0.26, 1.00)		Low (d)

[21] Molyneaux 2014 (Hantsoo 2014, Bloch 2012 and Yonkers 2008). In NICE 2015, the results for Bloch 2012 (SSRIs + psychotherapy versus placebo + psychotherapy) were assessed separately.
[22] Molyneaux 2014 (Hantsoo 2014 and Yonkers 2008). Excludes the Bloch 2012 study (SSRIs + psychotherapy versus placebo + psychotherapy).
[23] Method of ITT unclear.
[24] NICE 2015 (Hantsoo 2014 and Yonkers 2008).
[25] Completers: participants with at least three post-randomisation assessments.
[26] NICE 2015 (Hantsoo 2014).
[27] Molyneaux 2014 (Hantsoo 2014, Bloch 2012 and Yonkers 2008). In NICE 2015, the results for Bloch 2012 (SSRIs + psychotherapy versus placebo + psychotherapy) were assessed separately.
[28] Molyneaux 2014 (Hantsoo 2014 and Yonkers 2008). Excludes the Bloch 2012 study (SSRIs + psychotherapy versus placebo + psychotherapy).
[29] NICE 2015 (Hantsoo 2014 and Yonkers 2008).

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Response					
HRSD > 7 (6 weeks)	*Moderate*			(1 RCT)[30]	
	778 per 1000	397 per 1000 (202, 778)			
Depression mean scores					
Depression mean scores Post-treatment – available case analysis HRSD (6 weeks)	SMD -0.6 (-1.33, 0.12)		-	31 (1 RCT)[31]	●○○○ Very low (e,f)
Global severity and improvement mean scores					
Global severity mean scores Post-treatment – available case analysis CGI (8 weeks)	**SMD -0.9** **(-1.65, -0.16)**		-	31 (1 RCT)[32]	●●○○ Low (d,e)
Adverse events					
Decreased appetite Post-treatment – available case analysis (8 weeks)	*Study population*		RR 1.50 (0.27, 8.43)	70 (1 RCT)[33]	●○○○ Very low (e,f)
	57 per 1000	86 per 1000 (15, 482)			
	Moderate				
	57 per 1000	85 per 1000 (15, 481)			
Diarrhoea Post-treatment – available case analysis (6-8 weeks)	*Study population*		RR 1.02 (0.32, 3.30)	106 (2 RCTs)[34]	●○○○ Very low (e,f)
	93 per 1000	94 per 1000 (30, 306)			
	Moderate				
	84 per 1000	86 per 1000 (27, 277)			
Dizziness Post-treatment – available case analysis (8 weeks)	*Study population*		RR 2.00 (0.54, 7.37)	70 (1 RCT)[35]	●○○○ Very low (e,f))
	86 per 1000	171 per 1000 (46, 632)			
	Moderate				
	86 per 1000	172 per 1000 (46, 632)			
Headache Post-treatment – available case analysis (6-8 weeks)	*Study population*		RR 0.75 (0.37, 1.49)	106 (2 RCTs)[36]	●○○○ Very low (e,f)
	241 per 1000	181 per 1000 (89, 359)			
	Moderate				
	186 per 1000	140 per 1000 (69, 277)			
Nausea Post-treatment – available case analysis (6-8 weeks)	*Study population*		RR 0.97 (0.35, 2.71)	106 (2 RCTs)[37]	●○○○ Very low (e,f)
	111 per 1000	108 per 1000 (39, 301)			
	Moderate				
	86 per 1000	83 per 1000 (30, 233)			

[30] NICE 2015 (Hantsoo 2014).
[31] NICE 2015 (Yonkers 2008).
[32] NICE 2015 (Yonkers 2008).
[33] NICE 2015 (Yonkers 2008).
[34] NICE 2015 (Hantsoo 2014 and Yonkers 2008).
[35] NICE 2015 (Yonkers 2008).
[36] NICE 2015 (Hantsoo 2014 and Yonkers 2008).
[37] NICE 2015 (Hantsoo 2014 and Yonkers 2008).

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Response					
Somnolence Post-treatment – available case analysis (8 weeks)	*Study population*		RR 1.00 (0.32, 3.15)	70 (1 RCT)[38]	⨁◯◯◯ Very low (e,f)
	143 per 1000	143 per 1000 (46, 450)			
	Moderate				
	143 per 1000	143 per 1000 (46, 450)			
Dry mouth Post-treatment – available case analysis (8 weeks)	*Study population*		RR 9.00 (0.5, 161)	70 (1 RCT)[39]	⨁◯◯◯ Very low (e,f)
	0 per 1000	0 per 1000 (0, 0)			
	Moderate				
	0 per 1000	0 per 1000 (0, 0)			

Evidence Statements:

Treatment with an SSRI may improve response rate at 6-8 weeks post-treatment compared with placebo, in women with postnatal depression, from a rate of 37% to 52% (very low certainty evidence).

Treatment with an SSRI may improve remission rate at 6-8 weeks post-treatment compared with placebo, in women with postnatal depression, from a rate of 26% to 46% (very low certainty evidence).

Treatment with an SSRI appears to have no effect on depression mean score at 6 weeks post-treatment compared with placebo, in women with postnatal depression (very low certainty evidence).

Treatment with an SSRI may improve global severity mean score at 6 weeks post-treatment compared with placebo, in women with postnatal depression (very low certainty evidence).

Treatment with an SSRI does not appear to be associated with an increased risk of maternal adverse events at 6-8 weeks post-treatment compared with placebo, in women with postnatal depression (very low certainty evidence).

Footnotes:

* The 'assumed risk' for the study population is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The moderate risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).

a. Downgraded due to indirectness (in one of the studies included in the meta-analysis participants in both arms additionally received brief dynamic psychotherapy).

b. Downgraded due to risk of bias (incomplete outcome data owing to loss to follow-up)

c. Downgraded due to high imprecision (wide confidence intervals owing to the small number and small samples of included studies)

d. Total population size is less than 400 (a threshold rule-of-thumb).

e. Risk of bias due to high attrition.

f. Total population size is less than 400 (a threshold rule-of-thumb) and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25).

Source: Molyneaux 2014 – SoF table (p4-5) and Table 1; NICE 2015 – Table 291, table 294, Table 302 and Table 313.

Abbreviations: CGI, Clinical Global Impression scale; CGI-I, Clinical Global Impression – Improvement scale; CI, confidence interval; HRSD, Hamilton Rating Scale for Depression; ITT, intention-to-treat; PND, postnatal depression; RCT, randomized controlled trial; RR, relative risk; SMD, standardised mean difference; SSRI, selective serotonin reuptake inhibitor.

[38] NICE 2015 (Yonkers 2008).

[39] NICE 2015 (Yonkers 2008).

Table C3-32 Summary of findings (treatment) – SSRIs + psychological interventions versus placebo + psychological interventions

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Response					
Non-response to treatment Post-treatment – ITT analysis[40] MADRS or EPDS > 50% (8 weeks)	*Study population*		RR 0.60 (0.27, 1.32)	42 (1 RCT)[41]	●●○○ Low (a)
	500 per 1000	300 per 1000 (135, 660)			
	Moderate				
	500 per 1000	300 per 1000 (135, 660)			
Remission					
Non-remission to treatment Post-treatment – ITT analysis[40] (8 weeks)	*Study population*		RR 0.64 (0.32, 1.30)	42 (1 RCT)[42]	●●○○ Low (a)
	545 per 1000	349 per 1000 (175, 709)			
	Moderate				
	546 per 1000	349 per 1000 (175, 710)			
Depression mean scores					
Depression mean scores Post-treatment – ITT analysis EPDS (8-12 weeks)	**SMD -0.42 (-0.77, -0.07)**		-	127 (2 RCTs)[43]	●●○○ Low (b,c)
Depression mean scores Post-treatment – available case analysis EPDS (12 weeks)	**SMD -0.56 (-1.07, -0.04)**		-	61 (1 RCT)[44]	●●○○ Low (b,c)
Global severity					
Global severity mean scores Post-treatment – ITT analysis CGI (8 weeks)	**SMD -1.37 (-2.06, -0.67)**		-	40 (1 RCT)[45]	●●●○ Moderate (c)
Global improvement mean scores Post-treatment – ITT CGI (8 weeks)	SMD -0.29 (-0.91, 0.33)		-	40 (1 RCT)[46]	●●○○ Low (d)
Distress					
Distress mean scores Post-treatment – ITT analysis Mental Health Inventory (8 weeks)	SMD -0.15 (-0.77, 0.47)		-	40 (1 RCT)[47]	●●○○ Low (d)

[40] Calculated based on LOCF and WCS for those not included in LOCF.
[41] NICE 2015 (Bloch 2012). Bloch 2012 included in 'versus placebo' analysis in Molyneaux 2014 SR.
[42] NICE 2015 (Bloch 2012). Bloch 2012 included in 'versus placebo' analysis in Molyneaux 2014 SR.
[43] NICE 2015 (Bloch 2012 and Appleby 1997).
[44] NICE 2015 (Appleby 1997).
[45] NICE 2015 (Bloch 2012).
[46] NICE 2015 (Bloch 2012).
[47] NICE 2015 (Bloch 2012).

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Evidence Statement: *Treatment with an SSRI plus a psychological intervention appears to have no effect on response rate at 8 weeks post-treatment compared with placebo plus a psychological intervention, in women with postnatal depression (low certainty evidence).* *Treatment with an SSRI plus a psychological intervention appears to have no effect on remission rate at 8 weeks post-treatment compared with placebo plus a psychological intervention, in women with postnatal depression (low certainty evidence).* *Treatment with an SSRI plus a psychological intervention for postnatal depression may improve depression mean score at 8-12 weeks post-treatment compared with placebo plus a psychological intervention, in women with postnatal depression (low certainty evidence).* *Treatment with an SSRI plus a psychological intervention improves global severity mean score at 8 weeks post-treatment compared with placebo plus a psychological intervention, in women with postnatal depression (moderate certainty evidence).* *Treatment with an SSRI plus a psychological intervention appears to have no effect on distress mean score at 8 weeks post-treatment compared with placebo plus a psychological intervention, in women with postnatal depression (low certainty evidence).*					

Footnotes:

* The 'assumed risk' for the study population is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The moderate risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).

a. Number of events is less than 300 (a threshold rule of thumb) and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25).

b. Risk of bias due to high and unbalanced attrition rate.

c. Total population size is less than 400 (a threshold rule-of-thumb).

d. Total population size is less than 400 (a threshold rule-of-thumb) and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25).

Source: NICE 2015 – Table 292, Table 296 and Table 301.

Abbreviations: CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; ITT, intention-to-treat; LOCF, last observation carried forward; MADRS, Montgomery–Åsberg Depression Rating Scale; RCT, randomized controlled trial; RR, relative risk; SSRI, selective serotonin reuptake inhibitor; WCS, worst case scenario.

Table C3-33 Summary of findings (treatment) – SSRIs versus TCAs

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Response					
Response rate at post-treatment	*Study population*		RR 0.82	109	●○○○
Post-treatment	685 per 1000	562 per 1000 (418, 754)	(0.61, 1.10)	(1 RCT)[48]	Very low (a,b)
As defined in individual study	*Moderate*				
(8 weeks)	NR	NR			
Non-response to treatment	*Study population*		RR 1.39	109	●○○○
Post-treatment – ITT analysis	315 per 1000	438 per 1000 (264, 715)	(0.84, 2.27)	(1 RCT)[49]	Very low (c,d)
HRSD < 50% reduction	*Moderate*				
(8 weeks)	315 per 1000	438 per 1000 (265, 715)			
Non-response to treatment	*Study population*		RR 2.81	29	●○○○
Intermediate follow-up – available case analysis	0 per 1000	0 per 1000 (0, 0)	(0.12, 63.8)	(1 RCT)[49]	Very low (c,d)
HRSD < 50% reduction	*Moderate*				
(22 weeks post-intervention)	0 per 1000	0 per 1000 (0, 0)			
Remission					
Remission rate at post-treatment	*Study population*		RR 0.94	109	●○○○
Post-treatment	481 per 1000	452 per 1000 (303, 678)	(0.63, 1.41)	(1 RCT)[50]	Very low (a,b)
As defined in individual study	*Moderate*				
(8 weeks)	NR	NR			
Non-remission to treatment	*Study population*		RR 1.05	109	●○○○
Post-treatment – ITT analysis	519 per 1000	544 per 1000 (384, 778)	(0.74, 1.50)	(1 RCT)[49]	Very low (c,d)
HRSD > 7	*Moderate*				
(8 weeks)	519 per 1000	545 per 1000 (384, 779)			
Non-remission to treatment	*Study population*		RR 1.24	29	●○○○
Post-treat Intermediate follow-up – available case analysis	214 per 1000	266 per 1000 (73, 986)	(0.34, 4.64)	(1 RCT)[49]	Very low (c,d)
HRSD > 7	*Moderate*				
(22 weeks post-intervention)	214 per 1000	265 per 1000 (73, 984)			
Depression symptomatology					
Depression mean scores	SMD 0.03 (-0.40, 0.47)		-	83	●●○○
Post-treatment - available case analysis				(1 RCT)[49]	Low (c,e)
HRSD					
(8 weeks)					
Depression mean scores	SMD 0.20 (-0.53, 0.93)		-	29	●○○○
Intermediate follow-up - available case analysis				(1 RCT)[51]	Very low (c,d)
HRSD					
(22 weeks post-intervention)					

[48] Molyneaux 2014 (Wisner 2006).
[49] NICE 2015 (Wisner 2006).
[50] Molyneaux 2014 (Wisner 2006).
[51] NICE 2015 (Wisner 2006).

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Global assessment of functioning					
Global assessment of functioning mean score Post-treatment – available case analysis Global Assessment Scale (8 weeks)	SMD 0.06 (-0.38, 0.49)		-	83 (1 RCT)[51]	●●○○ Low (b)
Global assessment of functioning mean score Intermediate follow-up – available case analysis Global Assessment Scale (22 weeks)	SMD 0.03 (-0.69, 0.76)		-	29 (1 RCT)[51]	●●○○ Low (b)
Social problems					
Social problems Post-treatment – available case analysis Social problems questionnaire (8 weeks)	*Study population*		RR 0.91 (0.57, 1.45)	83 (1 RCT)[51]	●●○○ Low (b)
	489 per 1000	445 per 1000 (279, 710)			
	Moderate				
	489 per 1000	445 per 1000 (279, 710)			
Social problems Intermediate follow-up – available case analysis Social problems questionnaire (22 weeks)	*Study population*		RR 0.93 (0.29, 3.03)	29 (1 RCT)[51]	●●○○ Low (b)
	286 per 1000	266 per 1000 (83, 866)			
	Moderate				
	286 per 1000	266 per 1000 (83, 867)			
Global severity and improvement symptomatology					
Global severity and improvement symptomology Post-treatment – available case analysis CGI ≥ 4 (8 weeks)	*Study population*		RR 0.65 (0.06, 6.92)	83 (1 RCT)[51]	●●○○ Low (b)
	43 per 1000	28 per 1000 (3, 294)			
	Moderate				
	43 per 1000	28 per 1000 (3, 298)			

Evidence Statements:

Treatment with a SSRI appears to have no effect on response rate at 8 weeks or up to 22 weeks post-treatment compared with treatment with TCAs, in women with postnatal depression (very low certainty evidence).

Treatment with an SSRI appears to have no effect on remission rate at 8 weeks or up to 22 weeks post-treatment compared with treatment with TCAs, in women with postnatal depression (very low certainty evidence).

Treatment with an SSRI appears to have no effect on depression means at 8 weeks or up to 22 weeks post-treatment compared with treatment with TCAs, in women with postnatal depression (low certainty evidence).

Treatment with an SSRI appears to have no effect on global assessment of functioning means score at 8 weeks or up to 22 weeks post-treatment compared with treatment with TCAs, in women with postnatal depression (low certainty evidence).

Treatment with an SSRI appears to have no effect on social problems at 8 weeks or up to 22 weeks post-treatment compared with treatment with TCAs, in women with postnatal depression (low certainty evidence).

Treatment with an SSRI appears to have no effect on global severity and improvement symptomatology at 8 weeks post-treatment compared with treatment with TCAs, in women with postnatal depression (low certainty evidence).

Footnotes:

* The 'assumed risk' for the study population is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The moderate risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).

a. Downgraded due to risk of bias (incomplete outcome data owing to loss to follow-up)

b. Downgraded due to imprecision (only 1 study available for this comparison).

c. Risk of bias due to incomplete outcome data (discontinuation between groups unbalanced).

d. Total population size is less than 400 (a threshold rule-of-thumb) and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25).

e. Total population size is less than 400 (a threshold rule-of-thumb).

Source: Molyneaux 2014 – Table 3; NICE 2015 – Table 293, Table 205 and Table 300.

C3.4.2 Antipsychotics

No SRs or individual RCTs were identified that assessed the effect of antipsychotics on the treatment of antenatal or postnatal mental health problems or maternal side effects.

Table C3-34 Summary of findings (treatment) – antipsychotics

Evidence Statement: *There is no RCT evidence for antipsychotics as an intervention for women with mental health problems in the perinatal period.*

C3.4.3 Anticonvulsants

No SRs or individual RCTs were identified that assessed the effect of anticonvulsants on the treatment of antenatal or postnatal mental health problems or maternal side effects.

Table C3-35 Summary of findings (treatment) – anticonvulsants

Evidence Statement: *There is no RCT evidence for anticonvulsants as an intervention for women with mental health problems in the perinatal period.*

C3.4.4 Benzodiazepines and z-drugs

No SRs or individual RCTs were identified that assessed the effect of benzodiazepines and z-drugs on the treatment of antenatal or postnatal mental health problems or maternal side effects.

Table C3-36 Summary of findings (treatment) – benzodiazepines and z-drugs

Evidence Statement: *There is no RCT evidence for benzodiazepines and z-drugs as an intervention for women with mental health problems in the perinatal period.*

C3.4.5 Lithium

No SRs or individual RCTs were identified that assessed the effect of lithium on the treatment of antenatal or postnatal mental health problems or maternal side effects.

Table C3-37 Summary of findings (treatment) – lithium

Evidence Statement: *There is no RCT evidence for lithium as an intervention for women with mental health problems in the perinatal period.*

C3.5 TREATMENT WITH COMPLEMENTARY THERAPIES

C3.5.1 Omega-3 fatty acids

One comparison was available for the use of omega-3 fatty acids in perinatal depression in the foundation review by NICE 2015 – against placebo.

Table C3-38 summarises the evidence and provides Evidence Statements relating to the comparison between omega-3 fatty acids and placebo. Three of the four RCTs included in the NICE 2015 SR examined different ratios of EPA and DHA (Mozurkewich 2013 [0.2:1], Freeman 2008 [1.4:1] and Su 2008 [1.8:1]), while the remaining study by Rees 2008 examined 'fish oil'. The analyses presented in NICE 2015 provide very low certainty evidence that treatment with omega- fatty acids does not appear to increase remission, reduce depression scores, or cause mild/transient side effects compared with placebo in women with perinatal depression. While there is also very-low-certainty evidence that treatment with omega-3 fatty acids does not increase response compared with placebo, the reduction in risk is substantial and the finding is close to being statistically significant (P=0.09).

Table C3-38 Summary of findings (treatment) – omega-3 fatty acids versus placebo

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Response					
Non-response to treatment Post treatment – ITT analysis HRSD < 50% reduction (8 weeks)	*Study population*		RR 0.67 (0.42, 1.06)	36 (1 RCT)[52]	●○○○ Very low (a,b)
	833 per 1000	558 per 1000 (350, 883)			
	Moderate				
	833 per 1000	558 per 1000 (350, 883)			
Non-response to treatment Post-treatment – available case analysis HRSD < 50% reduction (8 weeks)	*Study population*		RR 0.53 (0.24, 1.15)	24 (1 RCT)[52]	●○○○ Very low (a,b)
	727 per 1000	385 per 1000 (175, 836)			
	Moderate				
	727 per 1000	385 per 1000 (174, 836)			
Remission					
Non-remission to treatment Post-treatment – ITT analysis HRSD > 7 (8 weeks)	*Study population*		RR 0.81 (0.58, 1.13)	36 (1 RCT) [52]	●○○○ Very low (a,b)
	889 per 1000	720 per 1000 (516, 1000)			
	Moderate				
	889 per 1000	720 per 1000 (516, 1000)			

[52] NICE 2015 (Su 2008).

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Non-remission to treatment Post-treatment – available case analysis HRSD > 7 (8 weeks)	*Study population*		RR 0.75 (0.45, 1.26)	24 (1 RCT) [52]	●○○○ Very low (a,b)
	818 per 1000	614 per 1000 (368, 1000)			
	Moderate				
	818 per 1000	614 per 1000 (368, 1000)			
Depression mean scores					
Depression mean scores Post treatment – ITT analysis EPDS or BDI (6-36 weeks)	SMD -0.08 (-0.61, 0.46)		-	228 (4 RCTs)[53]	●○○○ Very low (a,b,c)
Adverse events					
Any mild/transient side effects Post-treatment – available case analysis (6-8 weeks)	*Study population*		RR 1.15 (0.64, 2.06)	118 (3 RCTs) [54]	●●○○ Low (b)
	246 per 1000	282 per 1000 (157, 506)			
	Moderate				
	NR	NR			

Evidence Statements:

Treatment with omega-3 fatty acids appears to have no effect on response rate at 8 weeks post-treatment compared with placebo, in women with antenatal or postnatal depression (very low certainty evidence).

Treatment with omega-3 fatty acids appears to have no effect on remission rate at 8 weeks post-treatment compared with placebo, in women with antenatal or postnatal depression (very low certainty evidence).

Treatment with omega-3 fatty acids appears to have no effect on depression mean score at 6-36 weeks post-treatment compared with placebo, in women with antenatal or postnatal depression (very low certainty evidence).

Treatment with omega-3 fatty acids does not appear to be associated with an increased risk of mild/transient side effects at 6-8 weeks post-treatment compared with placebo, in antenatal or postnatal depression (very low certainty evidence).

Footnotes:

* The 'assumed risk' for the study population is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The moderate risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).

a. Risk of bias due to unclear selection bias, detection bias and attrition bias.

b. Total population size is less than 400 (a threshold rule-of-thumb) and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25).

c. There was evidence of substantial heterogeneity between effect sizes

Source: NICE 2015 – Table 290 and Table 312.

Abbreviations: CI, confidence interval; BDI, Beck Depression Inventory; EPDS, Edinburgh Postnatal Depression Scale; HRSD, Hamilton Rating Scale for Depression; PND, postnatal depression; RCT, randomised controlled trial; RR, relative risk; SMD, standardised mean difference.

C3.5.2 St John's wort

No SRs or individual RCTs were identified that assessed the effectiveness of St John's wort on the treatment of antenatal or postnatal mental health problems.

[53] NICE 2015 (meta-analysis of data from Rees 2008, Su 2008, Freeman 2008 and Mozurkewich 2013).

[54] NICE 2015 (meta-analysis of data from Rees 2008, Su 2008 and Freeman 2008).

Table C3-39 Summary of findings (treatment) – St John's wort

Evidence Statement:
There is no RCT evidence for St John's wort as an intervention for women with mental health problems in the perinatal period.

C3.5.3 Gingko biloba

No SRs or individual RCTs were identified that assessed the effectiveness of gingko biloba on the treatment of antenatal or postnatal mental health problems.

Table C3-40 Summary of findings (treatment) – gingko biloba

Evidence Statement:
There is no RCT evidence for gingko biloba as an intervention for women with mental health problems in the perinatal period.

C3.6 TREATMENT WITH PHYSICAL INTERVENTIONS

C3.6.1 Exercise

Of the three SRs identified in the literature search, NICE 2015 was chosen as the foundation review due to its currency and high quality. NICE 2015 included three RCTs that compared physical activity with treatment as usual, and one RCT that compared physical activity with a mutual support group. The NICE 2015 SR did not separate out different forms of physical activity (i.e. yoga was analysed with other types of exercise interventions).

Of the three RCTs that compared physical activity with treatment as usual, there were two RCTs from the United Kingdom that examined the effect of exercise consultations on postpartum women. One RCT (N=38) assessed individual exercise consultations in the home setting (with follow-up support calls) in women with symptoms of depression (>12 on the EPDS), while the other RCT (N=94) assessed individual and group exercise consultations with support follow-up calls in women with a diagnosis of MDD. A third RCT from the United States (N=92) compared group tai-chi/yoga, specifically designed for women in their second and third trimester of pregnancy, with waitlist control in pregnant women who met diagnostic criteria for depression.

NICE 2015 also included an Australian RCT that compared a 12-week group pram walking exercise program with a 12-week mutual support group in postpartum women with symptoms of depression (N=24). The mutual support group was facilitated by a nurse/social worker and involved unstructured discussion for social and emotional but not practical support.

C3.6.1.1 Physical interventions versus treatment as usual

NICE 2015 did not consider the timing or format of the intervention in their analyses.

There was no evidence for a statistically or clinically meaningful effect of physical activity on mean depression scores at the end of the intervention, although the effect favoured physical activity compared with control (**Table C3-41**). There was no statistically or clinically significant effect of physical activity on mean anxiety scores.

Table C3-41 Summary of findings (treatment) – physical activity versus treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression mean scores					
Post-intervention, first available endpoint data – available case analysis Follow-up: 12-26 weeks	SMD -0.23 (-0.52, 0.05)		-	191 (3 studies)	⨁⨁◯◯ Low (a,b)
Anxiety mean scores					
Post-treatment (0-9 weeks) – available case analysis	SMD 0.18 (-0.27, 0.63)		-	75 (1 study)	⨁◯◯◯ Very low (a,b)
Evidence Statements: *Physical activity (individual and group exercise consultations or Tai Chi/yoga) appears to have no effect on depression mean scores at endpoint or first measurement (low certainty evidence) compared with treatment as usual in pregnant or postpartum women who have a diagnosis of depression or symptoms of depression.* *Group physical activity (Tai Chi/yoga) appears to have no effect on anxiety mean scores at endpoint or first measurement (very low certainty evidence) compared with waitlist control in pregnant women who met diagnostic criteria for depression.*					
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. Unclear risk of bias in several domains b. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5) or RR 0.75/1.25 and optimal information size (400 participants) not met.					

Source: NICE 2015, Table 349

Abbreviations: CI, confidence interval; RR, relative risk; SMD, standardised mean difference.

C3.6.1.2 Physical activity versus mutual support

There was very low quality, single-study evidence for a large beneficial effect of physical activity compared with mutual support on mean depression scores at post-treatment and at short-term follow-up (**Table C3-42**). However, the confidence in this estimate was very low due to serious imprecision (very small population size) and risk of bias in several domains.

Table C3-42 Summary of findings (treatment) – physical activity versus mutual support

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression mean scores					
Post-treatment, 0-9 weeks – available case analysis	**SMD -1.05 (-2.02, -0.07)**		-	19 (1 study)	●○○○ Very low (a,b)
Short-term follow-up, 9-16 weeks – available case analysis	**SMD -1.09 (-2.07, -0.11)**		-	19 (1 study)	●○○○ Very low (a,b)
Evidence Statement: *Physical activity (pram walking exercise program) may improve depression mean scores at endpoint or first measurement (very low certainty evidence), and at short follow-up (9-16 weeks post-intervention) (very low certainty evidence) compared with mutual support group in postpartum women with symptoms of depression.*					

Footnotes:
* The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).
a. Risk of bias in several domains
b. Optimal information size for dichotomous outcomes (300 events) and for continuous outcomes (400 participants) not met.

Source: NICE 2015, Table 349
Abbreviations: CI, confidence interval; RR, relative risk; SMD, standardised mean difference.

C3.6.2 Yoga

Of the two SRs identified in the literature search, Gong 2015 was chosen as the foundation review due to its currency and comprehensiveness. Gong 2015 included four RCTs that assessed yoga in pregnant women with a diagnosis of depression. Three RCTs (N=200) compared the effect of 12 weeks of exercise-based yoga with a control group that involved massage and standard prenatal care, parenting education sessions, or a social support group. The fourth RCT (N=92) compared 12 weeks of integrated yoga (yoga with Tai Chi) with a social support group.

One RCT that was not included in Gong 2015 was identified in another SR. The RCT (Vieten 2008) assessed an eight-week mindfulness intervention that incorporated experiential exercises and was facilitated by a licensed clinical psychologist trained in mindfulness-based interventions, as well as a certified prenatal yoga instructor. This RCT, which reported no benefits of the intervention on depression or anxiety symptoms, is included in the current report as a mindfulness intervention (see **Section C3.1.12**.)

C3.6.2.1 Yoga versus control group

Overall, there was a statistically significant reduction in depressive symptoms in the yoga group compared with control (**Table C3-43**); however, this reduction may not be clinically significant. Results of the subgroup analyses showed that exercise-based yoga did not significantly reduce depression symptoms relative to control, whereas integrated yoga appeared to have a beneficial effect (based on one RCT).

Table C3-43 Summary of findings (treatment) – yoga versus control group

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)**
	Assumed risk Control	Corresponding risk Intervention			
Depression mean scores					
Post-treatment – available case analysis All studies of women with depressive symptoms CES-D	**SMD -0.46 (-0.90, -0.03)**		-	208 (4 studies)	●○○○ Very low
Post-treatment – available case analysis Subgroup analysis – exercise-based yoga CES-D	SMD -0.41 (-1.01, 0.18)		-	159 (3 studies)	●○○○ Very low
Post-treatment – available case analysis Subgroup analysis – integrated yoga CES-D	**SMD -0.64 (-1.11, -0.18)**		-	75 (1 study)	●○○○ Very low
Evidence Statements: *Exercise-based yoga appears to have no effect on depression mean scores at endpoint or first measurement (very low certainty evidence) compared with a control group (massage and standard prenatal care, parenting education sessions, or a social support group) in pregnant women with a diagnosis of depression.* *Integrated yoga (with Tai Chi) may improve depression mean scores at endpoint or first measurement (very low certainty evidence) compared with a social support group in pregnant women with a diagnosis of depression.*					
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). ** Risk of bias was assessed in Gong 2015 and was translated for the purposes of this report into a GRADE quality of the evidence rating. For each quality criteria rated by Gong 2015 as 'unclear', the evidence was downgraded.					

Source: Gong 2015, Figure 3, Figure 4

Abbreviations: CI, confidence interval; RR, relative risk; SMD, standardised mean difference.

C3.6.3 Acupuncture

Of the three SRs identified in the literature search, NICE 2015 was chosen as the foundation review due to its currency, high quality and comprehensiveness. NICE 2015 included two RCTS from the United States (N=210) that compared depression-specific acupuncture with non-depression specific acupuncture and with massage in pregnant women with a diagnosis of MDD. NICE 2015 also included one RCT from China (Chung 2012; N=20) that compared electro-acupuncture with non-invasive sham acupuncture in postpartum women with a diagnosis of MDD.

C3.6.3.1 Acupuncture versus massage

There was no statistically or clinically significant difference in effect for acupuncture compared with massage on depression outcomes at post-treatment (**Table C3-44**).

There was no statistically or clinically significant difference in effect for acupuncture (depression and non-depression specific acupuncture combined) compared with massage on mean depression scores at post-treatment or short term follow-up. There was very low certainty evidence for a moderate beneficial effect of acupuncture compared with massage on depression diagnosis at short term follow-up; however, this was not statistically significant and the confidence in the estimate of the effect is low due to very serious imprecision.

Table C3-44 Summary of findings (treatment) – acupuncture versus massage

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Non-response to treatment					
HRSD >=14 and >=50% reduction from baseline Post-treatment (0-8 weeks)	*Study population*		RR 0.8 (0.54, 1.19)	188 (2 studies)	⨁◯◯◯ Very low (a,b)
	442 per 1000 (298, 657)	355 per 1000 (224, 562)			
	Moderate				
	466 per 1000 (315, 694)	379 per 1000 (239, 600)			
Depression diagnosis					
Above depression threshold (DSM-IV) Short term follow-up (9-16 weeks) – available case analysis	*Study population*		RR 0.44 (0.09, 2.13)	46 (1 study)	⨁◯◯◯ Very low (a,b)
	286 per 1000	71 per 1000 (9, 660)			
	Moderate				
	286 per 1000	72 per 1000 (9, 661)			
Depression mean scores					
Post-treatment (0-8 weeks) – available case analysis	SMD 0.19 (-0.47, 0.85)		-	54 (1 study)	⨁◯◯◯ Very low (a,b)
Short term follow-up (9-16 weeks) – available case analysis	SMD -0.16 (-0.77, 0.45)		-	49 (1 study)	⨁◯◯◯ Very low (a,b)

Evidence Statements:

Acupuncture appears to have no effect on response to treatment (measured using the HRSD) at endpoint or first measurement (very low certainty evidence), compared with massage in pregnant women with a diagnosis of MDD.

Acupuncture appears to have no effect on depression diagnosis at short follow-up (9-16 weeks post-intervention) (very low certainty evidence), and appears to have no effect on depression mean scores at endpoint or first measurement (very low certainty evidence), or at short follow-up (9-16 weeks post-intervention) (very low certainty evidence), compared with massage in pregnant women with a diagnosis of MDD.

Footnotes:

* The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).

a. Risk of bias in several domains

b. Optimal information size (for dichotomous outcomes, OIS=300 events; for continuous outcomes, OIS=400 participants) not met

Source: NICE 2015, Table 346, Table 351

Abbreviations: CI, confidence interval; MDD, major depressive disorder; RR, relative risk; SMD, standardised mean difference.

C3.6.3.2 Depression-specific acupuncture versus non-depression-specific acupuncture

There was very low certainty evidence from two studies for a moderate beneficial effect of depression-specific acupuncture post-treatment; however, the confidence in this estimate was very low due to serious imprecision and risk of bias in several domains (**Table C3-45**).

There was no statistically or clinically significant difference between depression-specific acupuncture and non-depression-specific acupuncture on mean depression scores at post-treatment or short-term follow-up. There was very low quality, single-study evidence for a moderate-to-large effect in the favour of depression-specific acupuncture on depression diagnosis at the end of intervention and at short term follow-up; however, these effects were not statistically significant and confidence in this estimate is very low due to very serious imprecision.

Table C3-45 Summary of findings (treatment) – depression-specific acupuncture versus non-depression-specific acupuncture

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Treatment non-response					
HRSD >=14 and >=50% reduction from baseline	*Study population*		**RR 0.59 (0.4, 0.88)**	121 (2 studies)	●○○○ Very low (a,b)
	593 per 1000	350 per 1000 (237, 522)			
	Moderate				
	576 per 1000	340 per 1000 (230, 507)			
Depression diagnosis					
Post-treatment (0-8 weeks) – available case	*Study population*		RR 0.47 (0.11, 2.13)	35 (1 study)	●○○○ Very low (a,b)
	263 per 1000	124 per 1000 (29, 561)			
	Moderate				
	263 per 1000	124 per 1000 (29, 560)			
Short term follow-up (9-16 weeks) – available case	*Study population*		RR 0.64 (0.06, 6.39)	32 (1 study)	●○○○ Very low (a,b)
	111 per 1000	71 per 1000 (7, 710)			
	Moderate				
	111 per 1000	71 per 1000 (7, 709)			
Depression mean scores					
Post-treatment (0-8 weeks) – available case	SMD -0.38 (-1.06, 0.29)		-	35 (1 study)	●○○○ Very low (a,b)
Short term follow-up (9-16 weeks) – available case	SMD -0.12 (-0.82, 0.57)		-	32 (1 study)	●○○○ Very low (a,b)

Evidence Statements:

Depression-specific acupuncture may improve response to treatment (measured using the HRSD) at endpoint or first measurement (very low certainty evidence) compared with non-depression-specific acupuncture in pregnant women with a diagnosis of MDD.

Depression-specific acupuncture appears to have no effect on depression diagnosis (very low certainty evidence) or depression mean scores (very low certainty evidence) at endpoint or first measurement, or at short follow-up (9-16 weeks post-intervention) compared with non-depression-specific acupuncture in pregnant women with a diagnosis of MDD.

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. Risk of bias in several domains b. Optimal information size (for dichotomous outcomes, OIS=300 events; for continuous outcomes, OIS=400 participants) not met.					

Source: NICE 2015, Table 347, Table 352

Abbreviations: CI, confidence interval; HRSD, Hamilton Rating Scale for Depression; MDD, major depressive disorder; RR, relative risk; SMD, standardised mean difference.

C3.6.3.3 Electroacupuncture versus non-invasive sham acupuncture

There was no statistically or clinically significant effect for electroacupuncture on mean depression scores or mean anxiety scores at post-treatment (**Table C3-46**).

Table C3-46 Summary of findings (treatment) – electroacupuncture versus non-invasive sham acupuncture

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression mean scores					
Post-treatment (0-8 weeks) – available case analysis	SMD -0.21 (-1.09, 0.67)			20 (1 study)	●○○○ Very low (a,b)
Anxiety mean scores					
Available case analysis	SMD -0.02 (-0.9, 0.85)			20 (1 study)	●○○○ Very low (a,b)
Evidence Statement: *Electroacupuncture appears to have no effect on depression mean scores at endpoint or first measurement (very low certainty evidence) compared with non-invasive sham acupuncture in postpartum women with a diagnosis of MDD.*					
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. Risk of bias in several domains b. Optimal information size (for dichotomous outcomes, OIS=300 events; for continuous outcomes, OIS=400 participants) not met.					

Source: NICE 2015, Table 353, Table 358

Abbreviations: CI, confidence interval; MDD, major depressive disorder; SMD, standardised mean difference.

C3.6.4 Electroconvulsive therapy

No SRs or individual RCTs were identified that assessed the effect of ECT on the treatment of antenatal or postnatal mental health problems or maternal side effects.

Table C3-47 Summary of findings (treatment) – electroconvulsive therapy

Evidence Statement: *There is no RCT evidence for ECT as an intervention for women with mental health problems in the perinatal period.*

C3.6.5 Transcranial magnetic stimulation

No SRs or individual RCTs were identified that assessed the effect of TMS on the treatment of antenatal or postnatal mental health problems or maternal side effects.

Table C3-48 Summary of findings (treatment) – transcranial magnetic stimulation

Evidence Statement: *There is no RCT evidence for TMS as an intervention for women with mental health problems in the perinatal period.*

C4 RESULTS – PREVENTION

For the majority of psychosocial, psychological, physical and pharmacological interventions, the National Institute for Health and Care Excellence Clinical Guideline Number 192 (NICE 2015) was chosen as the foundation review, primarily due to its currency, comprehensiveness, and high quality. The EWG agreed that the current Review would reproduce the Summary of Findings (SoF) tables from NICE 2015 and that replication of data extraction tables, risk-of-bias assessment, or Evidence Profile (EP) tables was not required. Readers are referred to the NICE guideline CG192 and appendices for this information.

NICE 2015 did not specifically assess evidence relating to online (web-based or computer-based) interventions. As such, the current Evidence Review relied on other published SRs, together with a literature search update to identify recent RCTs not included in the published reviews. For all relevant RCTs relating to online interventions, a full assessment of the evidence was required, including data extraction, risk-of-bias assessment, and EP tables; these are available in the **Part C Appendix**. Evidence from additional SRs was used for cognitive behaviour and interpersonal therapies (CBT/IPT; prevention) and omega-3 fatty acids (prevention); the quality of the SR was assessed, the SR evidence was transcribed directly into an SoF table, and the certainty of the body of evidence was informed by the SR assessment of risk of bias of the individual studies.

C4.1 PREVENTION WITH PSYCHOSOCIAL INTERVENTIONS

C4.1.1 Psychoeducation

Of the three SRs of prevention using psychoeducation identified in the literature search, NICE 2015 was chosen as the foundation review due to its currency, high quality and comprehensiveness. NICE 2015 included three RCTs that compared psychologically informed psychoeducation interventions with treatment as usual or enhanced treatment as usual[55]. The intervention was administered antenatally in two RCTs, and both antenatally and postnatally in one RCT. In all RCTs, the intervention involved five or six sessions of face-to-face psychoeducation delivered to individuals (one RCT; setting not reported), a hospital group setting (one RCT), or both individuals and groups (one RCT; setting not reported).

Although the other SRs included several additional RCTs, these studies mainly assessed education on preparing for parenting and are captured in the section on non-mental-health-focused education and support (**Section C4.1.5**).

C4.1.1.1 Psychologically (CBT/IPT) informed psychoeducation versus treatment as usual or enhanced treatment as usual

NICE 2015 did not consider the timing of the intervention, format or mode of delivery in their analyses.

The evidence for psychologically (CBT/IPT) informed psychoeducation as a preventive intervention for women at-risk of developing postnatal depression was inconsistent (**Table C4-1**). There was evidence from three studies for moderate-to-large effects of psychoeducation on preventing depression diagnosis (P=0.08); however, the confidence in this effect estimate is low due to serious imprecision. This effect was also not maintained at intermediate follow-up. In addition, no

[55] Enhanced treatment as usual involved non-mental health-focused education and support in the form of a booklet.

clinically or statistically significant preventive effects were observed on depression symptomatology or depression mean symptoms at endpoint or intermediate follow-up.

Although there was evidence of higher attrition at endpoint (follow-up 26-27 weeks) in the intervention group relative to treatment as usual or enhanced treatment as usual, this effect was not statistically significant due to very serious imprecision.

Table C4-1 Summary of findings (prevention) – psychologically (CBT/IPT) informed psychoeducation versus treatment as usual or enhanced treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression diagnosis					
Post-treatment – ITT analysis (at-risk populations) SCAN, SCID or Structured Clinical Interview for Childhood Diagnoses (KID-SCID) (mean 27 weeks)	*Study population*		RR 0.69 (0.45, 1.05)	360 (3 studies)	●●○○ Low (a,b)
	229 per 1000	158 per 1000 (103, 241)			
	Moderate				
	333 per 1000	230 per 1000 (150, 350)			
Post-treatment – available case analysis (at-risk populations) SCAN, SCID or KIDSCID (mean 27 weeks)	*Study population*		RR 0.48 (0.23, 1.01)	320 (3 studies)	●●○○ Low (a,b)
	132 per 1000	63 per 1000 (30, 133)			
	Moderate				
	227 per 1000	109 per 1000 (52, 229)			
Intermediate follow-up (17-24 weeks post-intervention) – ITT analysis (at-risk populations) SCID (mean 20 weeks)	*Study population*		RR 0.77 (0.33, 1.75)	45 (1 study)	●●○○ Low (a,b)
	381 per 1000	293 per 1000 (126, 667)			
	Moderate				
	381 per 1000	293 per 1000 (126, 667)			
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk populations) SCID (mean 20 weeks)	*Study population*		RR 0.64 (0.17, 2.46)	37 (1 study)	●●○○ Low (a,b)
	235 per 1000	151 per 1000 (40, 579)			
	Moderate				
	235 per 1000	150 per 1000 (40, 578)			
Depression symptomatology					
Post-treatment – ITT analysis (at-risk populations) EPDS ≥11/12 (mean 27 weeks)	*Study population*		RR 0.85 (0.58, 1.25)	254 (2 studies)	●●○○ Low (a,b)
	299 per 1000	254 per 1000 (174, 374)			
	Moderate				
	370 per 1000	315 per 1000 (215, 462)			
Post-treatment – available case analysis (at-risk populations) EPDS ≥11/12 (mean 27 weeks)	*Study population*		RR 0.88 (0.49, 1.57)	221 (2 studies)	●●○○ Low (a,b)
	183 per 1000	161 per 1000 (90, 288)			
	Moderate				
	171 per 1000	150 per 1000 (84, 268)			
Intermediate follow-up (17-24 weeks post-intervention) – ITT analysis (at-risk populations) EPDS >12	*Study population*		RR 1.17 (0.62, 2.2)	45 (1 study)	●●○○ Low (a,b)
	429 per 1000	501 per 1000 (266, 943)			
	Moderate				

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
(mean 20 weeks)	429 per 1000	502 per 1000 (266, 944)			
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk populations) EPDS >12 (mean 20 weeks)	*Study population*		RR 1 (0.24, 4.18)	30 (1 study)	⬤⬤◯◯ Low (a,b)
	200 per 1000	200 per 1000 (48, 836)			
	Moderate				
	200 per 1000	200 per 1000 (48, 836)			
Depression mean scores					
Post-treatment – available case analysis (at-risk populations) EPDS	SMD -0.06 (-0.75, 0.62)		-	33 (1 study)	⬤⬤◯◯ Low (a)
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk populations) EPDS (mean 20 weeks)	SMD -0.02 (-0.74, 0.7)		-	30 (1 study)	⬤⬤◯◯ Low (a,b)
Evidence Statements: ***CBT/IPT-informed psychoeducation*** *Psychologically (CBT/IPT) informed psychoeducation (individual, face-to-face) may have an effect[56] on depression diagnosis (low certainty evidence) but does not change depression symptomatology (low certainty evidence) at endpoint or first measurement compared with treatment as usual or enhanced treatment as usual in women who are considered to be 'at risk' of developing mental health problems in the perinatal period.* ***IPT-informed psychoeducation*** *IPT-informed psychoeducation (individual, face-to-face) appears to have no effect on depression diagnosis (low certainty evidence) or depression symptomatology (low certainty evidence) at intermediate follow-up (17-24 weeks post-intervention) compared with treatment as usual in women who are considered to be 'at risk' of developing mental health problems in the perinatal period.* *IPT-informed psychoeducation (individual, face-to-face) appears to have no effect on depression mean scores at endpoint or first measurement (low certainty evidence), or at intermediate follow-up (17-24 weeks post-intervention) (low certainty evidence), compared with treatment as usual in women who are considered to be 'at risk' of developing mental health problems in the perinatal period.*					
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. Total number of events is less than 300 (a threshold rule-of-thumb) b. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25)					

Source: NICE 2015, Table 40

Abbreviations: CBT, cognitive behaviour therapy; CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; IPT, interpersonal psychotherapy; ITT, intention-to-treat; KID-SCID, Structured Clinical Interview for Childhood Diagnoses; RR, relative risk; SCAN, Schedules for Clinical Assessment in Neuropsychiatry; SCID, Structured Clinical Interview for DSM Disorders; SMD, standardised mean difference.

C4.1.2 Psychoeducational booklet

Of the two SRs of prevention using psychoeducational booklet interventions identified in the literature search, NICE 2015 was chosen as the foundation review because individual study data was extracted and a meta-analysis was undertaken. NICE 2015 included two large RCTs from Australia and the United States that

[56] RR 0.69 (95% CI 0.45, 1.05); P=0.08

compared a psychoeducational booklet on PND with treatment as usual or enhanced treatment as usual[57]. In the Australian RCT the psychoeducational booklet was provided in hospital to pregnant women with psychosocial risk factors and a family history of mental health problems. In the RCT from the US, the intervention was a psychoeducational booklet and telephone support given to postpartum women with psychosocial risk factors.

One additional RCT from Taiwan was not included in NICE 2015 but the study was relatively small in terms of sample size (N=70). The authors of this study concluded that informational support given to women in the postnatal period may contribute to psychological wellbeing.

C4.1.2.1 Psychoeducational booklet versus treatment as usual or enhanced treatment as usual

NICE 2015 did not consider the timing of the intervention in their analyses.

Moderate-to-low certainty evidence from up to two studies does not provide convincing evidence that a psychoeducational booklet prevents depression symptomatology (**Table C4-2**).

Table C4-2 Summary of findings (prevention) – psychoeducational booklet versus treatment as usual or enhanced treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression symptomatology					
Post-treatment – ITT analysis (at-risk populations) EPDS ≥10/12 (mean 3 weeks)	*Study population*		RR 0.9 (0.79, 1.03)	1,140 (2 studies)	●●●○ Moderate (a)
	419 per 1000	377 per 1000 (331, 431)			
	Moderate				
	409 per 1000	368 per 1000 (323, 421)			
Post-treatment – available case analysis (at-risk populations) EPDS ≥10/12 (mean 3 weeks)	*Study population*		RR 0.73 (0.51, 1.06)	838 (2 studies)	●○○○ Very low (a,b,c)
	208 per 1000	152 per 1000 (106, 220)			
	Moderate				
	218 per 1000	159 per 1000 (111, 231)			
Short Follow-up (9-16 weeks post-intervention) – ITT analysis (at-risk populations) EPDS ≥10 (mean 13 weeks)	*Study population*		RR 0.88 (0.64, 1.23)	540 (1 study)	●●○○ Low (a,b)
	222 per 1000	196 per 1000 (142, 273)			
	Moderate				
	222 per 1000	195 per 1000 (142, 273)			
Short Follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) EPDS ≥10 (mean 13 weeks)	*Study population*		RR 0.64 (0.38, 1.08)	479 (1 study)	●●○○ Low (b,c)
	132 per 1000	85 per 1000 (50, 143)			
	Moderate				
	132 per 1000	84 per 1000 (50, 143)			
Intermediate follow-up (17-24 weeks post-intervention) – ITT analysis (at-risk populations) EPDS ≥10 (mean 26 weeks)	*Study population*		RR 0.83 (0.65, 1.08)	540 (1 study)	●●○○ Low (b,c)
	333 per 1000	277 per 1000 (217, 360)			
	Moderate				
	333 per 1000	276 per 1000 (216, 360)			
	Study population		RR 0.64 (0.37, 1.1)	423 (1 study)	●●○○

[57] Enhanced treatment as usual involved non-mental health-focused education and support in the form of a booklet.

Outcomes (follow-up)	**Illustrative comparative risks***		**Relative effect (95% CI)**	**No. of participants (studies)**	**Certainty of the evidence (GRADE)**
	Assumed risk Control	**Corresponding risk Intervention**			
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk populations)	139 per 1000	89 per 1000 (51, 153)			Low (b,c)
	Moderate				
EPDS ≥10 (mean 26 weeks)	139 per 1000	89 per 1000 (51, 153)			
Evidence Statements: *A psychoeducational booklet has no effect on depression symptomatology at endpoint or first measurement (moderate certainty evidence) compared with treatment as usual or enhanced treatment as usual in pregnant or postpartum women who are considered to be 'at risk' of developing mental health problems in the perinatal period (psychosocial risk factors and/or a history of mental health problems).* *A psychoeducational booklet and telephone support appears to have no effect on depression symptomatology at short follow-up (9-16 weeks post-intervention) (low certainty evidence), or at intermediate follow-up (17-24 weeks post-intervention) (low certainty evidence), compared with enhanced treatment as usual (non-mental-health-focused education and support booklet) in postpartum women who are considered to be 'at risk' of developing mental health problems in the perinatal period (psychosocial risk factors).*					
Footnotes: * The 'assumed risk' for the study population is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The moderate risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. Risk of bias due to statistically significant group differences at baseline b. Total number of events is less than 300 (a threshold rule-of-thumb) c. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25)					

Source: NICE 2015, Table 41

Abbreviations: CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; ITT, intention-to-treat; RR, relative risk; SMD, standardised mean difference.

C4.1.3 Social/peer support

Of the two SRs of prevention using social/peer support identified in the literature search, NICE 2015 was chosen as the foundation review because the eligibility criteria were clearer. NICE 2015 included one RCT from the United Kingdom that compared social support with treatment as usual in women at risk of depression. The intervention involved antenatal and postnatal peer-mediated support, which included one-to-one befriending and psychoeducational group meetings.

The other SR included studies that were specifically excluded from NICE 2015 due to methodological reasons, or were classified as treatment rather than prevention studies, or are not generalisable to Australia. Of the three additional RCTs that were not included in NICE 2015, one focused on peer mentor support given to pregnant women living with human immunodeficiency virus (HIV) in South Africa, another assessed the effectiveness of a booklet and video provided by a nurse to pregnant adolescents in the United States, and the third assessed the impact of a supportive labour companion doula for low risk pregnant women from a low income, multi-cultural urban population in South Africa. As these three studies focus on very specific populations, the findings may not be sufficiently generalisable to the target population in Australia.

C4.1.3.1 Social support versus treatment as usual

Very low quality, single-study evidence revealed no preventive benefit of social support on depression diagnosis in women at risk of developing PND, when using an ITT approach (**Table C4-3**). Moreover, there are risk-of-bias concerns with this study due to non-blind outcome assessment. There was higher attrition in the intervention group relative to treatment as usual; however, this effect estimate was not statistically significant due to serious imprecision.

Table C4-3 Summary of findings (prevention) – social support versus treatment as usual

Outcomes (follow-up)	**Illustrative comparative risks***		**Relative effect (95% CI)**	**No. of participants (studies)**	**Certainty of the evidence (GRADE)**
	Assumed risk Control	**Corresponding risk Intervention**			
Depression diagnosis					
Post-treatment – ITT analysis (at-risk populations) Schedules for Clinical Assessment in Neuropsychiatry (SCAN) (mean 12 weeks)	*Study population*		RR 0.85 (0.65, 1.1)	117 (1 study)	●○○○ Very low (a,b,c)
	714 per 1000	607 per 1000 (464, 786)			
	Moderate				
	714 per 1000	607 per 1000 (464, 785)			
Post-treatment – available case analysis (at-risk populations) SCAN (mean 12 weeks)	*Study population*		**RR 0.37 (0.17, 0.8)**	65 (1 study)	●○○○ Very low (a,b)
	543 per 1000	201 per 1000 (92, 434)			
	Moderate				
	543 per 1000	201 per 1000 (92, 434)			
Evidence Statement: *Peer-mediated social support (one-to-one befriending and psychoeducational group meetings) appears to have no effect on depression diagnosis at endpoint or first measurement (very low certainty evidence) compared with treatment as usual in pregnant and postpartum women who are considered to be 'at risk' of developing mental health problems in the perinatal period.*					
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. Risk of bias due to non-blind outcome assessment b. Total number of events is less than 300 (a threshold rule-of-thumb) c. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25)					

Source: NICE 2015, Table 39

Abbreviations: CI, confidence interval; ITT, intention-to-treat; RR, relative risk; SCAN, Schedules for Clinical Assessment in Neuropsychiatry; SMD, standardised mean difference.

Note: Statistically significant differences are shown in bold.

C4.1.4 Home visits

Of the two SRs of prevention using home visits identified in the literature search, NICE 2015 was chosen as the foundation review due to its currency, high quality and comprehensiveness. NICE 2015 included five RCTs, one of which was from Australia. The five RCTs assessed home visits to provide emotional and practical support and information, predominantly for women with psychosocial risk factors, but also including adolescent mothers (three RCTs). The Australian RCT (N=120) assessed home visits provided by a physiotherapist and psychologist to women at risk of mental health problems due to preterm delivery. The Australian

intervention involved a psychological component directed at the mother and family, and a physiotherapy component directed at the infant. Of the remaining four RCTs, one involved antenatal home visits and three involved both antenatal and postnatal home visits.

C4.1.4.1 Home visits versus treatment as usual

NICE 2015 did not consider the timing of the intervention in their analyses.

Very low certainty evidence from up to two studies suggests that home visits are no more effective than treatment as usual at preventing depression symptomatology at endpoint or first measurement, or at very long follow-up (>104 weeks post-intervention), using an ITT approach (**Table C4-4**). However, confidence in this effect estimate is low due to risk-of-bias concerns and very serious imprecision. The two RCTs in the analysis were targeted at very different populations. One study involved home visitation, mentoring and case management delivered to adolescent mothers from an economically disadvantaged background in the United States, while the other (from Australia) involved an intervention with a physiotherapy and psychological component that was designed for families with a preterm infant (born at <30 weeks' gestational age).

Very low certainty evidence from one Australian study showed moderate-to-large effects of home visits on mean anxiety symptoms and on preventing anxiety symptomatology at endpoint and long-term follow-up in women who had a preterm delivery. However, confidence in these effect estimates is very low due to risk-of-bias concerns and imprecision.

A single study (very low certainty evidence) found a borderline statistically significant benefit of home visits relative to treatment as usual for preventing poor maternal sensitivity assessed using the CARE index (P=0.05) for women with psychosocial risk factors and (family) history of mental health problems; however, these estimates did not meet the criteria for clinically appreciable benefits. The intervention involved 18 months of weekly visits from a health visitor trained in understanding the processes of helping, skills of relating to parents effectively and methods of promoting parent–infant interaction using the Family Partnership Model.

Table C4-4 Summary of findings (prevention) – home visits versus treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression symptomatology					
Post-treatment– ITT analysis (at-risk populations) CES-D ≥21 or HADS– Depression >7 (52-117 weeks)	*Study population*		RR 0.94 (0.45, 1.96)	204 (2 studies)	⨁◯◯◯ Very low (a,b,c,d,e)
	434 per 1000	408 per 1000 (195, 851)			
	Moderate				
	429 per 1000	403 per 1000 (193, 841)			
Post-treatment– available case analysis (at-risk populations) CES-D ≥16/21 or HADS– Depression >7 (52-117 weeks)	*Study population*		RR 0.78 (0.44, 1.41)	684 (3 studies)	⨁◯◯◯ Very low (a,c,d,f)
	332 per 1000	259 per 1000 (146, 468)			
	Moderate				
	256 per 1000	200 per 1000 (113, 361)			
Very long Follow-up (>104 weeks post-intervention) – ITT analysis (at-risk populations) HADS– Depression ≥8 (mean 104 weeks)	*Study population*		RR 0.90 (0.59, 1.35)	120 (1 study)	⨁◯◯◯ Very low (a,c,d,e)
	458 per 1000	412 per 1000 (270, 618)			
	Moderate				
	158 per 1000	142 per 1000 (93, 213)			
Very long Follow-up (>104 weeks post-intervention) – available case analysis (at-risk populations) HADS– Depression ≥8 (mean 104 weeks)	*Study population*		RR 0.49 (0.13, 1.81)	77 (1 study)	⨁◯◯◯ Very low (a,c,d,e)
	158 per 1000	77 per 1000 (21, 286)			
	Moderate				
	158 per 1000	77 per 1000 (21, 286)			
Depression mean scores					
Post-treatment – available case analysis (at-risk populations) CES-D or HADS – Depression (mean 52 weeks)	SMD -0.38 (-0.75, -0.01)		-	621 (2 studies)	⨁◯◯◯ Very low (a,g)
Very long Follow-up (>104 weeks post-intervention) – available case analysis (at-risk populations) HADS – Depression (mean 104 weeks)	SMD -0.37 (-0.82, 0.08)		-	77 (1 study)	⨁◯◯◯ Very low (a,d,e,h)
Anxiety symptomatology					
Post-treatment – ITT analysis (at-risk populations) HADS – Anxiety >7 (mean 52 weeks)	*Study population*		**RR 0.63 (0.43, 0.91)**	120 (1 study)	⨁◯◯◯ Very low (a,c,e)
	627 per 1000	395 per 1000 (270, 571)			
	Moderate				
	627 per 1000	395 per 1000 (270, 571)			
Post-treatment – available case analysis (at-risk populations) HADS – Anxiety >7 (mean 52 weeks)	*Study population*		**RR 0.44 (0.23, 0.82)**	90 (1 study)	⨁◯◯◯ Very low (a,c,e)
	488 per 1000	215 per 1000 (112, 400)			
	Moderate				
	488 per 1000	215 per 1000 (112, 400)			
Long Follow-up (25-103 weeks post-intervention) – ITT analysis (at-risk populations) HADS– Anxiety ≥8 (mean 104 weeks)	*Study population*		**RR 0.74 (0.55, 0.98)**	120 (1 study)	⨁◯◯◯ Very low (a,c,e)
	712 per 1000	527 per 1000 (392, 698)			
	Moderate				
	712 per 1000	527 per 1000 (392, 698)			
Long Follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) HADS– Anxiety ≥8 (mean 104 weeks)	*Study population*		**RR 0.46 (0.25, 0.85)**	77 (1 study)	⨁◯◯◯ Very low (a,c,e)
	553 per 1000	254 per 1000 (138, 470)			
	Moderate				
	553 per 1000	254 per 1000 (138, 470)			

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Anxiety mean scores					
Post-treatment – available case analysis (at-risk populations) HADS – Anxiety (mean 52 weeks)	SMD -0.89 (-1.33, -0.46)		-	90 (1 study)	●○○○ Very low (a,e,h)
Long Follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) HADS – Anxiety (mean 104 weeks)	SMD -0.61 (-1.06, -0.15)		-	77 (1 study)	●○○○ Very low (a,e,h)
Maternal sensitivity mean scores					
Post-treatment – available case analysis (at-risk populations) CARE Index scale – Maternal sensitivity (mean 78 weeks)	SMD 0.36 (0, 0.72)		-	121 (1 study)	●○○○ Very low (d,e,h)

Evidence Statements:

Home visits appear to have no effect on depression symptomatology (very low certainty evidence) at endpoint or first measurement compared with treatment as usual in pregnant and postpartum women who are considered to be 'at risk' of developing mental health problems in the perinatal period (adolescence and psychosocial risk factors or preterm delivery)

Home visits appear to have no effect on depression symptomatology at very long follow-up (>104 weeks) (very low certainty evidence) compared with treatment as usual in postpartum women who are considered to be 'at risk' of developing mental health problems due to preterm delivery.

Home visits may improve depression mean scores at endpoint or first measurement (very low certainty evidence) compared with treatment as usual in pregnant and postpartum women who are considered to be 'at risk' of developing mental health problems in the perinatal period (adolescence and psychosocial risk factors or preterm delivery); however, the magnitude of the benefit may not be clinically significant.

Home visits may improve depression mean scores at very long follow-up (>104 weeks) (very low certainty evidence) compared with treatment as usual in postpartum women who are considered to be 'at risk' of developing mental health problems due to preterm delivery; however, the magnitude of the benefit may not be clinically significant.

Home visits may improve anxiety mean scores at endpoint or first measurement (very low certainty evidence), and at long follow-up (25-103 weeks post-intervention) (very low certainty evidence), compared with treatment as usual in postpartum women who are considered to be 'at risk' of developing mental health problems due to preterm delivery.

Home visits may improve anxiety symptomatology at endpoint or first measurement (very low certainty evidence), and at long follow-up (25-103 weeks post-intervention) (very low certainty evidence), compared with treatment as usual in postpartum women who are considered to be 'at risk' of developing mental health problems due to preterm delivery; however, the magnitude of the benefit may not be clinically significant.

Home visits may improve maternal sensitivity mean scores at endpoint or first measurement (very low certainty evidence) compared with treatment as usual in pregnant and postpartum women who are considered to be 'at risk' of developing mental health problems in the perinatal period (multiple psychosocial risk factors); however, the magnitude of the benefit may not be clinically significant.

Footnotes:

* The 'assumed risk' for the *Study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The moderate risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).

a. Risk of bias due to statistically significant group differences at baseline
b. There is evidence of considerable heterogeneity of study effect sizes
c. Total number of events is less than 300 (a threshold rule-of-thumb)
d. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25)
e. Paper omits data
f. There is evidence of moderate heterogeneity of study effect sizes
g. There is evidence of substantial heterogeneity of study effect sizes
h. Total population size is less than 400 (a threshold rule-of-thumb)

Source: NICE 2015, Table 43, Table 49, Table 52, Table 56

Abbreviations: CES-D, Center for Epidemiologic Studies Depression Scale; CI, confidence interval; HADS, Hospital Anxiety and Depression Scale; ITT, intention-to-treat; RR, relative risk; SD, standard deviation; SMD, standardised mean difference.

Note: Statistically significant differences are shown in bold.

C4.1.5 Non-mental-health-focused education/support

Of the two SRs of prevention using non-mental-health-focused education or support identified in the literature search, NICE 2015 was chosen as the foundation review due to its currency, high quality and comprehensiveness. NICE 2015 included four RCTs, one of which was from Australia. The included studies compared non-mental-health-focused education and support with treatment as usual or enhanced treatment as usual[58] for women with a range of risk factors including psychosocial risk factors, preterm delivery and low birthweight baby, and multiple (twin) pregnancy. In one RCT the intervention involved written and audiotaped education provided individually to postpartum women at hospital. In the other three RCTs the intervention involved face-to-face education and support provided antenatally and postnatally to women in a group format (one RCT) or in an individual and group format (two RCTs that also included home visits as part of the intervention). The setting varied within and across studies, and involved community, home, hospital and clinic.

NICE 2015 missed an RCT from the United States (N=167) that assessed a paraprofessional-delivered in-home intervention for young reservation-based American Indian mothers, and an Australian RCT (N=44) that assessed an educational intervention focusing on parenting and coping strategies delivered in an individual and group format to pregnant and postpartum women. The SR that included these additional studies found that education on preparing for parenting had a positive effect on depression compared with usual care; however, the effects were not statistically significant.

C4.1.5.1 Non-mental-health-focused education/support versus treatment as usual or enhanced treatment as usual

NICE 2015 did not consider the timing of the intervention, format or mode of delivery in their analyses.

Low certainty evidence from up to two studies suggests that non-mental-health-focused education and support may be more effective than treatment as usual or enhanced treatment as usual at preventing depression symptomatology for women with multiple births or at risk of developing postnatal depression (no further details reported). However, effects were not maintained at intermediate or long-term follow-ups, and there was no evidence for statistically or clinically significant preventive benefits for depression mean symptoms at any time point (**Table C4-5**).

There was single-study evidence for a moderate effect of non-mental-health-focused education and support for preventing anxiety symptomatology (at endpoint and short-term follow-up) in women with multiple births when an ITT analysis approach was used. However, confidence in these effect estimates was very low due to serious imprecision and selective reporting bias. In addition, there was no evidence for statistically or clinically significant effects on anxiety mean scores at endpoint, short-term or intermediate follow-up, or on anxiety symptomatology at intermediate follow-up.

A single study found no evidence at any time point for a clinically or statistically significant effect on mother-infant attachment problems for non-mental-health-focused education and support group and home visits relative to treatment as usual for women with an uncomplicated twin pregnancy, using an ITT approach. However, confidence in this effect estimate was very low due to very serious imprecision and risk of selective reporting bias.

[58] Enhanced treatment as usual involved non-mental health-focused education (one RCT) and non-mental health-focused education and support without the focus on healthy eating and exercise (one RCT).

Table C4-5 Summary of findings (prevention) – non-mental health-focused education and support versus treatment as usual or enhanced treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression symptomatology					
Post-treatment – ITT analysis (at-risk populations) EPDS >12 (6-13 weeks)	*Study population*		RR 0.7 (0.44, 1.14)	306 (2 studies)	●●○○ Low (a,b)
	320 per 1000	224 per 1000 (141, 365)			
	Moderate				
	316 per 1000	221 per 1000 (139, 360)			
Post-treatment – available case analysis (at-risk populations) EPDS >12 (6-13 weeks)	*Study population*		RR 0.57 (0.31, 1.05)	261 (2 studies)	●●○○ Low (a,b)
	188 per 1000	107 per 1000 (58, 197)			
	Moderate				
	188 per 1000	107 per 1000 (58, 197)			
Short Follow-up (9-16 weeks post-intervention) – ITT analysis (at-risk populations) EPDS >12 (mean 6 weeks)	*Study population*		RR 0.68 (0.44, 1.06)	162 (1 study)	●●○○ Low (a,b)
	402 per 1000	274 per 1000 (177, 427)			
	Moderate				
	402 per 1000	273 per 1000 (177, 426)			
Short Follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) – Non-mental-health-focused education and support EPDS >12 (mean 12 weeks)	*Study population*		RR 0.48 (0.21, 1.12)	128 (1 study)	●●○○ Low (a,b)
	222 per 1000	107 per 1000 (47, 249)			
	Moderate				
	222 per 1000	107 per 1000 (47, 249)			
Intermediate follow-up (17-24 weeks post-intervention) – ITT analysis (at-risk populations) EPDS >12 (20-24 weeks)	*Study population*		RR 0.91 (0.44, 1.89)	306 (2 studies)	●○○○ Very low (a,b,e)
	294 per 1000	268 per 1000 (129, 556)			
	Moderate				
	290 per 1000	264 per 1000 (128, 548)			
Intermediate follow-up (17-24 weeks post-intervention) - available case analysis (at-risk populations) EPDS >12 (20-24 weeks)	*Study population*		RR 0.84 (0.27, 2.63)	254 (2 studies)	●○○○ Very low (a,b,e)
	143 per 1000	120 per 1000 (39, 376)			
	Moderate				
	142 per 1000	119 per 1000 (38, 373)			
Long Follow-up (25-103 weeks post-intervention) – ITT analysis (at-risk populations) EPDS >12 (mean 52 weeks)	*Study population*		RR 0.84 (0.57, 1.25)	162 (1 study)	●●○○ Low (a,b)
	415 per 1000	348 per 1000 (236, 518)			
	Moderate				
	415 per 1000	349 per 1000 (237, 519)			
Long Follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) EPDS >12 (mean 52 weeks)	*Study population*		RR 0.87 (0.42, 1.83)	123 (1 study)	●●○○ Low (a,b)
	200 per 1000	174 per 1000 (84, 366)			
	Moderate				
	200 per 1000	174 per 1000 (84, 366)			
Depression mean scores					
Post-treatment – ITT analysis (at-risk populations) CES-D (mean 28 weeks)	SMD -0.13 (-0.37, 0.1)		-	275 (1 study)	●●○○ Low (c,d)
Post-treatment – available case analysis (at-risk populations) BDI or EPDS	SMD -0.14 (-0.34, 0.07)		-	370 (2 studies)	●●●○ Moderate (c)
Short Follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) EPDS (mean 12 weeks)	SMD -0.21 (-0.56, 0.13)		-	128 (1 study)	●●○○ Low (b,c)

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk populations) EPDS (mean 24 weeks)	SMD -0.3 (-0.64, 0.04)		-	133 (1 study)	●●○○ Low (b,c)
Long Follow-up (25-103 weeks post-intervention) -Available case analysis (at-risk populations) EPDS (mean 52 weeks)	SMD -0.08 (-0.44, 0.27)		-	123 (1 study)	●●○○ Low (c)
Anxiety symptomatology					
Post-treatment – ITT analysis (at-risk populations) HADS – Anxiety (above unspecified threshold) (mean 6 weeks)	*Study population*		RR 0.74 (0.44, 1.24)	162 (1 study)	●○○○ Very low (a,b,d)
	305 per 1000	226 per 1000 (134, 378)			
	Moderate				
	305 per 1000	226 per 1000 (134, 378)			
Post-treatment – available case analysis (at-risk populations) HADS – Anxiety (above unspecified threshold) (mean 6 weeks)	*Study population*		RR 0.93 (0.32, 2.72)	131 (1 study)	●○○○ Very low (a,b,d)
	95 per 1000	89 per 1000 (30, 259)			
	Moderate				
	95 per 1000	88 per 1000 (30, 258)			
Short Follow-up (9-16 weeks post-intervention) – ITT analysis (at-risk populations) HADS – Anxiety (above unspecified threshold) (mean 12 weeks)	*Study population*		RR 0.67 (0.38, 1.19)	162 (1 study)	●○○○ Very low (a,b,d)
	280 per 1000	188 per 1000 (107, 334)			
	Moderate				
	281 per 1000	188 per 1000 (107, 334)			
Short Follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) HADS – Anxiety (above unspecified threshold) (mean 12 weeks)	*Study population*		RR 0.11 (0.01, 1.96)	128 (1 study)	●○○○ Very low (a,b,d)
	63 per 1000	7 per 1000 (1, 124)			
	Moderate				
	64 per 1000	7 per 1000 (1, 125)			
Intermediate follow-up (17-24 weeks post-intervention) – ITT analysis (at-risk populations) HADS – Anxiety (above unspecified threshold) (mean 24 weeks)	*Study population*		RR 0.76 (0.44, 1.31)	162 (1 study)	●○○○ Very low (a,b,d)
	280 per 1000	213 per 1000 (123, 367)			
	Moderate				
	281 per 1000	214 per 1000 (124, 368)			
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk populations) HADS – Anxiety (above unspecified threshold) (mean 24 weeks)	*Study population*		RR 0.94 (0.25, 3.6)	130 (1 study)	●○○○ Very low (a,b,d)
	63 per 1000	60 per 1000 (16, 229)			
	Moderate				
	64 per 1000	60 per 1000 (16, 230)			
Anxiety mean scores					
Post-treatment – available case analysis (at-risk populations) STAI-S or HADS –Anxiety (mean 6 weeks)	SMD -0.1 (-0.3, 0.11)		-	370 (2 studies)	●●●○ Moderate (c)
Short Follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) HADS – Anxiety (mean 12 weeks)	SMD -0.2 (-0.54, 0.15)		-	128 (1 study)	●○○○ Very low (b,c,d)
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk populations) HADS – Anxiety (mean 24 weeks)	SMD -0.26 (-0.6, 0.09)		-	130 (1 study)	●○○○ Very low (b,c,d)

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Mother-infant attachment problems					
Post-treatment – ITT analysis (at-risk populations) Green scale: Mother-infant attachment problems (above unspecified threshold) (mean 6 weeks)	*Study population*		RR 0.9 (0.65, 1.25)	162 (1 study)	⊕◯◯◯ Very low (a,b,d)
	500 per 1000	450 per 1000 (325, 625)			
	Moderate				
	500 per 1000	450 per 1000 (325, 625)			
Post-treatment – available case analysis (at-risk populations) Green scale: Mother-infant attachment problems (above unspecified threshold) (mean 6 weeks)	*Study population*		RR 1.01 (0.64, 1.59)	133 (1 study)	⊕◯◯◯ Very low (a,b,d)
	359 per 1000	363 per 1000 (230, 571)			
	Moderate				
	359 per 1000	363 per 1000 (230, 571)			
Short Follow-up (9-16 weeks post-intervention) – ITT analysis (at-risk populations) Green scale: Mother-infant attachment problems (above unspecified threshold) (mean 12 weeks)	*Study population*		RR 1.08 (0.78, 1.49)	162 (1 study)	⊕◯◯◯ Very low (a,b,d)
	463 per 1000	500 per 1000 (361, 690)			
	Moderate				
	463 per 1000	500 per 1000 (361, 690)			
Short Follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) Green scale: Mother-infant attachment problems (above unspecified threshold) (mean 12 weeks)	Study population		RR 1.29 (0.78, 2.13)	126 (1 study)	⊕◯◯◯ Very low (a,b,d)
	290 per 1000	375 per 1000 (226, 618)			
	Moderate				
	290 per 1000	374 per 1000 (226, 618)			
Intermediate follow-up (17-24 weeks post-intervention) – ITT analysis (at-risk populations) Green scale: Mother-infant attachment problems (above unspecified threshold) (mean 24 weeks)	*Study population*		RR 0.85 (0.64, 1.14)	162 (1 study)	⊕◯◯◯ Very low (a,b,d)
	585 per 1000	498 per 1000 (375, 667)			
	Moderate				
	585 per 1000	497 per 1000 (374, 667)			
Intermediate follow-up (17-24 weeks post-intervention) – available case analysis (at-risk populations) Green scale: Mother-infant attachment problems (above unspecified threshold) (mean 24 weeks)	*Study population*		RR 0.89 (0.59, 1.34)	127 (1 study)	⊕◯◯◯ Very low (a,b,d)
	443 per 1000	394 per 1000 (261, 593)			
	Moderate				
	443 per 1000	394 per 1000 (261, 594)			
Positive mother-infant interaction mean scores					
Post-treatment – available case analysis (at-risk populations) Index of Parental Behavior in the NICU: Positive interaction with quiet alert infant	**SMD 0.57 (0.29, 0.85)**		-	211 (1 study)	⊕⊕◯◯ Low (c,d)

Evidence Statements:

Non-mental-health-focused education and support (individual and group, face-to-face, with or without home visits) appears to have no effect on depression symptomatology at endpoint or first measurement (low certainty evidence), or at short follow-up (9-16 weeks post-intervention) (low certainty evidence), or at intermediate follow-up (17-24 weeks post-intervention) (very low certainty evidence), compared with treatment as usual in pregnant and postpartum women who are considered to be 'at risk' of developing mental health problems in the perinatal period.

Non-mental-health-focused education and support (individual and group, face-to-face, with home visits) appears to have no effect on depression symptomatology (low certainty evidence) at long follow-up (25-103 weeks post-intervention) compared with treatment as usual in pregnant and postpartum women who are considered to be 'at risk' of developing mental health problems due to multiple (twin) pregnancy

Non-mental-health-focused education and support (individual and group, face-to-face, with home visits) appears to have no effect on depression mean scores (low certainty evidence) at endpoint or first measurement compared with enhanced treatment as usual (non-mental-health-focused education and support without the focus on healthy eating and exercise) in pregnant and postpartum women who are considered to be 'at risk' of developing mental health problems in the perinatal period (psychosocial risk factors).

Non-mental-health-focused education and support (individual and group, face-to-face, with home visits) appears to have no effect on depression mean scores (low certainty evidence) at short (9-16 weeks post-intervention), intermediate (17-24 weeks post-intervention), or long (25-103 weeks post-intervention) follow-up compared with treatment as usual in pregnant and postpartum women who are considered to be 'at risk' of developing mental health problems due to multiple (twin) pregnancy

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Non-mental-health-focused education and support (individual and group, face-to-face, with home visits) appears to have no effect on anxiety symptomatology (very low certainty evidence) at endpoint or first measurement, or at short follow-up (9-16 weeks post-intervention), or at intermediate follow-up (17-24 weeks post-intervention) compared with treatment as usual in pregnant and postpartum women who are considered to be 'at risk' of developing mental health problems due to multiple (twin) pregnancy. *Non-mental-health-focused education and support (individual and group, with or without home visits) has no effect on anxiety mean scores at endpoint or first measurement (moderate certainty evidence), and appears to have no effect on anxiety mean scores at short follow-up (9-16 weeks post-intervention) (very low certainty evidence), or at intermediate follow-up (17-24 weeks post-intervention) (very low certainty evidence) compared with treatment as usual in pregnant and postpartum women who are considered to be 'at risk' of developing mental health problems due to preterm delivery and low birthweight or multiple (twin) pregnancy.* *Non-mental-health-focused education and support (individual and group, with or without home visits) appears to have no effect on mother-infant attachment problems (very low certainty evidence) at endpoint or first measurement, at short follow-up (9-16 weeks post-intervention), or at intermediate follow-up (17-24 weeks post-intervention), compared with treatment as usual in pregnant and postpartum women who are considered to be 'at risk' of developing mental health problems due to multiple (twin) pregnancy* *Non-mental-health-focused education and support (individual, written and audiotaped) appears to have no effect on positive mother-infant interaction mean scores (low certainty evidence) at endpoint or first measurement compared with enhanced treatment as usual (non-mental-health-focused information) in postpartum women who are considered to be 'at risk' of developing mental health problems due to preterm delivery and low birthweight.*					

Footnotes:

* The 'assumed risk' for the study population is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The moderate risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).

a. Total number of events is less than 300 (a threshold rule-of-thumb)

b. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25)

c. Total population size is less than 400 (a threshold rule-of-thumb)

d. Paper omits data

e. There is evidence of substantial heterogeneity of study effect sizes

Source: NICE 2015, Table 42, Table 48, Table 55

Abbreviations: BDI, Beck Depression Inventory; CES-D, Center for Epidemiologic Studies Depression Scale; CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; HADS, Hospital Anxiety and Depression Scale; ITT, intention-to-treat; NICU, neonatal intensive care unit; RR, relative risk; SMD, standardised mean difference; STAI-S, State-Trait Anxiety Inventory-State.

C4.1.6 Pre-delivery discussion

The literature search identified no SRs that relate to this intervention.

Table C4-6 Summary of findings (prevention) – pre-delivery discussion

Evidence Statement: *There is no RCT evidence for pre-delivery discussion in pregnant women who are considered to be 'at risk' of developing mental health problems in the perinatal period.*

C4.1.7 Post-delivery discussion

Of the two SRs of post-delivery discussion identified in the literature search, NICE 2015 was chosen as the foundation review. NICE 2015 included one RCT from Australia (N=1,041) that assessed individual, face-to-face, midwife-led post-delivery discussion (single session) for women who had an operative delivery.

C4.1.7.1 Post-delivery discussion versus enhanced treatment as usual

A single study found no clinically or statistically significant benefits of a midwife-led post-delivery discussion relative to enhanced treatment as usual (a non-mental health-focused information booklet) on preventing depression or poor general mental health outcomes in women who had had an operative delivery (**Table C4-7**).

Table C4-7 Summary of findings (prevention) – post-delivery discussion versus enhanced treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression symptomatology					
Post-treatment – ITT analysis (at-risk populations) EPDS ≥13 (mean 26 weeks)	*Study population*		RR 0.98 (0.8, 1.2)	1,041 (1 study)	●●●○ Moderate (a)
	263 per 1000	258 per 1000 (210, 316)			
	Moderate				
	263 per 1000	258 per 1000 (210, 316)			
Post-treatment – available case analysis (at-risk populations) EPDS ≥13 (mean 26 weeks)	*Study population*		RR 1.2 (0.89, 1.62)	916 (1 study)	●●○○ Low (a,b)
	145 per 1000	174 per 1000 (129, 235)			
	Moderate				
	145 per 1000	174 per 1000 (129, 235)			
Very long Follow-up (>104 weeks post-intervention) – ITT analysis (at-risk populations) EPDS ≥13 (208-312 weeks)	*Study population*		RR 1.01 (0.91, 1.12)	1,041 (1 study)	●●●● High
	568 per 1000	574 per 1000 (517, 636)			
	Moderate				
	568 per 1000	574 per 1000 (517, 636)			
Very long Follow-up (>104 weeks post-intervention) – available case analysis (at-risk populations) EPDS ≥13 (208-312 weeks)	*Study population*		RR 0.95 (0.65, 1.4)	534 (1 study)	●●○○ Low (a,b)
	167 per 1000	158 per 1000 (108, 233)			
	Moderate				
	167 per 1000	159 per 1000 (109, 234)			
Depression mean scores					
Post-treatment – available case analysis (at-risk populations) EPDS (mean 26 weeks)	SMD 0.08 (-0.05, 0.21)			916 (1 study)	●●●● High
Very long Follow-up (>104 weeks post-intervention) – available case analysis (at-risk populations) EPDS (208-312 weeks)	SMD -0.08 (-0.25, 0.09)			534 (1 study)	●●●● High

Evidence Statement:

Individual, midwife-led post-delivery discussion has no effect on depression symptomatology at endpoint or first measurement (moderate certainty evidence) or at very long follow-up (>104 weeks post-intervention) (high certainty evidence) compared with a non-mental health-focused information booklet in women who are considered to be 'at risk' of developing mental health problems in the postnatal period due to an operative delivery.

Individual, midwife-led post-delivery discussion has no effect on depression mean scores at endpoint or first measurement (high certainty evidence), or at very long follow-up (>104 weeks post-intervention) (high certainty evidence), compared with a non-mental health-focused information booklet in women who are considered to be 'at risk' of developing mental health problems in the postnatal period due to an operative delivery

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. Total number of events is less than 300 (a threshold rule-of-thumb) b. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25)					

Source: NICE 2015, Table 44, Table 53

Abbreviations: CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; ITT, intention-to-treat; RR, relative risk; SMD, standardised mean difference.

C4.1.8 Post-miscarriage self-help

Only one SR of prevention using post-miscarriage self-help (NICE 2015) was identified in the literature search. NICE 2015 included one RCT conducted in European German-speaking countries that compared post-miscarriage self-help with waitlist control. The intervention involved a five-week cognitive behaviour internet-based self-help therapy for parents after the loss of a child during pregnancy (due to miscarriage, termination due to fetal abnormality, or stillbirth). The self-help intervention was based on CBT principles and participants were assigned written tasks (10 x 45-minute assignments) which were personalised by the therapist for each participant. At baseline, 37% study participants had an IES score >35 (baseline IES-R mean score 31.1 [SD 8.6]).

C4.1.8.1 Post-miscarriage self-help versus treatment as usual

A single study showed large effects of post-miscarriage self-help on preventing PTSD symptomatology and reducing mean PTSD for women who had lost a child during pregnancy (**Table C4-8**). However, confidence in these effect estimates was very low due to risk-of-bias concerns and imprecision.

Table C4-8 Summary of findings (prevention) – post-miscarriage self-help versus treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
PTSD symptomatology					
Post-treatment – ITT IES-R ≥35 (mean 5 weeks)	*Study population*		**RR 0.34 (0.18, 0.62)**	228 (1 study)	●○○○ Very low (a,c)
	310 per 1000	105 per 1000 (56, 192)			
	Moderate				
	310 per 1000	105 per 1000 (56, 192)			
PTSD mean scores					
Post-treatment – ITT IES-R (mean 5 weeks)	**SMD -0.88 (-1.15, -0.61)**		-	228 (1 study)	●○○○ Very low (a,b)

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression mean symptoms					
Post-treatment – ITT BSI: Depression[59] (mean 5 weeks)	**SMD -0.64 (-0.91, -0.37)**		-	228 (1 study)	●●○○ Low (a,b)
Anxiety mean scores					
Post-treatment – ITT BSI: Anxiety[60] (mean 5 weeks)	**SMD -0.47 (-0.73, -0.2)**		-	228 (1 study)	●●○○ Low (a,b)
Evidence Statement: *Internet-based cognitive behaviour self-help therapy appears to improve PTSD symptomatology (very low certainty evidence), PTSD mean scores (very low certainty evidence), depression mean symptoms (low certainty evidence), and anxiety mean scores (low certainty evidence), at endpoint or first measurement compared with waitlist control in women who are considered to be 'at risk' of developing mental health problems due to the loss of a child during pregnancy.*					
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. Risk of bias due to statistically significant group differences at baseline b. Imprecision - Total population size is less than 400 (a threshold rule-of-thumb) c. Imprecision - Total number of events is less than 300 (a threshold rule-of-thumb)					

Source: NICE 2015, Table 38, Table 47, Table 50, Table 51

Abbreviations: BSI, Brief Symptom Inventory; CI, confidence interval; IES-R, Impact of Events Scale – Revised; ITT, intention-to-treat; PTSD, post-traumatic stress disorder; RR, relative risk; SMD, standardised mean difference.

Note: Statistically significant differences are shown in bold.

C4.1.9 Seeing and/or holding stillborn infant

Only one SR of prevention associated with seeing and/or holding stillborn infant was identified in the literature search (NICE 2015) but no RCTs were included. NICE 2015 included three retrospective cohort studies and one nested cohort within a case-control study. All included studies compared mental health outcomes in women who saw and/or held their stillborn infants compared with those who did not. The length of time since the stillbirth varied considerably within and between studies, ranging from less than one year to 18 years. One study only recruited women who had previously experienced a stillbirth and were pregnant with another child.

C4.1.9.1 Seeing and/or holding stillborn infant versus not seeing and/or holding the stillborn infant

The evidence for benefits or harms associated with seeing and/or holding the stillborn infant was contradictory with evidence from a single cohort study from the United Kingdom suggestive of harms associated with these protocols following stillbirth and evidence from a Swedish study and Norwegian study suggestive of benefits associated with spending as much time with the stillborn infant as women wished or holding the stillborn infant. Potential reasons for these differences

[59] Outcome is a subscale of a global severity measure – not depression-specific

[60] Outcome is a subscale of a global severity measure – not anxiety specific

could be differences in gestational age at the time of stillbirth (none of the papers report the mean gestational age at stillbirth) and pregnancy status at the time of participation in the studies.

Table C4-9 Summary of findings (prevention) – seeing and/or holding stillborn infant versus not seeing and/or holding stillborn infant

Evidence Statement: *There is no RCT evidence for seeing and/or holding the stillborn infant versus not seeing and/or holding the stillborn infant in women who are considered to be 'at risk' of developing mental health problems in the perinatal period.*

C4.1.10 Mother-infant relationship interventions

Of the two SRs of prevention associated with mother-infant relationship interventions identified in the literature search, NICE 2015 was chosen as the foundation review due to its currency, high quality and comprehensiveness. The four RCTs included in NICE 2015 compared face-to-face mother-infant relationship interventions with treatment as usual in women with psychosocial risk factors (one RCT) or with premature or low birthweight babies (three RCTs). One RCT was from Australia (assessing an intervention largely based on the Mother-Infant Transaction Program) while the others were from Norway (Mother-Infant Transaction Program), the Netherlands (Infant Behavioural Assessment and Intervention Program), and South Africa (a relationship/attachment based intervention that closely follows the principles of The Social Baby). The intervention was delivered antenatally and postnatally at home in one RCT, and postnatally (at home, or both at hospital and home) in the remaining three RCTs.

One additional RCT from the United Kingdom was not included in NICE 2015 but the study was relatively small in terms of sample size (N=35). The antenatal parenting support intervention, which was underpinned by attachment theory and aimed at women with additional health and social care needs, appeared to have positive effects on the women's mental health and well-being overall at 8-12 weeks postnatally.

C4.1.10.1 Mother-infant relationship interventions versus treatment as usual

NICE 2015 did not consider the timing or setting of the intervention in their analyses.

A single study in women living in socioeconomically deprived community in South Africa provided low certainty evidence that a mother-infant relationship intervention aimed at improving the quality of the mother-infant relationship and infant attachment had no clinically or statistically significant effect on preventing mother-infant attachment problems (assessed using the Ainsworth Strange Situation Procedure); however, a moderate benefit was seen when an available case analysis approach was used. There was also evidence from two studies (one from Australia and one from the Netherlands) for a small benefit of a mother-infant relationship intervention on preventing poor mother-infant interaction mean scores for women who had had a preterm delivery and/or a low birthweight baby; however, this effect estimate did not reach criteria for a clinically meaningful benefit (SMD<0.5). There was no statistically significant difference between study groups for any other observed outcome measures.

The evidence for mother-infant relationship interventions preventing depression in women with psychosocial risk factors or who had a preterm delivery and/or low birthweight baby was inconsistent (**Table C4-10**); however, the interventions were not specifically intended for this purpose. A single study showed no

statistically or clinically significant effects of a mother-infant relationship intervention on depression diagnosis in women living in a socioeconomically deprived community in South Africa, using an ITT approach. However, a single study in Norwegian women who had a preterm delivery showed large harms on depression symptomatology associated with mother-infant relationship interventions (one and a half times more likely to score CES-D ≥16); however, the confidence in this effect estimate is very low due to risk-of-bias concerns and very serious imprecision, and the effect was not maintained at long follow-up (25-103 weeks post-intervention). There was no evidence for statistically or clinically significant effects of mother-infant relationship interventions on mean depression symptoms at short-term or long-term follow-up, and no evidence for clinically significant effects on depression mean symptoms at endpoint.

Table C4-10 Summary of findings (prevention) – mother-infant relationship interventions versus treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression diagnosis					
Post-treatment – ITT analysis (at-risk populations) SCID (mean 26 weeks)	*Study population*		RR 1 (0.76, 1.31)	449 (1 study)	●●○○ Low (a,b)
	323 per 1000	323 per 1000 (246, 423)			
	Moderate				
	323 per 1000	323 per 1000 (245, 423)			
Post-treatment – available case analysis (at-risk populations) SCID (mean 26 weeks)	*Study population*		RR 0.78 (0.47, 1.32)	354 (1 study)	●●○○ Low (a,b)
	158 per 1000	123 per 1000 (74, 208)			
	Moderate				
	158 per 1000	123 per 1000 (74, 209)			
Long Follow-up (25-103 weeks post-intervention) – ITT analysis (at-risk populations) SCID (mean 52 weeks)	*Study population*		RR 1 (0.77, 1.3)	449 (1 study)	●●○○ Low (a,b)
	332 per 1000	332 per 1000 (256, 431)			
	Moderate				
	332 per 1000	332 per 1000 (256, 432)			
Long Follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) SCID (mean 52 weeks)	*Study population*		RR 0.71 (0.41, 1.23)	346 (1 study)	●●○○ Low (a,b)
	155 per 1000	110 per 1000 (63, 190)			
	Moderate				
	155 per 1000	110 per 1000 (64, 191)			
Depression symptomatology					
Post-treatment – ITT analysis (at-risk populations) CES-D ≥16 (mean 27 weeks)	*Study population*		RR 1.52 (0.77, 3)	106 (1 study)	●○○○ Very low (a,d,g)
	200 per 1000	304 per 1000 (154, 600)			
	Moderate				
	200 per 1000	304 per 1000 (154, 600)			
Post-treatment – available case analysis (at-risk populations) CES-D ≥16 (mean 27 weeks)	*Study population*		RR 2.8 (0.6, 13.11)	87 (1 study)	●○○○ Very low (a,b,g)
	48 per 1000	133 per 1000 (29, 624)			
	Moderate				
	48 per 1000	134 per 1000 (29, 629			
Long Follow-up (25-103 weeks post-intervention) – ITT analysis (at-risk populations) CES-D ≥16 (mean 53 weeks)	*Study population*		RR 0.94 (0.56, 1.58)	106 (1 study)	●○○○ Very low (a,b,g)
	360 per 1000	338 per 1000 (202, 569)			
	Moderate				
	360 per 1000	338 per 1000 (202, 569)			

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Long Follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) CES-D ≥16 (mean 53 weeks)	*Study population*		RR 0.75 (0.25, 2.27)	80 (1 study)	●○○○ Very low (a,b,g)
	158 per 1000	118 per 1000 (39, 358)			
	Moderate				
	158 per 1000	119 per 1000 (40, 359)			
Depression mean scores					
Post-treatment – available case analysis (at-risk populations) EPDS (15-26 weeks)	**SMD -0.22 (-0.41, -0.02)**		-	417 (2 studies)	●●●● High
Short Follow-up (9-16 weeks post-intervention) – available case analysis (at-risk populations) EPDS (mean 28 weeks)	SMD -0.3 (-0.8, 0.19)		-	63 (1 study)	●●○○ Low (b,c)
Long Follow-up (25-103 weeks post-intervention) – available case analysis (at-risk populations) EPDS (mean 52 weeks)	SMD -0.14 (-0.35, 0.06)		-	354 (1 study)	●●●○ Moderate (c)
Mother-infant attachment problems					
Post-treatment – ITT analysis (at-risk populations) Ainsworth Strange Situation: Insecure (mean 78 weeks)	*Study population*		RR 0.85 (0.71, 1.02)	449 (1 study)	●●○○ Low (a,b)
	555 per 1000	471 per 1000 (394, 566)			
	Moderate				
	555 per 1000	472 per 1000 (394, 566)			
Post-treatment – available case analysis (at-risk populations) Ainsworth Strange Situation: Insecure (mean 78 weeks)	*Study population*		**RR 0.69 (0.5, 0.97)**	318 (1 study)	●●○○ Low (a)
	370 per 1000	256 per 1000 (185, 359)			
	Moderate				
	370 per 1000	255 per 1000 (185, 359)			
Positive mother-infant interaction mean scores					
Post-treatment – available case analysis (at-risk populations) Infant and Caregiver Engagement Phases (ICEP): Maternal positive engagement (% of time during behavioural observation) or Synchrony Scale (Milgrom & Meitz, 1988): Reciprocity/Synchrony (15-26 weeks)	**SMD 0.46 (0.16, 0.76)**		-	175 (2 studies)	●●○○ Low (c)
Maternal sensitivity mean scores					
Post-treatment – available case analysis (at-risk populations) Maternal Sensitivity and Responsivity Scales: Maternal sensitivity or synchrony (Milgrom & Meitz, 1988): Maternal Respond (15-26 weeks)	SMD 0.62 (-0.11, 1.35)		-	172 (2 studies)	●○○○ Very low (b,c,d)

Evidence Statements:

An individual, face-to-face mother-infant relationship intervention appears to have no effect on mother-infant attachment problems at endpoint or first measurement (low certainty evidence) compared with treatment as usual in pregnant and postpartum women who are considered to be 'at risk' of developing mental health problems in the perinatal period due to psychosocial risk factors.

Individual, face-to-face mother-infant relationship interventions may improve positive mother-infant interaction mean scores at endpoint or first measurement (low certainty evidence) compared with treatment as usual in postpartum women who are considered to be 'at risk' of developing mental health problems due to preterm delivery and/or low birthweight; however, the magnitude of the benefit is not clinically significant.

Individual, face-to-face mother-infant relationship interventions appear to have no effect on maternal sensitivity mean scores at endpoint or first measurement (very low certainty evidence) compared with treatment as usual in postpartum women who are considered to be 'at risk' of developing mental health problems due to preterm delivery and/or low birthweight.

An individual, face-to-face mother-infant relationship intervention improves depression mean scores at endpoint or first measurement (high certainty evidence) compared with treatment as usual in pregnant and postpartum women who are considered to be 'at risk' of developing mental health problems in the perinatal period due to psychosocial risk factors or preterm delivery and/or low birthweight; however, the magnitude on the benefit is not clinically significant.

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
An individual, face-to-face mother-infant relationship intervention has no effect on depression mean scores at long follow-up (25-103 weeks post-intervention) (moderate certainty evidence), and appears to have no effect on depression diagnosis at endpoint or first measurement (low certainty evidence), or at long follow-up (25-103 weeks post-intervention) (low certainty evidence), compared with treatment as usual in pregnant and postpartum women who are considered to be 'at risk' of developing mental health problems in the perinatal period due to psychosocial risk factors. *An individual, face-to-face mother-infant relationship intervention appears to have no effect on depression mean scores at short follow-up (9-16 weeks post-intervention) (low certainty evidence), and appears to have no effect (and may be harmful) on depression symptomatology at endpoint or first measurement (very low certainty evidence), or at long follow-up (25-103 weeks post-intervention) (very low certainty evidence), compared with treatment as usual in postpartum women who are considered to be 'at risk' of developing mental health problems due to preterm delivery and/or low birthweight.*					
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. Total number of events is less than 300 (a threshold rule-of-thumb) b. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25) c. Total population size is less than 400 (a threshold rule-of-thumb) d. There is evidence of considerable heterogeneity of study effect sizes e. High risk of selection bias due to statistically significant baseline difference with the intervention group having more mothers with earlier preterm birth and non-Norwegian origin f. Paper omits data g. Risk of bias due to statistically significant group differences at baseline					

Source: NICE 2015, Table 45, Table 54, Table 57

Abbreviations: CES-D, Center for Epidemiologic Studies Depression Scale; CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; HADS, Hospital Anxiety and Depression Scale; ICEP, Infant and Caregiver Engagement Phases; ITT, intention-to-treat; PICS, Pictoral Infant Communication Scales; RR, relative risk; SCID, Structured Clinical Interview for DSM Disorders; SMD, standardised mean difference; STSI, Short Temperament Scale for Infants.

Note: Statistically significant differences are shown in bold.

C4.1.11 Co-parenting interventions

The literature search identified no SRs that relate to this intervention.

Table C4-11 Summary of findings (prevention) – co-parenting interventions

Evidence Statement: *There is no RCT evidence for co-parenting interventions in women who are considered to be 'at risk' of developing mental health problems in the perinatal period.*

C4.1.12 Mindfulness

Three SRs of prevention using mindfulness interventions were identified in the literature search but only one RCT (identified in the Taylor 2016 SR) is considered relevant to the prevention of mental health problems in the perinatal period. The pilot RCT from the United States (N=55) assessed an adapted Mindfulness-Based Cognitive Therapy (MBCT) aimed at the prevention of depressive relapse/recurrence in pregnant women with a history of major depressive disorder but who were not currently clinically depressed. Although there was no significant difference in depressive symptoms between intervention and control participants at post-

intervention, the mindfulness intervention appeared to be protective against depressive relapse up to six months postpartum (estimated 18.4% for intervention participants compared with 50.2% for control participants; HR 3.87 [95% CI 1.39, 10.76]).

A SoF table has not been constructed for this outcome as it required data extraction, risk-of-bias assessment and formulation of comparative risks – all of which were out of scope for this PICO.

Table C4-12 Summary of findings (prevention) – mindfulness

Evidence Statement: *There is limited RCT evidence for mindfulness interventions in women who are considered to be 'at risk' of developing mental health problems in the perinatal period.*

C4.2 PREVENTION WITH PSYCHOLOGICAL INTERVENTIONS

C4.2.1 Structured psychological interventions (CBT or IPT)

Of the three SRs of prevention using structured psychological interventions identified in the literature search, Morrell 2016 was chosen as the foundation review. Of note, NICE 2015 did not include structured psychological interventions (CBT and IPT) as an explicit preventive intervention type for women considered to be 'at risk' of mental health problems in the perinatal period. However, all except one of the RCTs included in the three identified SRs were either excluded from NICE 2015 due to methodological reasons or were classified under alternative intervention types and are therefore accounted for in other sections of the current report (most often 'psychologically informed psychoeducation' or 'structured psychological interventions' for treatment rather than prevention). The only RCT that is not mentioned in the NICE 2015 SR was judged by Morrell 2016 to be at high risk of bias. Chabrol 2002 (N=258) assessed an individual CBT-based intervention that involved one cognitive behaviour prevention session during hospitalisation to pregnant women at risk of depression.

Morrell 2016 included five selective preventive intervention RCTs for structured psychological interventions, three of which were CBT-based and two were IPT-based. Comparisons were made with educational information in one RCT and usual care in the other four RCTs. One Australian RCT provided the intervention in a group format, while the other four RCTs incorporated individual sessions. None of the interventions were provided in the home setting. One RCT provided the intervention in the antenatal period only whereas two RCTs initiated the intervention postnatally and two RCTs provided the intervention across the perinatal period from pregnancy to after childbirth. Interventions were provided by a variety of service providers.

Morrell 2016 also included 13 indicated preventive intervention RCTs for structured psychological interventions, eight of which were CBT-based and five were IPT-based. Comparisons were made with educational information in two RCTs and with usual care in the other 11 RCTs. Five RCTs evaluated group sessions, seven evaluated individual sessions, and one evaluated both group and individual sessions. Three RCTs took place in the home setting. Two RCTs were undertaken in the antenatal period only, one in the postnatal period only, and the remainder were undertaken in both antenatal and postnatal periods. The interventions were provided by different health-care providers in all except one RCT where the intervention was provided by a group facilitator.

It should be noted that Morrell 2016 has several errors in referencing, but that best efforts have been used in the current Review to validate and cross-reference their reported data.

C4.2.1.1 Structured psychological interventions (CBT and IPT) versus usual care

Of the five preventive intervention RCTs, two were included in the network meta-analysis (NMA).[61] One RCT (Chabrol 2002) was included in an NMA that was used to compare the effects of a CBT-based intervention with thyroxine and midwife-led debriefing following childbirth, relative to usual care on EPDS threshold data. The NMA for EPDS threshold score at 6 weeks postnatally found that the CBT-based intervention had the biggest effect relative to usual care (although this was not statistically significant), and had the highest probability of being the best (probability 0.84). The authors concluded that, in general, the intervention effects were inconclusive.

Two selective preventive intervention RCTs (Zlotnick 2011 and Chabrol 2002) were included in an NMA that was used to compare the effects of a CBT-based intervention, IPT-based intervention, education on preparing for parenting, midwife-led debriefing after childbirth and peer support relative to usual care on EPDS mean scores. The NMA found that the interventions associated with the greatest reduction in EPDS mean score were the IPT-based intervention (at 3-4 months) and the CBT-based intervention (at 6-8 weeks). However, not all interventions provided information about intervention effects at each time, making it difficult to draw inferences across all interventions at each time. In general, the intervention effects were inconclusive and the CIs were wide. Although the structured psychological interventions appeared to be the most beneficial interventions, the evidence for the effect of CBT-based intervention came from one RCT (N=158) that was judged to be at high risk of bias, and the evidence for the effect of IPT-based interventions came from a small pilot RCT (N=163); as such, the results should be treated with caution.

Of the 13 indicated preventive intervention RCTs, six were included in the NMA.[62] One RCT (Morrell 2009a/2009b) was included in a NMA that was used to compare the effects of a CBT-based intervention, PCA-based intervention and education on preparing for parenting relative to usual care on EPDS threshold. The NMA for EPDS threshold score at 6 months postnatally found that the CBT-based intervention had reduced odds of high EPDS scores compared with usual care; however, the effect was not statistically significant. The authors concluded that the intervention effects were, in general, inconclusive.

Four indicated preventive intervention RCTs (Grote 2009; Morrell 2009a/2009b; Munoz 2007; Gorman 1997) were included in a NMA that was used to compare the effects of CBT-based intervention, IPT-based intervention, educational information, midwifery continuous care, peer support, PCA-based intervention and promoting parent–infant interaction relative to usual care on EPDS mean scores.[63] The NMA found that the interventions associated with the greatest reduction in EPDS mean scores were IPT-based interventions (at 6-7 months) and CBT-based interventions (at 12 months), although none of the effects were statistically significant. Not all interventions provided information about intervention effects at each time, making inferences across all treatments at each time difficult. In general, the intervention effects were inconclusive and the CIs were wide. The most beneficial treatments appeared to be IPT-based intervention, educational information, CBT-based intervention and PCA-based intervention.

[61] Two RCTs were excluded because of a lack of EPDS data, and one RCT was excluded because there was no usual care comparator.
[62] Six RCTs were excluded because of a lack of available EPDS data, and one RCT was excluded because it could not be connected to the main network of evidence.
[63] Austin 2008 and Ginsburg 2012 were also included in the NMA, providing data for the comparison of CBT-based interventions versus educational information interventions.

Table C4-13 Summary of findings (prevention) – structured psychological interventions versus usual care

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants** (studies)	[Risk of bias]*** Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression symptomatology					
CBT-based Selective preventive – 6 weeks postnatally EPDS threshold score	*Study population*		OR 0.46 (0.18, 1.10)	258 (1 study)	[High risk of bias] ●○○○ Very low
	NE	NE			
	Moderate				
	NE	NE			
CBT-based Indicated preventive – 6 months postnatally EPDS threshold score	*Study population*		OR 0.59 (0.26, 1.38)	595 (1 study)	[Low risk of bias] ●●●○ Moderate
	NE	NE			
	Moderate				
	NE	NE			
Depression mean scores					
CBT-based Selective preventive – 6 weeks postnatally EPDS	MD -1.75 (-4.25, 0.71)		-	258 (1 study)	[High risk of bias] ●○○○ Very low
CBT-based Indicated preventive – 3 months postnatally EPDS	MD -1.38 (-6.07, 3.87)		-	41 (1 study)	[Unclear risk of bias] ●○○○ Very low
IPT-based Selective preventive – 3 months postnatally EPDS	MD -1.85 (-5.60, 2.14)		-	54 (1 study)	[Unclear risk of bias] ●○○○ Very low
CBT-based Indicated preventive – 6 months postnatally EPDS	MD -0.34 (-3.06, 3.01)		-	595 (1 study)	[Low risk of bias] ●●●○ Moderate
IPT-based Indicated preventive – 6 months postnatally EPDS	MD -4.25 (-7.87, 0.43)		-	98 (2 studies)	[Unclear risk of bias] ●○○○ Very low
CBT-based Indicated preventive – 12 months postnatally EPDS	MD -2.18 (-5.39, 1.15)		-	636 (2 studies)	[Low risk of bias; unclear risk of bias] ●●○○ Low

Evidence Statements:

Therapies delivered to an individual

A single cognitive behaviour prevention session (individual) during hospitalisation appears to have no effect on depression symptomatology (very low certainty evidence) and appears to have no effect on depression mean scores (very low certainty evidence) at follow-up (6 weeks postnatally) compared with usual care, in pregnant women with significantly higher than average risk of PND due to one or more social risk factors.

A CBT-based intervention (individual with home visits) may have an effect on depression symptomatology (moderate certainty evidence) but has no effect on depression mean scores (moderate certainty evidence) at follow-up (6 months postnatally) compared with usual care, in postpartum women who scored 12 or more on the EPDS.

An IPT-based intervention (individual) for low-income pregnant women with intimate partner violence appears to have no effect on depression mean scores at follow-up (3 months postnatally) (very low certainty evidence) compared with usual care, in women with significantly higher than average risk of PND due to one or more social risk factors.

An IPT-based intervention (individual) appears to have no effect on depression mean scores at follow-up (6 months postnatally) (very low certainty evidence) compared with usual care, in pregnant and postpartum women at high risk of developing PND on the basis of psychological risk factors, above average scores on psychological measures or other indications of a predisposition to PND.

Therapies delivered to a group

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants** (studies)	[Risk of bias]*** Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
A CBT-based intervention (group) appears to have no effect on depression mean scores) at follow-up (3 months postnatally) (very low certainty evidence) compared with usual care, in low-income predominantly Latina women who screened positive for a major depressive episode and/or who scored 16 or more on the CES-D.					
Therapies delivered to a group or individual					
A CBT-based intervention (group or individual) appears to have no effect on depression mean scores at follow-up (12 months postnatally) (low certainty evidence) compared with usual care, in pregnant and postpartum women at high risk of developing PND on the basis of psychological risk factors, above average scores on psychological measures or other indications of a predisposition to PND.					
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). ** Refers to number randomised rather than analysed. Number analysed in the SR was not reported. *** Risk of bias as assessed in Morrell 2016. This was translated for the purposes of this report into a GRADE quality of the evidence rating using the following rules: Overall risk of bias low = GRADE high quality. Overall risk of bias low but only one study = GRADE moderate quality. Overall risk of bias high = GRADE very low quality. Overall risk of bias unclear = GRADE very low quality.					

Source: Morrell 2016, Table 36, Table 46

Abbreviations: CBT, cognitive behaviour therapy; CI, confidence interval; EPDS, Edinburgh Postnatal Depression Scale; IPT, interpersonal therapy; MD, mean difference; NE, not estimable (from data in SR); OR, odds ratio; PND, postnatal depression; SR, systematic review.

C4.2.2 Directive counselling

The literature search identified no SRs that relate to this intervention.

Table C4-14 Summary of findings (prevention) – directive counselling

Evidence Statement: *There is no RCT evidence for directive counselling in women who are considered to be 'at risk' of developing mental health problems in the perinatal period.*

C4.2.3 Non-directive counselling

The literature search identified no SRs with RCTs that relate to non-directive counselling for the prevention of mental health problems in the perinatal period.

Table C4-15 Summary of findings (prevention) – non-directive counselling

Evidence Statement: *There is no RCT evidence for non-directive counselling in women who are considered to be 'at risk' of developing mental health problems in the perinatal period.*

C4.2.4 Case management/individual treatment

Only one SR of prevention using case management or individual treatment (NICE 2015) was identified in the literature search. NICE 2015 included one RCT from the United States (N=34) that assessed in-hospital case management and individualised treatment in women who had preterm delivery and low birthweight babies. The intervention was coordinated by one clinician (care manager) from an interdisciplinary team which included pediatrics, psychology, nursing, and physical therapy. The intervention addressed four major domains including: infant behaviour and characteristics; family organisation and functioning; caregiving environment; and home discharge and community resources.

C4.2.4.1 Case management and individualised treatment versus treatment as usual

A single study provided very low certainty evidence for a large effect of case management and individualised treatment on preventing depression symptomatology for women who had a preterm delivery or low birthweight baby (**Table C4-16**), with women in the intervention group showing a 75% risk reduction for scoring above threshold on a depression scale (BDI ≥9). However, confidence in this effect estimate is very low due to risk-of-bias concerns (statistically significant group differences in maternal age at baseline with older mean age in the intervention group) and very serious imprecision.

Table C4-16 Summary of findings (prevention) – case-management and individualised treatment versus treatment as usual

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression symptomatology					
Post-treatment – ITT analysis (at-risk populations) BDI ≥9 (mean 5 weeks)	*Study population*		RR 0.25 (0.06, 1.05)	34 (1 study)	●○○○ Very low (a,b,c)
	438 per 1000	109 per 1000 (26, 459)			
	Moderate				
	438 per 1000	109 per 1000 (26, 460)			
Post-treatment – Available case analysis (at-risk populations) BDI ≥9 (mean 5 weeks)	*Study population*		RR 0.25 (0.06, 1.05)	34 (1 study)	●○○○ Very low (a,b,c)
	438 per 1000	109 per 1000 (26, 459)			
	Moderate				
	438 per 1000	109 per 1000 (26, 460)			
Maternal sensitivity					
Post-treatment – ITT analysis (at-risk populations) Behavioural observation: Maternal sensitivity (mean 5 weeks)	*Study population*		RR 1.4 (0.95, 2.05)	30 (1 study)	●○○○ Very low (b,c,d)
	667 per 1000	933 per 1000 (633, 1000)			
	Moderate				
	667 per 1000	934 per 1000 (634, 1000)			
Post-treatment – available case analysis (at-risk populations) Behavioural observation: Maternal sensitivity (mean 5 weeks)	*Study population*		RR 1.4 (0.95, 2.05)	30 (1 study)	●○○○ Very low (b,c,d)
	667 per 1000	933 per 1000 (633, 1000)			
	Moderate				
	667 per 1000	934 per 1000 (634, 1000)			

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Evidence Statement: *In-hospital case management and individualised treatment may have an effect[64] on depression symptomatology at endpoint or first measurement (very low certainty evidence) compared with treatment as usual in women who are considered to be 'at risk' of developing mental health problems due to preterm delivery.* *In-hospital case management and individualised treatment appears to have no effect on maternal sensitivity at endpoint or first measurement (very low certainty evidence) compared with treatment as usual in women who are considered to be 'at risk' of developing mental health problems due to preterm delivery.*					
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. Risk of bias due to statistically significant group differences at baseline b. Total number of events is less than 300 (a threshold rule-of-thumb) c. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25) d. High risk of selection bias due to unclear allocation concealment and statistically significant baseline difference in maternal age (29.7 in intervention group and 25.9 in control group)					

Source: NICE 2015, Table 46, Table 58

Abbreviations: BDI, Beck Depression Inventory; CI, confidence interval; ITT, intention-to-treat; RR, relative risk; SMD, standardised mean difference.

C4.2.5 Self-help and facilitated self-help

The literature search identified no SRs that relate to this intervention.

Table C4-17 Summary of findings (prevention) – self-help

Evidence Statement: *There is no RCT evidence for self-help and facilitated self-help in women who are considered to be 'at risk' of developing mental health problems in the perinatal period.*

C4.2.6 Post-traumatic birth counselling

The literature search identified no SRs that relate to this intervention.

Table C4-18 Summary of findings (prevention) – post-traumatic birth counselling

Evidence Statement: *There is no RCT evidence for post-traumatic birth counselling in women who are considered to be 'at risk' of developing mental health problems in the perinatal period.*

[64] RR 0.25 (95% CI 0.06, 1.05); P=0.06

C4.2.7 Post-miscarriage counselling

The literature search identified no SRs that relate to this intervention.

Table C4-19 Summary of findings (prevention) – post-miscarriage counselling

Evidence Statement: *There is no RCT evidence for post-miscarriage counselling in women who are considered to be 'at risk' of developing mental health problems in the perinatal period.*

C4.3 PREVENTION WITH ONLINE INTERVENTIONS

Only one SR of prevention using online interventions (Ashford 2016) was identified in the literature search. Ashford 2016 included four studies of web- or computer-based interventions for the prevention of mental health problems. No additional studies of web- or computer-based interventions were identified by the other SRs included in the current report.

Of the four prevention studies included in Ashford 2016, two were RCTs, one was a modified partially randomised patient preference trial, and one was a quasi-experimental pre-test/post-test study without a control group. The preventive interventions varied in the mental health issue and timeframe for which they were developed. For pregnant women, web- and computer-based interventions were developed for stress and anxiety and mental health of women diagnosed with preterm labour. For the postpartum period, web-based interventions were developed for stress and for overall psychological health of women and their partners following miscarriage. Due to the small number of studies and their heterogeneous methodological designs and quality, the authors of the Ashford 2016 SR considered that data synthesis in the form of a meta-analysis would be inappropriate and therefore information was synthesised and reported narratively.

Only one of the RCTs identified in the Ashford 2016 SR compared an online intervention with an offline intervention. The RCT (reported in a doctoral dissertation) compared the web-based 'LivingSMART' intervention from the United States with a face-to-face version of the same program. 'LivingSMART' was a postpartum stress management program based on Herbert Benson's theory of relaxation responses and stress management techniques. The web-based program offered audio instruction of the techniques and weekly emails. The educational information, handouts, additional resources, and relaxation training, was the same for each group. Four treatment sections were covered over a four-week period, with each week divided into three areas: information, relaxation, and exercises. Exercises and homework assignments were recommended for practice during the week. Social support was provided to each of the two groups. In the Internet-based program, participants were able to email or call the researcher with questions, comments and concerns, for the duration of the study. Individuals in the face-to-face program were able to have direct contact with the researcher as well as with other participants. Although the study did not recruit women with specific risk factors for mental health problems, 47.4% of study participants reported being diagnosed with depression or anxiety at some point in their lives, and 19% were currently taking medication for the treatment of depression or anxiety.

A literature search was conducted to identify RCTs of online interventions published after the literature search date of the Ashford 2016 SR. Only those studies that compared an online intervention with an offline version of the same intervention were considered eligible. No additional studies, published in full, were identified in the literature search update.

C4.3.1.1 Online intervention versus offline (face-to-face) intervention

On the basis of a single RCT, there was no significant reduction in mean depression scores in women who received a web-based postpartum stress management intervention or in women who received a face-to-face version of the same intervention. Post-intervention stress and anxiety mean scores were significantly lower than pre-intervention scores in the web-based intervention group, whereas the face-to-face control group showed no significant improvement post-intervention.[65] However, there was no significant between-group differences in post-intervention stress, anxiety or depression mean scores, and no significant differences between groups in mean change from baseline scores.[66] Attrition in the intervention group was 41.1% versus 33.3% in the control group.

Table C4-20 Summary of findings (prevention) – online intervention versus offline (face-to-face) intervention

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Depression mean scores					
One week post-treatment – available case analysis BDI-II	MD 2.05 (-1.67, 5.77)		-	38 (1 study)	●○○○ Very low (a,b,c)
Anxiety mean scores					
One week post-treatment – available case analysis STAI	MD 0.73 (-3.52, 4.98)		-	38 (1 study)	●○○○ Very low (a,b,c)
Perceived stress mean scores					
One week post-treatment – available case analysis PSS	MD -0.43 (-3.85, 2.99)		-	38 (1 study)	●○○○ Very low (a,b,c)
Evidence Statement: *A web-based postpartum stress management intervention appears to have no effect on depression mean scores (very low certainty evidence), anxiety mean scores (very low certainty evidence), or perceived stress mean scores (very low certainty evidence), at one week post-intervention compared with a face-to-face version of the program in postpartum women (with no specific risk factors for developing mental health problems in the perinatal period).*					
Footnotes: * The 'assumed risk' for the *study population* is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The *moderate* risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. High risk of bias due to unclear allocation concealment and high rate of attrition b. Study not targeted to women at risk of mental health problems in the perinatal period c. 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5)					

Source: Raw data (mean and SD) taken from King 2009, Table 1. Post-intervention between-group difference calculated post hoc using Review Manager 5.3.

Abbreviations: BDI, Beck Depression Inventory; CI, confidence interval; MD, mean difference; PSS, Perceived Stress Scale; STAI, State Trait Anxiety Inventory.

[65] Post hoc tests showed a 24% reduction in stress and a 25% reduction in anxiety in the online group, both of which were statistically significant. The control group findings showed a 10% reduction in anxiety and a 7% reduction in stress, which were not statistically significant. Both groups experienced a reduction in depression that was not statistically significant (16% reduction in the online group and 7% reduction in the control group).

[66] Between-group differences not reported in King 2009 thesis dissertation but was calculated for the purposes of the current Review using Review Manager 5.3.

C4.4 PREVENTION WITH PHARMACOLOGICAL INTERVENTIONS

C4.4.1 Antidepressants

Of the two SRs identified in the scoping and updated searches, NICE 2015 was chosen as the foundation review due to its assessment of the evidence using SoF tables. Both SRs included the same two RCTs – one comparing SSRIs (sertraline) with placebo and one comparing TCAs (nortriptyline).

C4.4.1.1 SSRIs

The NICE 2015 SR provided data for one comparison involving an SSRI: sertraline versus placebo.

Table C4-21 summarises the evidence and provides Evidence Statements relating to the comparison between sertraline and placebo. There was very low certainty evidence that prophylaxis with sertraline does not appear to prevent recurrence of postnatal depression; however, the reduction in risk was very substantial (RR 0.14) and the upper 95% CI only just crosses 1.0, indicating that there may be a significant effect, but the analysis is underpowered. In terms of adverse events, there was very low certainty evidence of no difference in dizziness between sertraline and placebo, and a significantly increased risk of drowsiness.

Table C4-21 Summary of findings (prevention) – SSRIs (sertraline) versus placebo

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE) Comments
	Assumed risk Control	Corresponding risk Intervention			
Recurrence					
Recurrence of depression	*Study population*		RR 0.14	22	●○○○
Post-treatment	500 per 1000	70 per 1000 (10, 535)	(0.02, 1.07)	(1 RCT)[67]	Very low (a,b)
HRSD ≥ 15 on two occasions and DSM-IV	*Moderate*				
(17 weeks)	500 per 1000	70 per 1000 (10, 535)			
Adverse events					
Dizziness	*Study population*		RR 4.57	22	●○○○
Post-treatment	125 per 1000	571 per 1000 (86, 1000)	(0.69, 30.2)	(1 RCT)[68]	Very low (a,b)
(17 weeks)	*Moderate*				
	125 per 1000	571 per 1000 (86, 1000)			
Drowsiness	*Study population*		**RR 1.93**	22	●○○○
Post-treatment	500 per 1000	965 per 1000 (500, 1000)	**(1.00, 3.74)**	(1 RCT)[69]	Very low (a,b)
(17 weeks)	*Moderate*				
	500 per 1000	965 per 1000 (500, 1000)			

[67] NICE 2015 (Wisner 2004).
[68] NICE 2015 (Wisner 2004).
[69] NICE 2015 (Wisner 2004).

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE) Comments
	Assumed risk Control	Corresponding risk Intervention			
Evidence Statements: *Prophylaxis with sertraline appears to have no effect on (but may reduce) the risk of recurrence of depression at 17 weeks post-treatment compared with placebo, in women with one or more psychological risk factors for the development of postnatal depression (very low certainty evidence).* *Prophylaxis with sertraline appears to have no effect on the risk of dizziness at 17 weeks post-treatment compared with placebo, in women with one or more psychological risk factors for the development of postnatal depression (very low certainty evidence).* *Prophylaxis with sertraline may increase the risk of drowsiness at 17 weeks post-treatment compared with placebo, in women with one or more psychological risk factors for the development of postnatal depression, from an absolute risk of 50% to 97% (very low certainty evidence).*					
Footnotes: * The 'assumed risk' for the study population is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The moderate risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI). a. Unclear attrition bias and independence of data assumption contravened. b. Total population size is less than 400 (a threshold rule-of-thumb) and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25).					

Source: NICE 2015 – Table 281 and Table 283.

Abbreviations: CI, confidence interval; DSM-IV, Diagnostic and Statistical Manual of Mental Disorders, 4th Edition; HRSD, Hamilton Rating Scale for Depression; PND, postnatal depression; RCT, randomised controlled trial; RR, relative risk; SMD, standardised mean difference.

C4.4.1.2 TCAs

The NICE 2015 SR provided data for one comparison involving a TCA: nortriptyline versus placebo.

Table C4-22 summarises the evidence and provides Evidence Statements relating to the comparison between nortriptyline and placebo. There was very low certainty evidence that prophylaxis with nortriptyline does not appear to prevent recurrence of postnatal depression. In terms of adverse events, there was very low certainty evidence of no difference in discontinuation due to adverse events, but the 95% CI was very wide, indicating a lack of power in the analysis. There was moderate certainty evidence that use of nortriptyline results in an increased risk of constipation compared with placebo.

Table C4-22 Summary of findings (prevention) – TCAs (nortriptyline) versus placebo

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Recurrence of depression					
Recurrence of major depression	*Study population*		RR 0.96	51	●●○○
Post-treatment	240 per 1000	230 per 1000 (86, 622)	(0.36, 2.59)	(1 RCT)[70]	Low (a)
HRSD ≥ 15 and RDC for major depression	*Moderate*				
(22 weeks)	240 per 1000	230 per 1000 (86, 622)			

[70] NICE 2015 (Wisner 2001).

Outcomes (follow-up)	**Illustrative comparative risks***		**Relative effect (95% CI)**	**No. of participants (studies)**	**Certainty of the evidence (GRADE)**
	Assumed risk Control	**Corresponding risk Intervention**			
Recurrence of major depression postpartum	*Study population*		RR 1.20 (0.57, 2.55)	51 (1 RCT)[70]	●●○○ Low (a)
Long-term follow-up – 25-103 weeks post intervention	320 per 1000	384 per 1000 (182, 816)			
HRSD ≥ 15 and RDC for major depression	*Moderate*				
(26 weeks)	320 per 1000	384 per 1000 (182, 816)			
Side effects					
Discontinuation due to adverse events	*Study population*		RR 0.32 (0.01, 7.53)	51 (1 RCT)[70]	●●○○ Low (a)
Post-treatment	40 per 1000	13 per 1000 (0, 301)			
(20 weeks)	*Moderate*				
	40 per 1000	13 per 1000 (0, 301)			
Constipation	*Study population*		**RR 3.21 (1.55, 6.64)**	51 (1 RCT)[70]	●●●○ Moderate (b)
Post-treatment	240 per 1000	770 per 1000 (372, 1000)			
(20 weeks)	*Moderate*				
	NR	NR			

Evidence Statements:

Prophylaxis with nortriptyline appears to have no effect on the risk of recurrence of depression at 22 weeks post-treatment, or 26 weeks post-intervention, compared with placebo, in women with one or more psychological risk factors for the development of postnatal depression (low certainty evidence).

Prophylaxis with nortriptyline increases the risk of constipation at 22 weeks post-treatment compared with placebo, in women with one or more psychological risk factors for the development of postnatal depression, from an absolute risk of 24% to 77% (moderate certainty evidence).

Footnotes:

* The 'assumed risk' for the study population is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The moderate risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).

a. Total population size is less than 400 (a threshold rule-of-thumb) and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25).

b. Total population size is less than 400 (a threshold rule-of-thumb).

Source: NICE 2015 – Table 282 and Table 284.

Abbreviations: CI, confidence interval; HRSD, Hamilton Rating Scale for Depression; PND, postnatal depression; RCT, randomised controlled trial; RDC, Research Diagnostic Criteria; RR, relative risk; SMD, standardised mean difference.

C4.4.2 Antipsychotics

No SRs or individual RCTs were identified that assessed the effect of antipsychotics on the prevention of antenatal or postnatal mental health problems in women who are considered to be 'at risk', or maternal side effects.

Table C4-23 Summary of findings (prevention) – antipsychotics

Evidence Statement: *There is no RCT evidence for antipsychotics as an intervention for pregnant women who are considered to be 'at risk' of developing mental health problems in the perinatal period.*

C4.4.3 Anticonvulsants

No SRs or individual RCTs were identified that assessed the effectiveness of anticonvulsants on the prevention of antenatal or postnatal mental health problems in women who are considered to be 'at risk'.

Table C4-24 Summary of findings (prevention) – anticonvulsants

Evidence Statement: *There is no RCT evidence for anticonvulsants as an intervention for pregnant women who are considered to be 'at risk' of developing mental health problems in the perinatal period.*

C4.4.4 Benzodiazepines and z-drugs

No SRs or individual RCTs were identified that assessed the effectiveness of benzodiazepines and z-drugs on the prevention of antenatal or postnatal mental health problems in women who are considered to be 'at risk'.

Table C4-25 Summary of findings (prevention) – benzodiazepines and z-drugs

Evidence Statement: *There is no RCT evidence for benzodiazepines and z-drugs as an intervention for pregnant women who are considered to be 'at risk' of developing mental health problems in the perinatal period.*

C4.4.5 Lithium

No SRs or individual RCTs were identified that assessed the effectiveness of lithium on the prevention of antenatal or postnatal mental health problems in women who are considered to be 'at risk'.

Table C4-26 Summary of findings (prevention) – lithium

Evidence Statement: *There is no RCT evidence for lithium as an intervention for pregnant women who are considered to be 'at risk' of developing mental health problems in the perinatal period.*

C4.5 PREVENTION WITH COMPLEMENTARY INTERVENTIONS

C4.5.1 Omega-3 fatty acids

One SR was identified in the scoping and updated searches that assessed the use of omega-3 fatty acids compared with placebo to prevent postnatal depression (Miller 2013). The single included study (Mozurkewich 2013) examined separately two types of omega-3 fatty acids: eicosapentaenoic acid (EPA)-rich fish oil supplements and docosahexaenoic acid (DHA)-rich fish oil supplements. Each will be considered separately.

C4.5.1.1 Eicosapentaenoic acid (EPA)

Table C4-27 summarises the evidence and provides Evidence Statements relating to the comparison between EPA and placebo. There was moderate certainty evidence that prophylaxis with EPA compared with placebo does not reduce depression mean score or prevent diagnosis with MDD at 6-8 weeks postnatal in women at risk of developing postnatal depression.

Table C4-27 Summary of findings (prevention) – omega-3 fatty acids (EPA) versus placebo

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Diagnosis of MDD					
Presence of MDD	*Study population*		RR 1.58	80	●●●○
Post-treatment - ITT analysis	49 per 1000	77 per 1000 (14, 438)	(0.28, 8.94)	(1 RCT)[71]	Moderate(a)
MINI	*Moderate*				
(6-8 weeks post-partum)	NR	NR			
Mean depression score					
Mean depression score	MD 0.70		-	80	●●●○
Post-treatment – ITT analysis	(-1.78, 3.18)			(1 RCT)[72]	Moderate(a)
BDI					
(6-8 weeks post-partum)					
Evidence Statements:					
Prophylaxis with EPA has no effect on the risk of being diagnosed with major depressive disorder at 6-8 weeks postpartum compared with placebo, in women at risk of developing postnatal depression (moderate certainty evidence).					
Prophylaxis with EPA has no effect on depression mean score at 6-8 weeks postpartum compared with placebo, in women at risk of developing postnatal depression (moderate certainty evidence).					

Footnotes:

* The 'assumed risk' for the study population is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The moderate risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).

a. Total population size is less than 400 (a threshold rule-of-thumb) and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25).

Source: Miller 2013 – Analysis 2.1 and 2.2.

Abbreviations: CI, confidence interval; BDI, Beck Depression Inventory; EPDS, Edinburgh Postnatal Depression Scale; HRSD, Hamilton Rating Scale for Depression; MDD, major depressive disorder; RCT, randomised controlled trial; RR, relative risk; SMD, standardised mean difference.

C4.5.1.2 Docosahexaenoic acid (DHA)

Table C4-28 summarises the evidence and provides Evidence Statements relating to the comparison between DHA and placebo. There was moderate certainty evidence that prophylaxis with EPA compared with placebo does not reduce depression mean score or prevent diagnosis with MDD at 6-8 weeks postnatal in women at risk of developing postnatal depression.

[71] Miller 2013 (Mozurkewich 2013).

[72] Miller 2013 (Mozurkewich 2013).

Table C4-28 Summary of findings (prevention) – omega-3 fatty acid (DHA) versus placebo

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Diagnosis of MDD					
Presence of MDD	*Study population*		RR 1.08	79	●●●○
Post-treatment - ITT analysis	49 per 1000	53 per 1000 (8, 357)	(0.16, 7.28)	(1 RCT)[73]	Moderate(a)
MINI	*Moderate*				
(6-8 weeks post-partum)	NR	NR			
Mean depression score					
Mean depression score	MD -0.20		-	79	●●●○
Post-treatment – ITT analysis	(-2.61, 2.21)			(1 RCT)[74]	Moderate(a)
BDI					
(6-8 weeks post-partum)					
Evidence Statements:					
Prophylaxis with DHA has no effect on the risk of being diagnosed with major depressive disorder at 6-8 weeks postpartum compared with placebo, in women at risk of developing postnatal depression (moderate certainty evidence).					
Prophylaxis with DHA has no effect on depression mean score at 6-8 weeks postpartum compared with placebo, in women at risk of developing postnatal depression (moderate certainty evidence).					

Footnotes:

* The 'assumed risk' for the study population is calculated using the mean baseline risk from the studies in the meta-analysis (i.e. total number of events in the control/comparison group divided by the total number of patients in the control/comparison group). The moderate risk scenario is calculated using the median control/comparison group risk from the studies in the meta-analysis. The 'corresponding risk' (and its 95% CI) is based on the assumed risk in the control/comparison group and the relative effect of the intervention (and its 95% CI).

a. Total population size is less than 400 (a threshold rule-of-thumb) and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25).

Source: NICE 2015 – Analysis 3.1 and 3.2.

Abbreviations: CI, confidence interval; BDI, Beck Depression Inventory; EPDS, Edinburgh Postnatal Depression Scale; HRSD, Hamilton Rating Scale for Depression; MDD, major depressive disorder; PND, postnatal depression; RCT, randomised controlled trial; RR, relative risk; SMD, standardised mean difference.

[73] Miller 2013 (Mozurkewich 2013).

[74] Miller 2013 (Mozurkewich 2013).

C4.5.2 St John's wort

No SRs or individual RCTs were identified that assessed the effectiveness of St John's wort on the prevention of antenatal or postnatal mental health problems in women who are considered to be 'at risk'.

Table C4-29 Summary of findings (prevention) – St John's wort

Evidence Statement:
There is no RCT evidence for St John's wort as an intervention for pregnant women who are considered to be 'at risk' of developing mental health problems in the perinatal period.

C4.5.3 Gingko biloba

No SRs or individual RCTs were identified that assessed the effectiveness of gingko biloba on the prevention of antenatal or postnatal mental health problems in women who are considered to be 'at risk',.

Table C4-30 Summary of findings (prevention) – gingko biloba

Evidence Statement:
There is no RCT evidence for Gingko biloba as an intervention for pregnant women who are considered to be 'at risk' of developing mental health problems in the perinatal period.

C4.6 PREVENTION WITH PHYSICAL INTERVENTIONS

C4.6.1 Exercise

Only one SR of prevention using physical interventions (Daley 2015) was identified in the literature search. The SR identified one RCT (N=34) that assessed experiential exercise as part of a mindfulness intervention in women at risk of antenatal depression. However, on the basis of participant baseline symptoms, the RCT was considered to have recruited depressed women and the intervention was therefore classified as a treatment rather than preventive intervention.

Table C4-31 Summary of findings (prevention) – exercise

Evidence Statement:
There is no RCT evidence for exercise as an intervention for pregnant women who are considered to be 'at risk' of developing mental health problems in the perinatal period.

C4.6.2 Yoga

The literature search identified no SRs that relate to this intervention.

Table C4-32 Summary of findings (prevention) – yoga

Evidence Statement:
There is no RCT evidence for yoga as an intervention for pregnant women who are considered to be 'at risk' of developing mental health problems in the perinatal period.

C4.6.3 Acupuncture

Only one SR of prevention using acupuncture (NICE 2015) was identified in the literature search. The RCT from Brazil (N=29) assessed acupuncture delivered over a 12-week period to postpartum women with preterm infants with very low birthweight. Placebo acupuncture was applied using the same needles customised to not perforate skin, and a toothpick to create the sensation of needle perforation.

C4.6.3.1 Acupuncture versus placebo acupuncture

There was no statistically or clinically significant effect of acupuncture on mean anxiety scores at the end of intervention (**Table C4-33**).

Table C4-33 Summary of findings (prevention) – acupuncture versus placebo acupuncture

Outcomes (follow-up)	Illustrative comparative risks*		Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE)
	Assumed risk Control	Corresponding risk Intervention			
Anxiety mean scores					
Post intervention- Available case analysis STAI (12 weeks)	SMD 0.56 (-0.19, 1.3)		-	29 (1 study)	●○○○ Very low (a,b)
Evidence Statement: *Acupuncture (delivered over 12 weeks) appears to have no effect on anxiety mean scores at endpoint or first measurement (very low certainty evidence) compared to placebo acupuncture, in women who are considered to be 'at risk' of developing mental health problems due to preterm delivery and low birthweight.*					
Footnotes: *The basis for the assumed risk (for example the median control group risk across studies) is provided in footnotes. The corresponding risk (and its 95% CI) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI). a. High risk of bias in several domains b. Total population size is less than 400 (a threshold rule-of-thumb) and 95% CI crosses both line of no effect and measure of appreciable benefit or harm (SMD -0.5/0.5 or RR 0.75/1.25)					

Source: NICE 2015, Table 341

Abbreviations: CI, confidence interval; RR relative risk; SMD, standardised mean difference; STAI, State-Trait Anxiety Inventory.

C4.6.4 Electroconvulsive therapy

No SRs or individual RCTs were identified that assessed the effectiveness of ECT on the prevention of mental health problems during pregnancy.

Table C4-34 Summary of findings (prevention) – electroconvulsive therapy

Evidence Statement: *There is no RCT evidence for ECT as an intervention for pregnant women who are considered to be 'at risk' of developing mental health problems in the perinatal period.*

C4.6.5 Transcranial magnetic stimulation

No SRs or individual RCTs were identified that assessed the effectiveness of TMS on the prevention of mental health problems during pregnancy.

Table C4-35 Summary of findings (prevention) – transcranial magnetic stimulation

Evidence Statement: *There is no RCT evidence for TMS as an intervention for pregnant women who are considered to be 'at risk' of developing mental health problems in the perinatal period.*

C4.7 ECONOMIC EVIDENCE

No relevant Australian economic evaluations of interventions for the treatment or prevention of perinatal mental health problems were identified in the literature search. Although there are existing economic evaluations from other countries, they have limited local context applicability, since the design, delivery, and cost of healthcare systems around the world very greatly.

An Australian economic analysis has been published of the psychoeducational intervention 'What Were We Thinking (WWWT)', which recruited English-speaking first-time mothers who had recently given birth and attended participating Maternal and Child Health Centres (MCHCs) in Victoria (Ride 2016). The intervention focused on parental partner relationships, management of infant behaviour and parental fatigue. Participants did not strictly meet the criteria for this Review because they did not necessarily have risk factors for mental health problems or mental health symptomatology.

A community-based before and after controlled study found that WWWT reduced postnatal mental health problems among women with no history of psychiatric disorders. The Sleep, Parenting and Relationships in a Community Setting (SPARCS) cluster-randomised trial was subsequently conducted to examine the clinical and cost-effectiveness of WWWT for the prevention of depression, anxiety and adjustment disorders in women at six months postpartum. Participants at intervention centres received WWWT-informed care during MCH visits, and both parents were invited to attend an extra six-hour first-time-parent group session on a Saturday, during which the WWWT intervention was delivered, and received printed materials to take home. Participants at control MCHCs were provided usual MCH care. At follow-up, the unadjusted 30-day prevalence of DSM-IV diagnoses of depression, anxiety and adjustment disorders was 8.57% in the control group, and 8.70% in the intervention group.

The economic evaluation took a public-sector perspective (incorporating healthcare, early childhood and social service costs), plus the participant's out-of-pocket costs. There was no statistically significant difference in either effectiveness or costs; however, the authors claimed that the point estimate of cost-effectiveness suggested WWWT might be a cost-effective intervention for the prevention of postnatal mental health problems in first-time mothers, albeit with a high degree of uncertainty surrounding the result.

C5 REFERENCES

Armstrong, K. L., J. A. Fraser, M. R. Dadds and J. Morris (1999). "A randomized, controlled trial of nurse home visiting to vulnerable families with newborns." Journal of Paediatric Child Health 35: 237-244.

Ashford, M. T., E. K. Olander and S. Ayers (2016). "Computer- or web-based interventions for perinatal mental health: A systematic review." Journal of Affective Disorders 197: 134-146.

Bloch, M., H. Meiboom, M. Lorberblatt, I. Bluvstein, I. Aharonov and S. Schreiber (2012). "The effect of sertraline add-on to brief dynamic psychotherapy for the treatment of postpartum depression: A randomized, double-blind, placebo-controlled study." Journal of Clinical Psychiatry 73: 235-241.

Chabrol, H., F. RTeissedre, J. M. Sant, N. Teisseyre, B. Rog and E. Mullet (2002). "Prevention and treatment of post-partum depression: a controlled randomized study on women at risk." Psychological Medicine 32: 1039-1047.

Chung, K. F., W. F. Yeung, Z. J. Zhang, K. P. Yung, S. C. Man and C. P. Lee (2012). "Randomized noninvasive sham-controlled pilot trial of electroacupuncture for postpartum depression. Journal of Affective Disorders." Journal of Affective Disorders 142: 115-121.

Daley, A. J., L. Foster, G. Long, C. Palmer, O. Robinson, H. Walmsley and R. Ward (2015). "The effectiveness of exercise for the prevention and treatment of antenatal depression: Systematic review with meta-analysis." BJOG: An International Journal of Obstetrics and Gynaecology 122(1): 57-62.

Freeman, M. P., M. Davis, P. Sinha, K. L. Wisner, J. R. Hibbeln and A. J. Gelenberg (2008). "Omega-3 fatty acids and supportive psychotherapy for perinatal depression: A randomized placebo-controlled study." Journal of Affective Disorders 110: 142-148.

Gong, H., C. Ni, X. Shen, T. Wu and C. Jiang (2015). "Yoga for prenatal depression: A systematic review and meta-analysis." BMC Psychiatry 15(1).

Gorman, L. (1997). Prevention of Postpartum Difficulties in a High-Risk Sample. Doctoral Dissertation, University of Iowa.

Grote, N. K., H. A. Swartz, S. L. Geibel, A. Zuckoff, P. R. Houck and E. Frank (2009). "A randomized controlled trial of culturally relevant, brief interpersonal psychotherapy for perinatal depression." Psychiatric Services 60: 313-321.

Hantsoo, L., D. Ward-O'Brien, K. A. Czaekowski, R. Guerguieva, L. H. Price and C. N. Epperson (2014). "A randomised, placebo-controlled, double-blind trial of sertraline for postpatrum depression." Psychopharmacology 231: 939-948.

Kallen, B. (1988). "Comments on teratogen update: lithium." Teratology 38(597).

Lee, E. W., F. C. Denison, K. Hor and R. M. Reynolds (2016). "Web-based interventions for prevention and treatment of perinatal mood disorders: A systematic review." BMC Pregnancy and Childbirth 16(1).

Miller, B., J., L. Murray, M. Beckmann Michael, T. Kent and B. Macfarlane (2013) "Dietary supplements for preventing postnatal depression." Cochrane Database of Systematic Reviews DOI: 10.1002/14651858.CD009104.pub2.

Molyneaux, E., L. M. Howard, H. R. McGeown, A. M. Karia and K. Trevillion (2014). "Antidepressant treatment for postnatal depression." The Cochrane database of systematic reviews 9: CD002018.

Morrell, C. J., R. Warner, P. Slade, S. Dixon, S. Walters, G. Paley, and T. Brugha (2009a). “Psychological interventions for postnatal depression: Cluster randomised trial and economic evaluation. The PoNDER trial.” Health Technology Assessment 13:i-153.

Morrell, C. J., P. Slade, R. Warner, G. Paley, S. Dixon and S. J. Walters (2009b). "Clinical effectiveness of health visitor training in psychologically informed approaches for depression in postnatal women: pragmatic cluster randomised trial in primary care." BMJ 338: a3045.

Morrell, C. J., P. Sutcliffe, A. Booth, J. Stevens, A. Scope, M. Stevenson, R. Harvey, A. Bessey, A. Cantrell, C. L. Dennis, S. Ren, M. Ragonesi, M. Barkham, D. Churchill, C. Henshaw, J. Newstead, P. Slade, H. Spiby and S. Stewart-Brown (2016). "A systematic review, evidence synthesis and meta-analysis of quantitative and qualitative studies evaluating the clinical effectiveness, the cost-effectiveness, safety and acceptability of interventions to prevent postnatal depression." Health Technology Assessment 20(37): 1-414.

Mozurkewich, E. L., C. M. Clinton, J. L. Chilimigras, S. Hamilton, L. Allbaugh and D. Berman (2013). "The Mothers, Omega-3, and Mental Health Study: a double-blind, randomized controlled trial." American Journal of Obstetrics and Gynecology 208: 21-29.

Munoz, R. F., H. N. Le, C. G. Ippen, M. A. Diaz, G. G. Urizar and J. Soto (2007). "Prevention of postpartum depression in low-income women: development of the Mamas y Bebes/Mothers and Babies Course." Cognitive and Behavioral Practice 14: 70-83.

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

Rees, A. M., M. P. Austin and G. B. Parker (2008). "Omega-3 fatty acids as a treatment for perinatal depression: Randomized double-blind placebo-controlled trial." Australian and New Zealand Journal of Psychiatry 42: 199-205.

Ride, J., P. Lorgelly, T. Tran, K. Wynter, H. Rowe and J. Fisher (2016). "Preventing postnatal maternal mental health problems using a psychoeducational intervention: the cost-effectiveness of What Were We Thinking." BMJ Open 6(11): e012086.

Sharp, D. J., C. Chew-Graham, A. Tylee, G. Lewis, L. Howard, and I. Anderson (2010). "A pragmatic randomised controlled trial to compare antidepressants with a community-based psychosocial intervention for the treatment of women with postnatal depression: the RESPOND trial." Health Technology Assessment 14: iii-iv, ix-xi, 1-153.

Su, K. P., S. Y. Huang, T. H. Chiu, K. C. Huang, C. L. Huang and H. C. Chang (2008). "Omega-3 fatty acids for major depressive disorder during pregnancy: Results from a randomized, double-blind, placebo-controlled trial." Journal of Clinical Psychiatry 69: 644-651.

Taylor, B. L., K. Cavanagh and C. Strauss (2016). "The effectiveness of mindfulness-based interventions in the perinatal period: A systematic review and meta-analysis." PLoS ONE 11(5).

Vieten, C. and J. Astin (2008). "Effects of a mindfulness-based intervention during pregnancy on prenatal stress and mood: results of a pilot study." Archives of Womens Mental Health 11: 67-74.

Wisner, K. L., B. H. Hunusa, J. m. Perel, K. S. Peindl, C. M. Piontek, and D. K. Y. Sir (2006). "Postpartum depression: a randomised trial of sertraline versus nortriptyline." Journal of Clinical Psychopharmacology 26: 353-360.

Yonkers, K. A., H. Lin, H. B. Howell, A. C. Health and L. S. Cohen (2008). "Pharmacologic treatment of postpartum women with new-onset major depressive disorder: A randomized controlled trial with paroxetine." Journal of Clinical Psychiatry 69: 659-665.

Zlotnick, C., N. M. Capezza and D. Parker (2011). "An interpersonally based intervention for low-income pregnant women with intimate partner violence: a pilot study." Archives of Womens Mental Health 14: 55-65.

Australian Perinatal Mental Health Guideline Evidence Review

Technical Report Part D

Harms of treatment and prevention interventions

Prepared by

hereco
clinical evidence for the real world

June 2017

TABLE OF CONTENTS

D1 INTRODUCTION 1

D2 METHODOLOGY 2

D2.1 Clinical questions 2

D2.2 Criteria for determining study eligibility 3

D2.3 Literature search 5

D2.3.1 Search strategy 5

D2.3.2 Study eligibility 6

D2.4 Assessment of the evidence 9

D2.5 Evidence to recommendations process 9

D2.5.1 Grading of the certainty of the evidence 9

D2.5.2 Determining the absolute increase in risk 10

D2.5.3 Drafting of Evidence Statements 11

D3 RESULTS 13

D3.1 Pharmacological 13

D3.1.1 Antidepressants 13

D3.1.2 Antipsychotics 49

D3.1.3 Anticonvulsants 78

D3.1.4 Benzodiazepines and z-drugs 90

D3.1.5 Lithium 104

D3.2 Complementary 109

D3.2.1 Omega-3 fatty acids 109

D3.2.2 St John's wort 116

D3.2.3 Gingko biloba 119

D3.3 Physical 119

D3.3.1 Electroconvulsive therapy 119

D3.3.2 Transcranial magnetic stimulation 121

D4 REFERENCES 123

TABLE OF TABLES

Table D 2-1 NHMRC Evidence Hierarchy: designation of levels of evidence according to type of research question 4
Table D2-2 PICO criteria used to inform the literature search 5
Table D 2-3 Hierarchy of evidence for the literature review 6
Table D 2-4 Evidence selection criteria - general 8
Table D 2-5 Evidence selection criteria – intervention-specific 8
Table D 3-1 Summary of results of the Evidence Review for antidepressants 15
Table D3-2 Evidence Profile table: SSRI harms 16
Table D3-3 Evidence Profile table: paroxetine harms 29
Table D3-4 Evidence Profile table: fluoxetine harms 31
Table D3-5 Evidence Profile table: sertraline harms 33
Table D3-6 Evidence Profile table: citalopram harms 35
Table D3-7 Evidence Profile table: escitalopram harms 37
Table D3-8 Evidence Profile table: fluvoxamine harms 38
Table D3-9 Evidence Profile table: SNRI/venlafaxine harms 40
Table D3-10 Evidence Profile table: NaSSA/mirtazapine harms 42
Table D3-11 Evidence Profile table: TCA harms 44
Table D3-12 Evidence Profile table: bupropion harms 47
Table D 3-13 Summary of results of the Evidence Review for antipsychotics 50
Table D3-14 Evidence Profile table: any antipsychotics 51
Table D3-15 Evidence Profile table: SGAs 58
Table D3-16 Evidence Profile table: FGAs 61
Table D3-17 Evidence Profile table: aripiprazole 64
Table D3-18 Evidence Profile table: flupenthixol 66
Table D3-19 Evidence Profile table: haloperidol, infant harms 67
Table D3-20 Evidence Profile table: olanzapine, infant harms 68
Table D3-21 Evidence Profile table: perphenazine, infant harms 70
Table D3-22 Evidence Profile table: quetiapine, infant harms 71
Table D3-23 Evidence Profile table: risperidone, infant harms 73
Table D3-24 Evidence Profile table: ziprasidone, infant harms 75
Table D3-25 Evidence Profile table: zuclopenthixol, infant harms 77
Table D 3-26 Summary of results of the Evidence Review for anticonvulsants 79
Table D3-27 Evidence Profile table: sodium valproate harms 80
Table D3-28 Evidence Profile table: carbamazepine harms 85
Table D3-29 Evidence Profile table: lamotrigine harms 88
Table D 3-30 Summary of results of the Evidence Review for benzodiazepines and z-drugs 90
Table D3-31 Evidence Profile table: benzodiazepines ± z-drugs 91
Table D3-32 Evidence Profile table: diazepam 96
Table D3-33 Evidence Profile table: temazepam 97
Table D3-34 Evidence Profile table: z-drugs 98
Table D3-35 Evidence Profile table: zolpidem 99
Table D3-36 Evidence Profile table: zopiclone 102
Table D 3-37 Summary of results of the Evidence Review for lithium 105
Table D3-38 Evidence Profile table: lithium harms 106
Table D 3-39 Summary of results of the Evidence Review for omega-3 fatty acids 109
Table D3-40 Evidence Profile table: omega-3 fatty acids 110
Table D 3-41 Summary of results of the Evidence Review for St John's wort 116
Table D3-42 Evidence Profile table: St John's wort 117
Table D3-43 Evidence Profile table: ECT harms 120
Table D3-44 Evidence Profile table: TMS harms 122

ABBREVIATIONS

AD	antidepressant
ADHD	attention deficit hyperactivity disorder
ADSI	Ankara Developmental Screening Inventory
AOR	adjusted odds ratio
ARR	adjusted relative risk
ASD	autism spectrum disorder
Benzo	benzodiazepine
BSID	Bayley Scales of Infant Development
CI	confidence interval
ECT	electroconvulsive therapy
FGA	first generation antipsychotics
GMDS	Griffiths Mental Development Scales
IQ	intelligence quotient
IUGR	intrauterine growth restriction
K-ABC	Kaufman Assessment Battery for Children
LFGA	large for gestational age
MD	mean difference
meds	medication
NA	not available
NaSSA	noradrenergic and specific serotonergic antidepressants
NE	not estimable
NR	not reported
OBS	observational studies
OR	odds ratio
P & L	pregnancy and lactation
P	pregnancy
PNAS	poor neonatal adaptation syndrome
PPH	persistent pulmonary hypertension
PPVT	Peabody Picture Vocabulary Test
PS	propensity score
RCT	randomised controlled trial
RD	risk difference
RE	risk estimate
RR	relative risk
RR	risk ratio
Rx	prescription
SFGA	small for gestational age
SGA	second generation antipsychotic
SMD	standardised mean difference
SNRI	serotonin and noradrenalin reuptake inhibitor
SR	systematic review
SRI	selective reuptake inhibitor
SRS	social responsiveness scale
SSRI	selective serotonin reuptake inhibitor
TCA	tricyclic antidepressant

D1 INTRODUCTION

The aim of this Evidence Review is to assess the evidence relating to the identification and treatment or prevention of mental health problems in women during pregnancy or the postnatal period. The following Technical Reports and associated Appendices are related to this assessment:

- Part B Technical Report and Part B Appendix – Psychosocial Assessment and Screening
- Part C Technical Report and Part C Appendix – Treatment and Prevention
- Part D Technical Report and Part D Appendix – Harms.

This Technical Report and associated Appendix (Part D) present the findings of the assessment of evidence of the harms of *interventions* used for the treatment and prevention of mental health problems in women during the antenatal or postnatal period.

D2 METHODOLOGY

D2.1 CLINICAL QUESTIONS

The four main questions relating to the harms associated with interventions for the treatment of mental health problems in pregnant or postpartum women, or prevention of mental health problems in pregnant or postpartum women identified as being at risk of developing mental health problems, were each broken down into four sub-questions based on the different populations that may potentially experience harm. It should be noted that each sub-question is broken down further into individual interventions and outcomes. The detailed definitions associated with these interventions and outcomes can be found in **Section D2.2**. All questions were addressed via systematic review.

Harms to the fetus, infant or child include any direct harms (e.g. malformations, miscarriage, perinatal mortality, neurodevelopmental disorders) and any birth outcomes that may cause subsequent harm (e.g. prenatal birth, small for gestational age, convulsions). Harm to the mother has been limited to postpartum haemorrhage; maternal side effects of treatment have been assessed in **Part C** of the **Technical Report**.

D2.1.1.1 Pharmacological interventions

Main question:

6. What are the harms that occur as a result of perinatal exposure to pharmacological interventions used for the treatment of mental health problems?

Sub-questions:

6a. What are the harms that occur to the fetus as a result of perinatal exposure to a pharmacological intervention used for the treatment of mental health problems?

6b. What are the harms that occur to the infant as a result of perinatal exposure to a pharmacological intervention used for the treatment of mental health problems?

6c. What are the harms that occur to the child as a result of perinatal exposure to a pharmacological intervention used for the treatment of mental health problems?

6d. What are the harms that occur to the mother as a result of perinatal exposure to a pharmacological intervention used for the treatment of mental health problems?

D2.1.1.2 Complementary interventions

Main question:

7. What are the harms that occur as a result of perinatal exposure to a complementary intervention used for the treatment of mental health problems?

Sub-questions:

7a. What are the harms that occur to the fetus as a result of perinatal exposure to a complementary intervention used for the treatment of mental health problems?

7b. What are the harms that occur to the infant as a result of perinatal exposure to a complementary intervention used for the treatment of mental health problems?

7c. What are the harms that occur to the child as a result of perinatal exposure to a complementary intervention used for the treatment of mental health problems?

7d. What are the harms that occur to the mother as a result of perinatal exposure to a complementary intervention used for the treatment of mental health problems?

D2.1.1.3 Physical interventions

Main question:

8. What are the harms that occur as a result of perinatal exposure to a physical intervention used for the treatment of mental health problems?

Sub-questions:

8a. What are the harms that occur to the fetus as a result of perinatal exposure to a physical intervention used for the treatment of mental health problems?

8b. What are the harms that occur to the infant as a result of perinatal exposure to a physical intervention used for the treatment of mental health problems?

8c. What are the harms that occur to the child as a result of perinatal exposure to a physical intervention used for the treatment of mental health problems?

8d. What are the harms that occur to the mother as a result of perinatal exposure to a physical intervention used for the treatment of mental health problems?

D2.2 CRITERIA FOR DETERMINING STUDY ELIGIBILITY

To determine whether an intervention causes harm, a systematic review (SR) of randomised controlled trials (RCTs) provides the highest level of evidence, as shown in the evidence hierarchy for examination of intervention questions (**Table D 2-1**). However, in cases where it is not possible or ethical to perform a RCT (as is the case when examining harms to the fetus, infant or child following maternal exposure), observational evidence should be used. The highest level of evidence in this case is a SR of prospective cohort studies, as shown in the hierarchy for examination of aetiology questions. Thus, where available, RCT evidence was used, although the majority evidence came from observational studies. Wherever possible, only observational studies with concurrent control groups were included.

For each of the intervention-based questions to be assessed by the Evidence Review (effectiveness of treatment and prevention, and harms), the EWG agreed to the appropriate level of evidence for inclusion. For the review of the harms of pharmacological, complementary and physical interventions, the EWG agreed that SRs of observational studies should be used as the basis of the review where available, with individual observational studies and SRs of case series/reports to be assessed only where higher level evidence was unavailable or inadequate. There were exceptions to this: (i) fetal, infant and child harms associated with the use of anticonvulsants during pregnancy, and postpartum haemorrhage were limited to SRs of observational studies; and (ii) the assessment of evidence for harms related to omega-3 fatty acids were limited to SRs of RCTs.

Table D 2-1 NHMRC Evidence Hierarchy: designation of levels of evidence according to type of research question[1]

Level	Intervention	Aetiology
I	A systematic review of level II studies	A systematic review of level II studies
II	A randomised controlled trial	A prospective cohort study
III-1	A pseudorandomised controlled trial (i.e. alternate allocation or some other method)	All or none[2]
III-2	A comparative study with concurrent controls: • Non-randomised, experimental trial[3] • Cohort study • Case-control study • Interrupted time series with a control group	A retrospective cohort study
III-3	A comparative study without concurrent controls: • Historical control study • Two or more single arm studies[4] • Interrupted time series without a parallel control group	A case-control study
IV	Case series with either post-test or pre-test/post-test outcomes	A cross-sectional study or case series

Table D2-2 summarises the criteria used to determine study eligibility. The population of interest varies depending on the outcome being measured: (i) for outcomes that are identified at or occur around birth, pregnant women and/or neonates are the appropriate population; (ii) for outcomes that occur around the time of breast feeding, post-partum women and/or infants are the appropriate population, and (iii) for neurodevelopmental outcomes that are measured in the years after birth, infants/children are the appropriate population. For fetal, infant or child harm, the exposure status of the mother is coupled with the outcome status of the fetus, infant or child. It should be noted that because the outcome was harm to the fetus, infant, child or mother (and the effect of the intervention on the fetus, infant or child independent of the mother's mental health status is under investigation) the maternal population for inclusion was not always specifically limited to women with mental health disorders, although that population was used preferentially where available.

Pharmacological, complementary and physical interventions that are known to be used in pregnant and postnatal women with mental health disorders were selected for assessment. For anticonvulsants, this was limited to the three drugs most commonly used as mood stabilisers: sodium valproate, carbamazepine and lamotrigine. While classified as physical therapies with exercise, yoga and acupuncture in Part C of the Technical Report, electroconvulsive therapy (ECT) and transcranial magnetic stimulation (TMS) have been included in the assessment of harms to the fetus, infant and child due to their direct impact on maternal systemic physiology.

Two types of comparator were included: (i) comparison to no intervention and (ii) comparison to other interventions. Comparison to no intervention provides a measure of whether an intervention may cause a harm, whereas comparison to another intervention provides a measure of whether the intervention of interest causes more or less harm than the comparator intervention.

The included outcomes were grouped into three categories:

- Malformations – which occur as a result of antenatal exposure, generally in the first trimester.

[1] NHRMC (2009) NHMRC additional levels of evidence and grades for recommendations for developers of guidelines. Accessed on 12 May 2017 from https://www.nhmrc.gov.au/_files_nhmrc/file/guidelines/developers/nhmrc_levels_grades_evidence_120423.pdf.

[2] All or none of the people with the risk factor(s) experience the outcome; and the data arises from an unselected or representative case series which provides an unbiased representation of the prognostic effect. For example, no smallpox develops in the absence of the specific virus; and clear proof of the causal link has come from the disappearance of small pox after large-scale vaccination.

[3] This also includes controlled before-and-after (pre-test/post-test) studies, as well as adjusted indirect comparisons (ie. utilise A vs B and B vs C, to determine A vs C with statistical adjustment for B).

[4] Comparing single arm studies ie. case series from two studies. This would also include unadjusted indirect comparisons (ie. utilise A vs B and B vs C, to determine A vs C but where there is no statistical adjustment for B).

- Pregnancy and birth outcomes – featl, infant or child harms which can occur as a result of antenatal exposure both early in pregnancy (e.g. miscarriage) and later in pregnancy (e.g. poor neonatal adaptation syndrome [PNAS] and respiratory distress), and maternal harm which can occur as a result of antenatal exposure.
- Neurodevelopmental outcomes – which may potentially occur as a result of antenatal or postnatal exposure.

Table D2-2 PICO criteria used to inform the literature search

Population	Exposure	Comparator	Outcomes
Pregnant women Post-partum women Infants or children exposed during pregnancy or postnatally	Pharmacological therapies Antidepressants Antipsychotics Mood stabilisers (including anticonvulsants,[5] benzodiazepines and z-drugs) Lithium Complementary therapies Omega-3 fatty acids St John's wort Gingko biloba Physical therapies Electroconvulsive therapy Transcranial magnetic stimulation	No exposure Exposure to an active comparator	**Fetal, infant or child harms** Malformations Major malformations Cardiac malformations Septal malformations Pregnancy and birth outcomes Neonatal mortality/still birth Miscarriage Preterm birth SFGA/IUGR PNAS Persistent pulmonary hypertension Respiratory distress Tremors Convulsions Neurodevelopmental outcomes Autism spectrum disorder ADHD Other neurodevelopmental disorders measured with validated instruments Intelligence quotient Behavioural problems Depression Anxiety **Maternal harm** Postpartum haemorrhage

Abbreviations: ADHD, attention deficit hyperactivity disorder; IUGR, intrauterine growth restriction; PNAS, poor neonatal adaptation syndrome; SFGA, small for gestational age.

D2.3 LITERATURE SEARCH

D2.3.1 Search strategy

A two-tiered search strategy was undertaken. An initial search was undertaken to identify SRs that assessed various treatments for the main mental health disorders seen during the perinatal period; these included depression, anxiety, schizophrenia and bipolar disorder. Full details of the SR search can be found in **Appendix D1.1.1** and **Appendix D1.2.1**. It should be noted that this search was conducted to identify studies not only for the assessment of harms, but also for screening interventions, and the efficacy of treatment and prevention for psychosocial and psychologic interventions, as well as additional physical interventions.

From this search, an initial list was assembled of SRs that assessed the harms to the infant associated with the pharmacological, complementary and physical therapies outlined in **Table D2-2**.The individual studies included in each SR were identified and, where possible, a 'foundation review' was identified. The process for identifying the foundation reviews is outlined in **Appendix D2**. The foundation review was defined as

[5] Sodium valproate, carbamazepine and lamotrigine only.

the SR that included the most recent and comprehensive set of data for a particular intervention and outcome, and if suitable could be included in the Evidence Review; if not suitable for inclusion, the foundation review could be used to identify relevant individual studies. Further details on the criteria for determining the suitability for inclusion of foundation SRs is provided in **Section D2.3.2**.

Based on the findings of the SR search, a second series of literature searches were carried out. These 'updated' searches aimed to identify additional SRs, and individual RCTs and observational studies, and were based on the interventions of interest as follows:

- Where a suitable foundation review was identified, the search was limited from the year of the foundation review's literature search up to October 2016. Date-limited searches were conducted for all pharmacological agents except z-drugs, and the complementary therapy omega-3 fatty acids.
- Where no suitable foundation review was identified, no initial date limit was set, and the search was conducted up to October 2016. Extended date searches were conducted for z-drugs, the complementary therapies St John's wort and Gingko biloba, and the physical therapies electroconvulsive therapy (ECT) and transcranial magnetic stimulation (TMS).

Full details of the updated searches can be found in **Appendix D1.1.2** and **Appendix D1.2.2**. It should be noted that these updated searches also aimed to identify evidence of efficacy for the pharmacologic, complementary and selected physical interventions.

Searches were conducted in the MEDLINE, Embase and PsychINFO databases (via the OVID and/or Embase.com interfaces), the Cochrane Library, and included examination of the reference lists of included SRs and individual studies.

D2.3.2 Study eligibility

The aim of the literature search was to identify the highest possible quality evidence for each intervention/outcome. As noted previously, SRs of RCTs provide the highest level of evidence for assessment of the effects of interventions; however, it may not be feasible or ethical to conduct an RCT to examine harms to offspring or women exposed to interventions used for treating or preventing mental health disorders in pregnant or postnatal women. In this case, a SR of observational studies provides an alternative. For each intervention/outcome assessed, a hierarchy of evidence was applied (see **Table D 2-3**). Starting from SRs of RCTs, evidence at each level in the hierarchy were searched, until relevant evidence was found.

The level of evidence identified for each intervention/outcome pairing had a direct impact on the grading of the quality of the evidence, as will be described in **Section D2.5.1**.

Table D 2-3 Hierarchy of evidence for the literature review

SR of RCTs
Individual RCT
SR of comparative observational studies
Individual comparative observational studies
SR of case series/single-arm studies
SR of case reports
Individual case series/report

Abbreviations: RCT, randomised controlled trial; SR, systematic review.

Citations identified in the literature searches were reviewed and evidence selection criteria were applied hierarchically. As shown in **Table D 2-4**, there was a set of standard evidence selection criteria that applied to both the SR search for all interventions, and the updated searches for pharmacological, complementary and the physical interventions ECT and TMS.

In addition, due to the volume and types of evidence available for certain interventions, additional intervention-specific criteria were applied. A large volume of evidence was identified for fetal, infant and child outcomes for antidepressants and so strict inclusion criteria were applied in order to identify 'higher quality' evidence. In order to be included in the assessment of antidepressant harms, studies had to have attempted to match or adjust the analysis for most of the main known confounders, which included maternal age, parity, smoking and alcohol. In addition, studies had to have specifically addressed confounding by indication by (i) limiting the analysis to women with a psychiatric condition, (ii) adjusting for psychiatric condition-related variables (e.g. psychiatric diagnosis, presence/number of psychiatric visits), or (iii) performing sibling analyses, in which outcomes in exposed/unexposed pairs of siblings are compared (with the assumption being that confounding by indication should be minimised because each discordant pair has the same mother). Similar criteria were applied for antipsychotics and benzodiazepines/z-drugs; however, these criteria could be relaxed for individual treatments where the 'higher quality' evidence wasn't available.

The assessment of evidence for anticonvulsants was limited to SRs of observational studies only. This is because there is a large volume of SR evidence available regarding the fetal, infant and child harms associated with anticonvulsants. This evidence is limited to a maternal population with epilepsy, and so the high level of confounding by indication known to be associated with psychiatric disorders is not present.

The assessment of evidence for omega-3 fatty acids was also limited to SRs due to the large volume available. There is a large amount of RCT evidence available for the use of omega-3 fatty acids in pregnancy, as there are no known harms, and it is believed that omega-3 fatty acids are beneficial to the mother and offspring when taken during pregnancy. For this reason, examination of the evidence for omega-3 fatty acids was limited to SRs of RCTs.

No additional evidence selection criteria were applied for St John's wort, Gingko biloba, ECT and TMS.

The ultimate aim of the evidence selection criteria was to limit the assessment of evidence to the 'highest quality' studies for each intervention grouping and type. 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. Full details of the exclusion of studies are provided in **Appendix D1.3**.

Table D 2-4 Evidence selection criteria - general

Criterion	Description
SR search	
Not a SR	Excludes individual clinical studies, narrative reviews, editorials, animal studies and in vitro studies
Wrong population	Excludes studies that are not conducted in pregnant or postnatal women, or children exposed to intervention antenatally or postnatally
Wrong intervention/exposure	Excludes studies that do not examine one of the exposures included in Table D2-2 (as well as other psychosocial, psychological and physical interventions defined in Part C of the Technical Report)
Wrong outcome	Excludes studies that do not examine one of the outcomes included in Table D2-2 (as well as other efficacy/safety/harm outcomes defined in Part C of the Technical Report)
Not in English	Excludes SRs not available in English.
Updated searches - all	
Not a clinical study	Excludes narrative reviews, editorials, animal studies and in vitro studies
Not a SR	Excludes reviews described as systematic that are not, or that limit identification of evidence to MEDLINE /PubMed only.
Wrong population	Excludes studies that are not conducted in pregnant or postnatal women, or children exposed to intervention antenatally or postnatally
Wrong intervention/exposure	Excludes studies that do not examine one of the exposures included in Table D2-2
Wrong/no comparator	Excludes studies that do not compare the exposure with no exposure or a relevant active exposure
Wrong study type	Excludes individual studies (assessment of postpartum haemorrhage [see Part C of the Technical Report] limited to SRs only)
Protocol only	Excludes publications describing a study protocol only
Duplicate data	Excludes studies that include data that has already been included from another publication
Not in English	Excludes studies not available in English
Abstract only	Excludes studies available as a conference abstract only. Where identified, an additional search will be conducted to see if the study has subsequently published.

Table D 2-5 Evidence selection criteria – intervention-specific

Criterion	Description
Updated searches - antidepressants	
Not adjusted for potential confounders	Excludes individual studies that have not attempted to minimise confounding either by study design or statistical methods
Not limited to/adjusted for maternal mental health disorder	Excludes studies that have not specifically attempted to minimise confounding by indication by limiting the included population, or matching or adjusting for disorder-related variables
Updated searches - antipsychotics	
Not adjusted for potential confounders	Excludes individual studies that have not attempted to minimise confounding either by study design or statistical methods
Not limited to/adjusted for maternal mental health disorder	Excludes studies that have not specifically attempted to minimise confounding by indication by limiting the included population, matching on disorder-related variable, or adjusting for disorder-related variables. However, where no such information was available for a specific antipsychotic, this criterion was relaxed.
Updated searches - anticonvulsants	
Wrong study type	Excludes individual studies (assessment of anticonvulsants limited to SRs only)
Updated searches -benzodiazepine and z-drugs	
Not adjusted for potential confounders	Excludes individual studies that have not attempted to minimise confounding either by study design or statistical methods
Not limited to/adjusted for maternal mental health disorder	Excludes studies that have not specifically attempted to minimise confounding by indication by limiting the included population, matching on disorder-related variable, or adjusting for disorder-related variables. However, where no such information was available for a specific antipsychotic, this criterion was relaxed.
Updated searches – omega-3 fatty acids	
Wrong study type	Excludes SRs of observational studies, and individual RCTs or observational studies (assessment of omega-3 fatty acids limited to SRs of RCTs only)

Abbreviations: RCT, randomised controlled trial; SR, systematic review.

D2.4 ASSESSMENT OF THE EVIDENCE

The highest quality evidence for each intervention/outcome was selected from the available body of evidence. Where there were no existing SR/meta-analyses appropriate for inclusion, and multiple individual studies were identified, it was necessary to perform a meta-analysis de novo for this literature review. Meta-analyses were performed using Review Manager 5.3. The results most completely adjusted for potential confounding were used preferentially where available, and the inverse variance method with a random effects model (REM) was used; meta-analyses were not performed using raw, unadjusted data from observational studies.

The full assessment of the evidence for harms for each intervention can be found in **Appendix D4**.

D2.5 EVIDENCE TO RECOMMENDATIONS PROCESS

The aim of the Evidence Review process was to identify the highest quality evidence of the harms of maternal exposure to various pharmacological, complementary and physical mental health disorder interventions. This evidence was then described and graded, and recommendations developed.

Grading of Recommendations, Assessment, Development and Evaluations (GRADE) methodology was used to grade the quality of the evidence for each intervention and outcome and translate this into recommendations and practice points. For further details about GRADE see http://www.gradeworkinggroup.org/.

According to the GRADE process, the body of evidence is summarised in either an Evidence Profile (EP), or Summary of Findings (SoF) table. For the purpose of the assessment of infant and maternal harm, the evidence was presented in EP tables, because they provide greater transparency regarding the decisions that have gone into grading the evidence. An EP table explicitly provides the following information:

- Quality assessment – this section provides information on the size of the evidence base, as well as the assessment of the quality of the evidence. The evidence is assessed according to five domains: risk of bias, inconsistency, indirectness, imprecision and publication bias. The aim of this section is to generate a 'score' for the overall quality of the evidence for each intervention/outcome.
- Summary of findings – this section provides details on the study event rates for the intervention and comparator groups in the study, the risk estimate, and the anticipated absolute effects.

It should be noted that modifications to these were required in order to accommodate the evidence base for harms, which largely consisted of observational studies. Each of these will be described in detail below. The EP tables for each intervention can be found in **Section D3**.

D2.5.1 Grading of the certainty of the evidence

The certainty of evidence assessment for GRADE involves consideration of the following five domains: risk of bias, inconsistency, indirectness, imprecision and publication bias. For an evidence base drawn from RCTs, the grading of the certainty of the evidence starts at 'high' (⊕⊕⊕⊕). However, for an evidence base drawn from observational studies (which mostly form the basis for the assessment of harms in this guideline), the grading of the certainty of the evidence starts at 'low' (⊕⊕◯◯). For the purpose of this Evidence Review, it is assumed that this 'low' grading already takes into account the general biases associated with observational study design. The certainty of the evidence is then downgraded depending on whether there is any *additional* risk of bias, and how it scores on the other four domains. There is also the opportunity to upgrade the certainty of the evidence in specific circumstances (see below).

A number of 'general rules' for handling the assessment of the certaimty of the evidence were agreed *a priori* with the EWG and Harms Expert Committee. These included:

- The certainty of the evidence could be downgraded for one or more of the five domains examined in GRADE: risk of bias, inconsistency, indirectness, imprecision, and publication bias.
- An additional downgrading by one or two levels for risk of bias could be undertaken if there were specific study-, exposure- or outcome-related concerns.
- The certainty of the evidence was downgraded one level for inconsistency where there was moderate heterogeneity within a meta-analysis (I^2 between 25% and 59%). The certainty of the evidence was downgraded two levels for inconsistency where there was substantial heterogeneity within a meta-analysis ($I^2 \geq 60\%$).
- The certainty of the evidence was downgraded one level for indirectness where the exposed population (with a mental health disorder) was compared with a non-exposed population without a mental health disorder, except in the case where the underlying condition was accounted for in the analysis using statistical methods.
- The certainty of the evidence was downgraded one level for imprecision for any one of the following reasons: (i) where the 95% confidence interval (CI) of the relative risk (RR) crossed 1.00, and where either or both the lower and upper 95% CI crossed 0.75 or 1.25; this indicated that the results included a measure of appreciable benefit and/or harm; (ii) where only a p value was provided; and (iii) where there were no events for the analysis.
- The certainty of the evidence was never downgraded due to publication bias as a comprehensive literature search was conducted to identify all relevant studies and few of the studies were identified as having been commercially funded.[6]
- The certainty of the evidence could potentially be upgraded for the following reasons, as long as it had not already been downgraded for one of the domains above: (i) large magnitude of effect, (ii) dose-response gradient, or (iii) effect of plausible residual confounding.

A number of additional 'intervention-specific rules' were also agreed with the EWG and Harms Expert Committee due to the different evidence base identified for some of the intervention types. These will be outlined in the results section where appropriate.

In some cases, downgrading resulted in the evidence base being considered as lower than 'very low' (●○○○), the lowest certainty category used by GRADE. For the purpose of this Evidence Review, an additional category was added – inadequate (○○○○). This circumstance is distinct from situations where there is no evidence. It was agreed by the EWG that evidence-based recommendations could not be made based on evidence that is inadequate; however, it was acknowledged there may be cases where it is appropriate to use this evidence to make consensus-based recommendations or practice points.

D2.5.2 Determining the absolute increase in risk

This section of the EP table generally includes the event rates seen in the intervention and comparator groups. As this is an assessment of harms, and the body of evidence is largely based on observational studies, it was not considered appropriate to include event rates. Instead, for evidence based on observational studies, the size of the exposed and unexposed/active comparator populations was included instead.

The absolute increase in risk could be calculated for dichotomous outcomes that were reported as RRs or risk differences (RD). As the evidence is based largely on data from cohort and case-control studies, in many cases the results were presented as odds ratios (ORs) instead of RRs. Where the baseline risk was <7% (identified by the risk in an unexposed group with a mental health disorder, where available), it was assumed that the OR approximates the RR and the results were interpreted as RRs. The absolute increase in risk was calculated by determining the baseline (unexposed or active treatment) risk, and multiplying by

[6] The following studies were commercially funded: Cole 2007a (GlaxoSmithKline), Cole 2007b (Genzyme) and Nulman 2015 (Wyeth-Ayerst Canada and Shopper Drug Mart, Canada). See the individual risk of bias assessments for further details.

the RR. Where the certainty of the evidence base was considered inadequate (◯◯◯◯), the absolute additional risk associated with the intervention was not calculated as the results are highly uncertain.

D2.5.3 Drafting of Evidence Statements

Whilst not a requirement of GRADE, Evidence Statements (ES) for each PICO have been developed for the purpose of the current Guideline. This has been done to facilitate the explicit weighing of benefits and harms across multiple outcomes, for the mother versus the infant, in the antenatal versus the postnatal periods.

It should be noted that evidence from RCTs can be used to infer that an intervention *causes* an outcome/harm, while observational studies provide evidence only of an *association* between an intervention and an outcome, which is not sufficient alone to prove causality. Causal inference in epidemiology requires consideration of a number of criteria including the following which, if present, may strengthen the possibility of a causal relationship, although it should be noted there are counterarguments against most of them:[7]

- A temporal relationship – exposure to the intervention precedes the condition.
- Strength of the association – the stronger the association, the more likely it is causal.
- Dose response – increasing the amount of exposure increases the risk proportionally.
- Consistency – the association is consistent when results are replicated in studies using different methods.
- Biologic plausibility – the association agrees with currently accepted understanding of biologic processes.
- Experimentation – the condition can be altered (prevented or ameliorated) by an appropriate experimental intervention.
- Specificity – a single putative cause produces a specific effect.
- Biologic coherence – the association is consistent with the natural history of the disease.
- Analogy – there are similar associations in other populations or under different settings.

The wording of the ES has thus been chosen carefully to avoid undue use of double negatives, and to convey the confidence of the findings, keeping in mind that the findings relate to the presence or absence of *associations* between exposure and the outcomes (most of which are pre-specified as 'harms' not benefits). The specific rules around the wording of the ES are as follows:

- If the effect estimate and CI include 1.00 (relative measures; RR, OR) or 0 (absolute measures; RD, mean difference [MD], standardised mean difference [SMD]):
    - and moderate or high certainty evidence is available: the phrasing "is no association between [exposure] and an increased risk of [outcome]" is used.
    - and low or very low certainty evidence is available: the phrasing "does not appear to be an association between [exposure] and an increased risk of [outcome]" is used
    - and inadequate certainty evidence is available: the phrasing "any association between [exposure] and an increased risk of [outcome] is uncertain" is used.
- If the effect estimate and CI do not include 1.00 (relative measures; RR, OR) or 0 (absolute measures; RD, MD, SMD):
    - and moderate or high certainty evidence is available: the phrasing "is an association between [exposure] and an increased risk of [outcome]" is used.
    - a low or very low quality evidence is available: the phrasing "may be an association between [exposure] and an increased risk of [outcome]" is used and the absolute risk estimates cited.

[7] See Kovesdy and Kalantar-Zadeh (2012) Observational studies vs. randomized controlled trials: avenues to causal inference in nephrology. Adv Chronic Kidney Dis 19(1): 11-18.

  - Where low quality evidence is available, but the evidence shows a large magnitude of effect:[8] the phrasing "is an association between [exposure] and an increased risk of [outcome] is used.
  - and inadequate quality evidence is available: the phrasing "appears to be an association between [exposure] and an increased risk of [outcome], but due to the inadequate quality of the evidence this association is uncertain" with no citing of the absolute risk estimates.
- Where there 'is' or 'may be' an association, and where an absolute increase or decrease in risk is available, this is also captured in the ES.

[8] 95% exceeds the minimum level of appreciable harm (RR > 1.25 or SMD < -0.5).

D3 RESULTS

The results of the assessment of evidence are presented in the following sections:

- Pharmacological therapies:
  - Antidepressants: Section D3.1.1
  - Antipsychotics: Section D3.1.2
  - Anticonvulsants: Section D3.1.3
  - Benzodiazepines and z-drugs: Section D3.1.4
  - Lithium: Section D3.1.5
- Complementary therapies:
  - Omega-3 fatty acids: Section D3.2.1
  - St John's wort: Section D3.2.2
  - Gingko biloba: Section D3.2.3
- Physical therapies:
  - Electroconvulsive therapy: Section D3.3.1
  - Transcranial magnetic stimulation: Section D3.3.2.

The following sections of the Appendix to Part provide detailed information on how this evidence was selected and evaluated:

- Included studies: **Appendix D2**
- Data extraction: **Appendix D3**
- Assessment of evidence: **Appendix D4**
- Risk of bias assessment: **Appendix D5**.

D3.1 PHARMACOLOGICAL

D3.1.1 Antidepressants

The following section presents the Evidence Profile tables for the specific antidepressant classes and individual medications examined. Due to the large amount of evidence available for the assessment of antidepressants, only evidence from studies that adjusted for confounding and attempted to minimise the effect of confounding by indication have been included here. A summary of the characteristics of the individual included studies can be found in **Table AppD2-5** in **Appendix D2.1.1.2**. A detailed discussion of the evidence for each group or individual intervention type and outcome can be found in **Appendix D4.1.1**.

It should be noted that no certainty assessments based on assessment of individual studies were downgraded due to indirectness, because all included studies had been selected to minimise indirectness: they either limited the comparison to a population with depression/psychiatric disorder, or adjusted the analysis for depression/psychiatric disorder, thus attempting to minimise confounding by indication.

Table D 3-1 presents a summary of the results of the Evidence Review of antidepressants as well as the location of the detailed assessment of the certainty of evidence in the evidence profile tables. Due to the unsuitability of the identified SRs, relevant individual studies were identified and de novo meta-analyses were performed where appropriate. While evidence was identified for a number of groupings of antidepressants, only groupings with a pharmacological or chemical basis (i.e. groups based on similar modes of action such as receptor type [eg, SSRIs, SNRIs] or similar chemical structure [eg, TCAs]) were assessed in the EP tables. However, the evidence base for excluded groupings such as any antidepressants, non-SSRIs and co-exposures) is presented and discussed in **Appendix D4.1.1**.

The most evidence was available for SSRIs as a class, as demonstrated by the number of outcomes that were able to be assessed. The results suggest that antidepressants are, or may be, associated with adverse pregnancy and birth outcomes. There appeared to be no effect of SSRIs as a group, fluvoxamine, SNRIs and TCAs on malformations, although septal malformations may be associated with use of fluoxetine. For other individual SSRIs and antidepressants, the evidence on malformations was imprecise due to the low numbers available for the analysis. Where evidence was available on malformations, the certainty was considered very low. With the exception of no effect on IQ for SSRIs as a group, all available evidence for neurodevelopmental outcomes was uncertain, the main reason being that studies did not adequately account for confounding by indication due to depression severity both during pregnancy, and in the period following birth, which for some of the neurodevelopmental outcomes was up to 14 years. The evidence for the maternal harm postpartum haemorrhage was considered to be inadequate for SSRIs (although the finding was statistically significant), while for SNRIs the evidence was of very low certainty, and suggested SNRIs may be associated with postpartum haemorrhage.

A number of comparisons were made against other treatments; however, most of these were based on inadequate evidence. The exception was the risk of poor neonatal adaptation syndrome (PNAS) associated with the use of SSRIs compared with SNRIs, which showed that the risk may be greater for SSRIs.

Table D 3-1 Summary of results of the Evidence Review for antidepressants

Intervention	Increased/may be increased risk of harm Outcome Certainty of evidence	Appears to be no increased risk of harm Outcome Certainty of evidence	Decreased/may be decreased risk of harm Outcome Certainty of evidence	Uncertain Outcome ○○○○	Evidence profile table
SSRIs[9]	Miscarriage ●●○○ Preterm birth ●●○○ PNAS ●●○○ PNAS (SSRI vs SNRI) ●○○○ PPH ●○○○ Respiratory distress ●○○○ Convulsions ●●○○	Major malformation ●○○○ Cardiac malformation ●○○○ Neonatal mortality ●○○○ IQ ●○○○ Behavioural problems[10] ●○○○		Cardiac malformation (vs non-SSRI) Septal malformation ASD ADHD Other disorders[11] Depression Anxiety Postpartum haemorrhage	**Table D3-2**
Paroxetine	Miscarriage ●●○○			Major malformation Cardiac malformation Cardiac malformation (vs other ADs) ASD	**Table D3-3**
Fluoxetine	Septal malformation ●○○○	Miscarriage ●○○○		Major malformation Cardiac malformation ASD	**Table D3-4**
Sertraline		Miscarriage ●○○○		Major malformation Cardiac malformation ASD	**Table D3-5**
Citalopram		Miscarriage ●○○○		Major malformation Cardiac malformation ASD	**Table D3-6**
Escitalopram				Major malformation Cardiac malformation	**Table D3-7**
Fluvoxamine		Major malformation ●○○○ Cardiac malformation ●○○○ Miscarriage ●○○○		ASD	**Table D3-8**
SNRIs/ venlafaxine	Miscarriage ●●○○ Postpartum haemorrhage ●○○○	Major malformation ●○○○		Cardiac malformation ASD ADHD	**Table D3-9**
NaSSA/ mirtazapine				Major malformation (vs other ADS) Stillbirth (vs other ADS) Miscarriage (vs other ADS) Preterm birth (vs other ADS)	**Table D3-10**
TCAs	Miscarriage ●○○○	Major malformation ●○○○ Neonatal mortality ●○○○		Cardiac malformation ASD ADHD	**Table D3-11**
Bupropion				Cardiac malformation Cardiac malformation (vs other ADs) ADHD	**Table D3-12**

Abbreviations: AD, antidepressant; ADHD, attention deficit hyperactivity disorder; ASD, autism spectrum disorder; IQ, intelligence quotient; NaSSA, noradrenergic and specific serotonergic antidepressants; PNAS, poor neonatal adaptation syndrome; PPH, persistent pulmonary hypertension; SNRI, serotonin and noradrenalin reuptake inhibitor; SRI, selective reuptake inhibitor; SSRI, selective serotonin reuptake inhibitor, TCA, tricyclic antidepressant.

Note: All comparisons are against non-exposure, unless otherwise stated. Certainty of evidence gradings are as follows: ●●●● – high certainty; ●●●○ – moderate certainty; ●●○○ – low certainty; ●○○○ – very low certainty; ○○○○ – inadequate certainty.

[9] Also includes some data on SRIs (SSRIs and SNRIs)

[10] Includes internalising and externalising behaviours.

[11] Includes speech/language, scholastic and motor disorders.

Table D3-2 Evidence Profile table: SSRI harms

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias**[12]	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)** **Unexposed**	**Exposed**	**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects** **Risk with control**[13]	**Risk with intervention**[14]
Major malformations: see Section AppD4.1.1.3.2											
48,717 (3 – OBS)[15]	Serious(a)	None	None	None	None	●○○○ Very low	**Unexposed** NA	**SSRIs**[16] **(first trimester)** NA	RR 1.02 (0.91, 1.14)	28 per 1000[17]	29 per 1000 (25, 32)
Evidence Statement: *Maternal use of SSRIs during the first trimester of pregnancy does not appear to be associated with an increased risk of major malformation in the newborn (very low certainty evidence)*											
Cardiac malformations: see Section AppD4.1.1.4.2											
286,647 (6 – OBS)[18]	Serious(a)	None	None	None	None	●○○○ Very low	**Unexposed** NA	**SSRIs (first trimester)** NA	RR 1.04 (0.94, 1.15)	6 per 1000[19]	6 per 1000 (6, 7)
3,768 (1 – OBS)[20]	Serious(a)	NA	None	Serious(b)	None	○○○○ Inadequate	**Non-SSRIs** 992	**SSRIs (first trimester)** 2,776	RR 1.48 (0.58, 3.73)	Unknown	-
Evidence Statements: *Maternal use of SSRIs during the first trimester of pregnancy does not appear to be associated with an increased risk of cardiac malformation in the newborn (very low certainty evidence)* *Due to the inadequate certainty of the evidence, any additional risk of cardiac malformations in the newborn associated with maternal use of SSRIs during the first trimester of pregnancy, compared with maternal use of non-SSRIs during the same period, is uncertain.*											
Septal malformations: see Section AppD4.1.1.5.2											
16,831 (1 – OBS)[21]	Serious(a)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** 1,651	**Non-sertraline SSRIs** 236	RR 1.13 (0.81, 1.58)	3 per 1000[22]	3 per 1000 (2, 5)
Evidence Statements: *Due to the inadequate certainty of the evidence, any association between maternal use of non-sertraline SSRIs during the first trimester of pregnancy and septal malformation in the newborn is uncertain.*											

[12] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns are noted and may result in further downgrading of the quality of the evidence.
[13] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[14] Calculated by multiplying relative effect by control risk; it is not considered appropriate to calculate the risk with intervention where the quality of the evidence is inadequate.
[15] Based on a de novo meta-analysis of data from Ban 2014a, Bérard 2015 and Simon 2002.
[16] One study included non-sertraline studies only (Bérard 2015).
[17] Ban 2014a.
[18] Based on a de novo meta-analysis of data from Ban 2014a, Bérard 2015, Furu 2015, Huybrechts 2014a, Margulis 2013 and Petersen 2016.
[19] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013.
[20] Petersen 2016.
[21] Bérard 2015.
[22] The Bérard 2015 study used an insured population and as such the prevalence of septal malformations in this study (1.83%) is not likely to be representative of the general population with depression. To estimate the prevalence, it is assumed that 50% of cardiac malformations are septal, resulting in an estimate of 0.3%.

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[12]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[13]**	**Risk with intervention[14]**
Neonatal mortality:[23] see Section AppD4.1.1.6.2											
NR (1 – OBS)[24]	None	NA	None	Serious (b)	None	●○○○ Very low	**Unexposed** NA	**SSRIs** **(first trimester)** NA	RR 1.2 (0.6, 2.3)	5 per 1000[25]	6 per 1000 (3, 12)
Evidence Statement: *Maternal use of SSRIs during the first trimester of pregnancy does not appear to be associated with an increased risk of neonatal mortality (very low certainty evidence).*											
Miscarriage: see Section AppD4.1.1.7.2											
NR (2 – OBS)[26]	None	NA	None	None	None	●●○○ Low	**Unexposed** NA	**SSRIs** **(first trimester)** NA	**RR 1.34** **(1.16, 1.54)**	81 per 1000[27]	109 per 1000 (94, 125)
5,001 (1 – OBS)[28]	None	NA	None	None	None	●●○○ Low	**Unexposed** NA	**SSRIs** **(up to 20 weeks)** NA	**OR 1.61** **(1.28, 2.04)**	81 per 1000[29]	Not estimable
Evidence Statement: *Maternal use of SSRIs during the first 20 weeks of pregnancy is associated with an increased risk of miscarriage, from an absolute risk of 8% to 11% (low certainty evidence).*											
Pre-term birth: see Section AppD4.1.1.8.2											
< 37 weeks *1,787* (1 – OBS)[30]	None	NA	None	None	None	●●○○ Low	**Unexposed** 1,566	**SSRIs** **(late gestation)** 221	**RR 2.68** **(1.83, 3.93)**	60 per 1000[31]	161 per 1000 (110, 236)

[23] Includes stillbirth and neonatal death up to 28 days.
[24] Ban 2012.
[25] Ban 2012.
[26] Based on a de novo meta-analysis of data from Almeida 2016 and Ban 2012.
[27] Almeida 2016 and Ban 2012.
[28] Nakhai-Pour 2010; population likely overlaps with that of Almeida 2016.
[29] Almeida 2016 and Ban 2012.
[30] Grzeskowiak 2012.
[31] Malm 2015.

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[12]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)** *P value*	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[13]**	**Risk with intervention[14]**
< 37 weeks *1,622* (1 – OBS)[32]	None	NA	None	None	None	●●○○ Low	**Unexposed** 805	**SSRI (any time)** 817	RD 0.007 (-0.018, 0.034)	60 per 1000[33]	-
Evidence Statement: *Maternal use of SSRIs during late pregnancy is associated with an increased risk of preterm birth, from an absolute risk of 6% to 16% (low certainty evidence).*											
Small for gestational age: see Section AppD4.1.1.9.2											
1,787 (1 – OBS)[34]	None	NA	None	Serious(b)	None	●○○○ Very low	**Unexposed** 1,566	**SSRI (any time)** 221	OR 1.13 (0.65, 1.94)	Unknown	-
1,622 (1 – OBS)[35]	None	NA	None	None	None	●●○○ Low	**Unexposed** 805	**SSRI (any time)** 817	**RD 0.033 (0.007, 0.059)**	Unknown	-
Evidence Statement: *Maternal use of SSRIs at any time during pregnancy does not appear to be associated with an increased risk of the newborn being small for gestational age (low certainty evidence).*											
Poor neonatal adaptation syndrome: see Section AppD4.1.1.10.1 and AppD4.1.1.10.2											
312 (2 – OBS)[36]	Unknown[37]	Serious (c)	Serious (d)	None	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (any time)** NA	**RR 4.74 (2.14, 10.5)**	Unknown	-
247 *(1 – OBS)[38]*	Serious(e)	NA	None	None	None	●○○○ Very low	**SNRI** 24	**SSRI (third trimester)** 188	**OR 2.75 (1.13, 6.71)**	Unknown	-
Evidence Statement: *Maternal use of SSRIs at any time during pregnancy appears to be associated with an increased risk of poor neonatal adaptation syndrome in the newborn, but due to the inadequate certainty of the evidence this association is uncertain.* *Maternal use of SSRIs during the third trimester of pregnancy may be associated with an increased risk of poor neonatal adaptation syndrome compared with maternal use of SNRIs during the same period (increase in absolute risk not estimable) (very low certainty)*											

[32] Oberlander 2006.
[33] Malm 2015.
[34] Grzeskowiak 2012.
[35] Oberlander 2006.
[36] Based on an existing meta-analysis by Grigoriadis 2013b. No individual studies comparing exposure to non-exposure met the 'higher quality' criteria.
[37] Individual included studies not reported.
[38] Kieviet 2015.

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[12]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)** *P value*	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[13]**	**Risk with intervention[14]**
Persistent pulmonary hypertension: see Section AppD4.1.1.11.1 and AppD4.1.1.11.2											
NR (3 – OBS)[39]	None[40]	None	Serious(d)	None	None	●○○○ Very low	**Unexposed** NA	**SSRI (any time)** NA	**RR 2.41 (1.35, 3.95)**	3 per 1000[41]	7 per 1000 (4, 12)
NR (3 – OBS)[42]	None[43]	Very serious(f)	Serious(d)	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (early pregnancy)[44]** NA	RR 1.45 (0.84, 2.49)	3 per 1000[45]	-
NR (4 – OBS)[46]	None[47]	Serious(c)	Serious(d)	None	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (late pregnancy)[48]** NA	**RR 2.72 (1.63, 4.54)**	3 per 1000[49]	
786,446 (2 – OBS)[50]	None	Very serious(f)	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (late exposure)[51]** NA	RR 1.80 (0.65, 4.95)	3 per 1000[52]	-
Full-term deliveries only *621,399* (1 – OBS)[53]	None	NA	None	None	None	●●○○ Low	**Unexposed** 567,118	**SSRI (late exposure)[54]** 54,281	**RR 1.27 (1.00, 1.61)**	3 per 1000[55]	4 per 1000 (3, 5)

[39] Based on an existing meta-analysis by McDonagh 2014. Included because the individual studies comparing exposure to non-exposure did not adjust for a major potential confounder, caesarean birth.
[40] Based on the description provided by McDonagh 2014.
[41] Huybrechts 2015.
[42] Based on an existing meta-analysis by McDonagh 2014. Included because the individual studies comparing exposure to non-exposure did not adjust for a major potential confounder, caesarean birth.
[43] Based on the description provided by McDonagh 2014.
[44] Not defined.
[45] Huybrechts 2015.
[46] Based on an existing meta-analysis by McDonagh 2014. Included because the individual studies comparing exposure to non-exposure did not adjust for a major potential confounder, caesarean birth.
[47] Based on the description provided by McDonagh 2014.
[48] Mostly > 20 weeks.
[49] Huybrechts 2015.
[50] Based on a de novo meta-analysis of data from Huybrechts 2015 and Kieler 2012.
[51] Defined as 90 days before delivery for Huybrechts 2015 and from 140 days after start of pregnancy for Kieler 2012.
[52] Huybrechts 2015.
[53] Huybrechts 2015.
[54] Defined as 90 days before delivery.
[55] Huybrechts 2015.

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[12]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)** **Unexposed**	**Exposed**	**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects** **Risk with control[13]**	**Risk with intervention[14]**
Without cardiac malformation or lung hypoplasia *722,830* (1 – OBS)[56]	None	NA	None	Serious(b)	None	⨁◯◯◯ Very low	**Unexposed** 657,515	**SSRI (late exposure)[54]** 65,316	RR 1.08 (0.92, 1.27)	3 per 1000[57]	3 per 1000 (3, 4)
Full-term deliveries and excluding cardiac malformation or lung hypoplasia 621,399 (1 – OBS)[58]	None	NA	None	None	None	⨁⨁◯◯ Low	**Unexposed** 567,118	**SSRI (late exposure) [54]** 54,281	**RR 1.28 (1.01, 1.64)**	3 per 1000[59]	4 per 1000 (3, 5)
No meconium aspiration *NR* (1 – OBS)[60]	None	NA	None	None	None	⨁⨁◯◯ Low	**Unexposed** NA	**SSRI (early exposure)[61]** NA	**RR 1.3 (1.1, 1.7)**	-[62]	-
Evidence Statement: *Maternal use of SSRIs during late pregnancy may be associated with an increased risk of persistent pulmonary hypertension in the newborn, from an absolute risk of 0.3% to 0.4% (low certainty evidence)*											
Respiratory distress: see Section AppD4.1.1.12.2											
25,381 (1 – OBS)[63]	Serious(g)	NA	None	None	None	⨁◯◯◯ Very low	**Unexposed** 9,652	**SSRI (any time)** 15,729	**RR 1.40 (1.20, 1.62)**	32 per 1000[64]	45 per 1000 (38, 52)
1,622 (1 – OBS)[65]	None	NA	None	None	None	⨁⨁◯◯ Low	**Unexposed** NR	**SSRI (any time)** NR	**RD 0.044 (0.013, 0.077)**	32 per 1000[64]	33 per 1000 (32, 34)
Evidence Statement: *Maternal use of SSRIs at any time during pregnancy may be associated with an increased risk of respiratory distress in neonates, from an absolute risk of 3% to 5% (very low certainty evidence)*											

[56] Huybrechts 2015.
[57] Huybrechts 2015.
[58] Huybrechts 2015.
[59] Huybrechts 2015.
[60] Kieler 2012.
[61] Defined as from 140 days after start of pregnancy for Kieler 2012.
[62] Limited to population of women with previous psychiatric hospitalisation. No data available for baseline risk in this population.
[63] Malm 2015.
[64] Malm 2015.
[65] Oberlander 2006.

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[12]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)** *P value*	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[13]**	**Risk with intervention[14]**
Convulsions: see Section AppD4.1.1.14.2											
228,876 (1 – OBS)[66]	None	NA	None	Serious(b)	None	●○○○ Very low	**Unexposed** NA	**SSRI (third-trimester and 1 filled prescription)** NA	RR 1.4 (0.7, 2.8)	3 per 1000[67]	4 per 1000 (2, 8)
228,876 (1 – OBS)[68]	None	NA	None	None	None	●●○○ Low	**Unexposed** NA	**SSRI (third-trimester and 2 filled prescriptions)** NA	**RR 2.8 (1.4, 5.5)**	3 per 1000[69]	8 per 1000 (6, 17)
228,876 (1 – OBS)[70]	None	NA	None	None	None	●●○○ Low	**Unexposed** NA	**SSRI (third-trimester and 3+ filled prescriptions)** NA	**RR 4.9 (2.6, 9.5)**	3 per 1000[71]	15 per 1000 (8, 29)
Note: Hayes 2012 also show (without presenting risk estimates) that these same analyses conducted for first and second trimester exposure to SSRIs did not result in significant associations with convulsions.											
1,622 (1 – OBS)[72]	None	NA	None	None	None	●●○○ Low	**Unexposed** NA	**SSRI (any time)** NA	RD 0.00077 (-0.0010, 0.0036)	3 per 1000[73]	-
Evidence Statement: *Maternal use of SSRIs during the third trimester of pregnancy is associated with an increased risk of convulsions in the newborn, and the risk increases with increasing exposure, from an absolute risk of 0.3% up to 0.4% for one prescription filled, and up to 1.5% for three prescriptions filled (low certainty evidence).*											
Autism spectrum disorder: see Section AppD4.1.1.15.2											
29,737 (3 – OBS)[74]	Very serious(h)	None	None	None	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (any time)** NA	**RR 1.38 (1.02, 1.87)**	9 per 1000[75]	12 (9, 17)

[66] Hayes 2012.
[67] Hayes 2012.
[68] Hayes 2012.
[69] Hayes 2012.
[70] Hayes 2012.
[71] Hayes 2012.
[72] Oberlander 2006.
[73] Hayes 2012.
[74] Based on a de novo meta-analysis of data from Malm 2016, Harrington 2014 and Sørensen 2013.
[75] Sørensen 2013 and Malm 2016.

Certainty assessment							**Summary of findings**				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[12]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)** *P value*	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[13]**	**Risk with intervention[14]**
229 (1 – OBS)[76]	Very serious(h)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (first trimester)** NA	RR 1.70 (0.66, 4.38)	9 per 1000[75]	-
229 (1 – OBS)[77]	Very serious(h)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (second trimester)** NA	RR 1.12 (0.40, 3.14)	9 per 1000[75]	-
229 (1 – OBS)[78]	Very serious(h)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (third trimester)** NA	RR 1.43 (0.52, 3.93)	9 per 1000[75]	-
144,507 (1 – OBS)[79]	Very serious(h)	NA	None	None	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (second or third trimester)** NA	**RR 2.17 (1.20, 3.93)**	9 per 1000[75]	20 per 1000 (11, 35)
Childhood autism											
5,799 (1 – OBS)[80]	Very serious(h)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (any time)** NA	RR 1.0 (0.4, 2.6)	Unknown	-
Pervasive developmental disorder											
623 (2 – OBS)[81]	Very serious (i)	None	None	None	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (any time)** NA	**RR 1.05 (1.01, 1.09)**	Unknown	-
178 (1 – OBS)[82]	Very serious (i)	NA	None	None	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (any time)** NA	RR 1.01 (0.98, 1.05)	Unknown	-

[76] Harrington 2014.
[77] Harrington 2014.
[78] Harrington 2014.
[79] Boukhris 2016.
[80] Sørensen 2013.
[81] Based on a de nova meta-analysis of data from Johnson 2016 and El Marroun 2014.
[82] Johnson 2016.

Certainty assessment							**Summary of findings**				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[12]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)** *P value*	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[13]**	**Risk with intervention[14]**
Autistic traits – SRS											
445 (1 – OBS)[83]	Very serious (i)	NA	None	None	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (any time)** NA	**β 0.10 (0.02, 0.18)**	NA	-
Social cognition – SRS											
445 (1 – OBS)[84]	Very serious (i)	NA	None	None	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (any time)** NA	β 0.10 (-0.02, 0.22)	NA	-
Social communication – SRS											
445 (1 – OBS)[85]	Very serious (i)	NA	None	None	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (any time)** NA	**β 0.12 (0.03, 0.21)**	NA	-
Autistic mannerism – SRS											
445 (1 – OBS)[86]	Very serious (i)	NA	None	None	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (any time)** NA	**β 0.09 (0.01, 0.17)**	NA	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of SSRIs at any time during pregnancy and autism spectrum disorder in the child, is uncertain.*											
Attention deficit hyperactivity disorder: see Section AppD4.1.1.16.2											
23,709 (1 – OBS)[87]	Very serious(h)	None	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (any time)** NA	RR 0.98 (0.75, 1.28)	10 per 1000[88]	-
NR (1 – OBS)[89]	Very serious(h)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (first trimester)** NA	RR 1.62 (0.79, 3.32)	10 per 1000[88]	-

[83] El Marroun 2014.
[84] El Marroun 2014.
[85] El Marroun 2014.
[86] El Marroun 2014.
[87] Malm 2016.
[88] Based on Malm 2016.
[89] Figueroa 2010.

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[12]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[13]**	**Risk with intervention[14]**
NR (1 – OBS)[90]	Very serious(h)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (second trimester)** NA	RR 1.59 (0.58, 4.35)	10 per 1000[88]	-
NR (1 – OBS)[91]	Very serious(h)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (third trimester)** NA	RR 0.38 (0.14, 1.03)	10 per 1000[88]	-
NR (1 – OBS)[92]	Very serious(h)	NA	None	None	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (after pregnancy)** NA	**RR 2.04 (1.43, 2.91)**	10 per 1000[88]	20 per 1000 (14, 29)
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of SSRIs at any time during or after pregnancy and attention deficit hyperactivity disorder in the child, is uncertain.*											
Other disorders: see Section AppD4.1.1.17.2											
Speech/ language disorder *25,133* (1 – OBS)[93]	Very serious(j)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (any time)** NA	RR 1.20 (0.97, 1.49)	Unknown	-
Speech/ language disorder *NR* (1 – OBS)[94]	Very serious(j)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**SSRI – 1 purchase (any time)** NA	RR 0.86 (0.67, 1.10)	Unknown	-
Speech/ language disorder *NR* (1 – OBS)[95]	Very serious(j)	NA	None	None	None	○○○○ Inadequate	**Unexposed** NA	**SSRI – 2+ purchases (any time)** NA	**RR 1.37 (1.11, 1.70)**	Unknown	-
Speech/ language disorder *NR* (1 – OBS)[96]	Very serious(j)	NA	None	None	None	○○○○ Inadequate	**Unexposed** NA	**SSRI monotherapy only – 2+ purchases (any time)** NA	**RR 1.34 (1.07, 1.68)**	Unknown	-

[90] Figueroa 2010.
[91] Figueroa 2010.
[92] Figueroa 2010.
[93] Brown 2016.
[94] Brown 2016.
[95] Brown 2016.
[96] Brown 2016.

Certainty assessment							**Summary of findings**				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[12]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[13]**	**Risk with intervention[14]**
Speech/ language disorder *NR* (1 – OBS)[97]	Very serious(j)	NA	None	None	None	○○○○ Inadequate	**Unexposed/ <u>additional adjustment for suicidal behaviour</u>** NA	**SSRI – 2+ purchases (any time)** NA	**RR 1.34 (1.07, 1.68)**	Unknown	-
Scholastic disorder *25,133* (1 – OBS)[98]	Very serious(j)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (any time)** NA	RR 1.00 (0.63, 1.59)	Unknown	-
Scholastic disorder *NR* (1 – OBS)[99]	Very serious(j)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**SSRI – 1 purchase (any time)** NA	RR 0.86 (0.52, 1.42)	Unknown	-
Scholastic disorder *NR* (1 – OBS)[100]	Very serious(j)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**SSRI – 2+ purchases (any time)** NA	RR 1.15 (0.72, 1.84)	Unknown	-
Motor disorder *25,133* (1 – OBS)[101]	Very serious(j)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**SSRI (any time)** NA	RR 1.18 (0.81, 1.72)	Unknown	-
Motor disorder *NR* (1 – OBS)[102]	Very serious(j)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**SSRI – 1 purchase (any time)** NA	RR 0.86 (0.57, 1.30)	Unknown	-
Motor disorder *NR* (1 – OBS)[103]	Very serious(j)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**SSRI – 2+ purchases (any time)** NA	RR 1.33 (0.93, 1.91)	Unknown	-
<u>Evidence Statement:</u> *Due to the inadequate certainty of the evidence, any association between maternal use of SSRIs at any time during pregnancy and speech/language, scholastic or motor disorders in the child, is uncertain.*											

[97] Brown 2016.
[98] Brown 2016.
[99] Brown 2016.
[100] Brown 2016.
[101] Brown 2016.
[102] Brown 2016.
[103] Brown 2016.

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[12]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[13]**	**Risk with intervention[14]**
Intelligence Quotient: see Section AppD4.1.1.18.2											
Total IQ 90 (1 – OBS)[104]	None	NA	None	Unknown(b)	None	●○○○ Very Low	**Unexposed** NA	**SRIs[105]** **(any time)** NA	P ≥ 0.05	NA	-
Verbal IQ 90 (1 – OBS)[106]	None	NA	None	Unknown(b)	None	●○○○ Very Low	**Unexposed** NA	**SRIs** **(any time)** NA	P ≥ 0.05	NA	-
Performance IQ 90 (1 – OBS)[107]	None	NA	None	Unknown(b)	None	●○○○ Very Low	**Unexposed** NA	**SRIs** **(any time)** NA	P ≥ 0.05	NA	-
Evidence Statement: *Maternal use of SRIs at any time during pregnancy does not appear to be associated with a reduction in IQ in children aged 3 to 6 years (very low certainty evidence)*											
Behavioural problems: see Section AppD4.1.1.19.2											
Total problems (CBCL) 90 (1 – OBS)[108]	None	NA	None	Unknown(b)	None	●○○○ Very low	**Unexposed** NA	**SRIs[109]** **(any time)** NA	P ≥ 0.05	NA	-
Internalising behaviours											
90 (1 – OBS)[110]	None	NA	None	Unknown(b)	None	●○○○ Very low	**Unexposed** NA	**SRIs** **(any time)** NA	P ≥ 0.05	NA	-

[104] Nulman 2015
[105] Includes SSRIs and SNRIs.
[106] Nulman 2015
[107] Nulman 2015
[108] Nulman 2015
[109] Includes SSRIs and SNRIs.
[110] Nulman 2015

Certainty assessment							**Summary of findings**				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[12]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[13]**	**Risk with intervention[14]**
Externalising behaviours											
90 (1 – OBS)[111]	None	NA	None	Unknown(b)	None	●○○○ Very low	**Unexposed** NA	**SRIs** **(any time)** NA	P ≥ 0.05	NA	-
Evidence Statement: *Maternal use of SRIs at any time during pregnancy does not appear to be associated with an increased risk of behavioural problems in children aged 3 to 6 years (very low certainty evidence)*											
Depression: see Section AppD4.1.1.20.2											
NR (1 – OBS)[112]	Very serious(k)	NA	None	None	None	○○○○ Inadequate	**Unexposed** NA	**SSRI mono- or polytherapy** **(any time)** NA	**HR 1.84** **(1.14, 2.97)**	3 per 1000[113]	6 per 1000 (3, 9)
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of SSRIs at any time during pregnancy and depression in children aged up to 14 years, is uncertain.*											
Anxiety: see Section AppD4.1.1.21.2											
NR (1 – OBS)[114]	Very serious(k)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**SSRI mono- or polytherapy** **(any time)** NA	RR 1.30 (0.84, 2.01)	3 per 1000[115]	4 per 1000 (3, 6)
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of SSRI mono- or polytherapy during pregnancy and an increased risk of anxiety in children aged up to 14 years is uncertain.*											
Postpartum haemorrhage: see Section AppD4.1.1.15.1											
NR (4/10 – OBS)[116]	None	Very serious(f)	None	None	None	○○○○ Inadequate	**Unexposed** NR	**SRIs** **(any time)** NR	**OR 1.23** **(1.06, 1.44)**	Unknown	-
NR (3/7 – OBS)[117]	None	Very serious(f)	None	None	None	○○○○ Inadequate	**Unexposed** NR	**SRIs** **(recent users)** NR	**OR 1.30** **(1.06, 1.60)**	Unknown	-

[111] Nulman 2015
[112] Malm 2016.
[113] Malm 2016.
[114] Malm 2016.
[115] Malm 2016.
[116] Represents studies/estimates. Included studies not specified.
[117] Represents studies/estimates. Included studies not specified.

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[12]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[13]**	**Risk with intervention[14]**
NR (2/4 – OBS)[118]	None	Very serious(f)	None	None	None	○○○○ Inadequate	**Unexposed** NR	**SRIs** **(current users)** NR	OR 1.39 (0.96, 1.61)	Unknown	-
Evidence Statements: *There appears to be an association between maternal use of SRIs at any time during pregnancy and an increased risk of postpartum haemorrhage, but due to the inadequate certainty of the evidence, this association is uncertain.*											
Footnotes: a. Downgraded one level due to moderate risk of bias; potential for selection bias due to exclusion of planned abortions, miscarriages and still born from the analysis. b. Downgraded one level due to imprecision; 95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25, no measure of precision available, or no events. c. Downgraded one level due to moderate heterogeneity (I^2 = 25% to 59%). d. Downgraded one level due to indirectness caused by use of non-depressed control group. e. Downgraded one level due to moderate risk of bias; use of a non-validated outcome assessment tool. f. Downgraded two levels due to substantial heterogeneity (I^2 > 60%). g. Downgraded one level due to moderate risk of bias; potential for selection bias between exposed and psychiatric disorder/unexposed populations. h. Downgraded two levels due to high risk of bias; lack of adjustment for potential confounding by maternal disease severity in the antenatal and postnatal period. i. Downgraded two levels due to high risk of bias; self-rated outcomes that were inconsistent and lack of/inadequate adjustment for maternal disease severity in the postnatal period. j. Downgraded two levels due to high risk of bias; potential for selection bias between exposed and psychiatric disorder/unexposed populations and lack of adjustment for potential confounding by maternal disease severity in the postnatal period. k. Downgraded two levels due to high risk of bias; potential for selection bias due to age unbalanced populations and lack of adjustment for potential confounding by maternal disease severity in the antenatal or postnatal period.											

Abbreviations: AD, antidepressant; CI, confidence interval; NA, not available; NE, not estimable; NR, not reported; OBS, observational studies; OR, odds ratio; RD, risk difference; RR, relative risk; SRS, social responsiveness scale; SSRI, selective serotonin reuptake inhibitor; TCA, tricyclic antidepressant.

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

[118] Represents studies/estimates. Included studies not specified.

Table D3-3 Evidence Profile table: paroxetine harms

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[119]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[120]**	**Risk with intervention[121]**
Major malformations: see Section AppD4.1.1.3.2											
27,362 (2–OBS)[122]	Serious(a)	None	None	Serious (b)	None	○○○○ Inadequate	**Unexposed** NA	**Paroxetine (first trimester)** NA	RR 1.09 (0.82, 1.45)	28 per 1000[123]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of paroxetine during the first trimester of pregnancy and major malformation in the newborn, is uncertain.*											
Cardiac malformations: see Section AppD4.1.1.4.2											
214,345 (2 – OBS)[124]	Serious (a)	Serious (c)	None	Serious (b)	None	○○○○ Inadequate	**Unexposed** NA	**Paroxetine** NA	RR 1.20 (0.69, 2.09)	6 per 1000[125]	-
5,013 (1 – OBS)[126]	Serious(a)	NA	None	Serious(b)	None	○○○○ Inadequate	**Other AD monotherapy (first trimester)** NA	**Paroxetine monotherapy (first trimester)** NA	RR 1.46 (0.74, 2.88)	Unknown	-
5,956 (1 – OBS)[127]	Serious(a)	NA	None	Serious(b)	None	○○○○ Inadequate	**Other AD mono- or polytherapy** NA	**Paroxetine mono- or polytherapy (first trimester)** NA	RR 1.68 (0.95, 2.97)	Unknown	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of paroxetine during the first trimester of pregnancy and cardiac malformation in the newborn, is uncertain.* *Due to the inadequate certainty of the evidence, any additional risk of cardiac malformation in the newborn that may be associated with maternal use of paroxetine in the first trimester, compared with maternal use of other antidepressant mono- or polytherapy during the same period, is uncertain.*											

[119] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns are noted and may result in further downgrading of the quality of the evidence.
[120] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[121] Calculated by multiplying relative effect by control risk.
[122] Based on a de novo meta-analysis of data from Ban 2014a and Ramos 2008.
[123] Ban 2014a.
[124] Based on a de novo meta-analysis of data from Ban 2014a and Huybrechts 2014a.
[125] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013.
[126] Cole 2007b.
[127] Cole 2007b.

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[119]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[120]**	**Risk with intervention[121]**
Miscarriage: see Section AppD4.1.1.7.2											
4,924 (1 – OBS)[128]	None	NA	None	None	None	●●○○ Low	**Unexposed** NA	**Paroxetine (up to 20 weeks)** NA	**OR 1.75 (1.31, 2.34)**	81 per 1000[129]	NE
Evidence Statement: *Maternal use of paroxetine up to the first 20 weeks of pregnancy is associated with an increased risk of miscarriage (increase in absolute risk not estimable) (low certainty evidence)*											
Autism spectrum disorder: see Section AppD4.1.1.15.2											
143,460 (1 – OBS)[130]	Very serious(d)	NA	None	None	None	○○○○ Inadequate	**Unexposed** 142,716	**Paroxetine (second or third trimester)** 744	**RR 1.99 (1.00, 3.96)**	9 per 1000[131]	18 per 1000 (9, 36)
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of paroxetine during the second or third trimester of pregnancy and autism spectrum disorder is uncertain.*											

Footnotes:
a. Downgraded one level due to moderate risk of bias; potential for selection bias due to exclusion of planned abortions, miscarriages and still born from the analysis.
b. Downgraded one level due to imprecision; 95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25, no measure of precision available, or no events.
c. Downgraded two levels due to substantial heterogeneity (I^2 > 60%).
d. Downgraded two levels due to high risk of bias; lack of adjustment for confounding for maternal disease severity in the antenatal and postnatal period.

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.
Abbreviations: AD, antidepressant; CI, confidence interval; NA, not available; NE, not estimable; NR, not reported; OBS, observational studies; OR, odds ratio; RR, relative risk.

[128] Nakhai-Pour 2010; population likely overlaps with that of Almeida 2016.
[129] Almeida 2016 and Ban 2012.
[130] Bérard 2016.
[131] Based on the pooled prevalence from Sørensen 2013 and Malm 2016.

Table D3-4 Evidence Profile table: fluoxetine harms

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[132]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[133]**	**Risk with intervention[134]**
Major malformations: see Section AppD4.1.1.3.2											
27,022 (1 – OBS)[135]	Serious(a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed** NA	**Fluoxetine (first trimester)** NA	RR 0.85 (0.66, 1.09)	28 per 1000[136]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of fluoxetine during the first trimester of pregnancy and major malformation in the newborn, is uncertain.*											
Cardiac malformations: see Section AppD4.1.1.4.2											
216,249 (2 – OBS)[137]	Serious(a)	Serious(c)	None	Serious (b)	None	○○○○ Inadequate	**Unexposed** NA	**Fluoxetine (first trimester)** NA	RR 1.01 (0.72, 1.42)	6 per 1000[138]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of fluoxetine during the first trimester of pregnancy and cardiac malformation in the newborn, is uncertain.*											
Miscarriage: see Section AppD4.1.1.7.2											
4,862 (1 – OBS)[139]	None	NA	None	Serious (b)	None	●○○○ Very low	**Unexposed** NA	**Fluoxetine (up to 20 weeks)** NA	OR 1.44 (0.86, 2.43)	81 per 1000[140]	Not estimable
Evidence Statement: *Maternal use of fluoxetine up to the first 20 weeks of pregnancy does not appear to be associated with an increased risk of miscarriage (very low certainty).*											

[132] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns are noted and may result in further downgrading of the quality of the evidence.
[133] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[134] Calculated by multiplying relative effect by control risk.
[135] Ban 2014a.
[136] Ban 2014a.
[137] Based on a de novo meta-analysis of data from Ban 2014a and Huybrechts 2014a.
[138] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013.
[139] Nakhai-Pour 2010; population likely overlaps with that of Almeida 2016.
[140] Almeida 2016 and Ban 2012.

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[132]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[133]**	**Risk with intervention[134]**
Autism spectrum disorder: see Section AppD4.1.1.15.2											
142,887 (1 – OBS)[141]	Very serious(d)	NA	None	None	None	○○○○ Inadequate	**Unexposed** NA	**Fluoxetine (second or third trimester)** NA	**RR 4.99 (1.45, 17.2)**	9 per 1000[142]	45 per 1000 (13, 155)
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of fluoxetine during the second or third trimester of pregnancy and autism spectrum disorder is uncertain.*											

Footnotes:
a. Downgraded one level due to moderate risk of bias; potential for selection bias due to exclusion of planned abortions, miscarriages and still born from the analysis.
b. Downgraded one level due to imprecision; 95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25, no measure of precision available, or no events.
c. Downgraded one level due to moderate heterogeneity (I^2 = 25% to 59%).
d. Downgraded two levels due to high risk of bias; lack of adjustment for confounding for maternal disease severity in the antenatal and postnatal period.

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.
Abbreviations: CI, confidence interval; NA, not available; OBS, observational studies; OR, odds ratio; RR, relative risk.

[141] Bérard 2016.
[142] Based on the pooled prevalence from Sørensen 2013 and Malm 2016.

Table D3-5 Evidence Profile table: sertraline harms

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias**[143]	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control**[144]	**Risk with intervention**[145]
Major malformations: See AppD4.1.1.3.2											
39,824 (2 – OBS)[146]	Serious (a)	None	None	Serious (b)	None	○○○○ Inadequate	**Unexposed** NA	**Sertraline (first trimester)** NA	RR 1.13 (0.88, 1.45)	28 per 1000[147]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of sertraline during the first trimester of pregnancy and major malformation in the newborn, is uncertain.*											
Cardiac malformations: see Section AppD4.1.1.4.2											
231,444 (3 – OBS)[148]	Serious (a)	None	None	Serious (b)	None	○○○○ Inadequate	**Unexposed** NA	**Sertraline (first trimester)** NA	RR 1.12 (0.92, 1.36)	6 per 1000[149]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of sertraline during the first trimester of pregnancy and cardiac malformation in the newborn, is uncertain.*											
Septal malformations: see Section AppD4.1.1.5.2											
15,234 (1 – OBS)[150]	Serious (a)	None	None	None	None	●○○○ Very low	**Unexposed** NA	**Sertraline** NA	**RR 1.34 (1.02, 1.76)**	3 per 1000[151]	4 per 1000 (3, 5)
Evidence Statement: *Maternal use of sertraline during the first trimester of pregnancy may be associated with an increased risk of septal malformation in the newborn, from an absolute risk of 0.3% to 0.4% (very low certainty evidence)*											

[143] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns are noted and may result in further downgrading of the quality of the evidence.
[144] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[145] Calculated by multiplying relative effect by control risk.
[146] Based on a de novo meta-analysis of data from Ban 2014a and Bérard 2015.
[147] Ban 2014a.
[148] Based on a de novo meta-analysis of data from Ban 2014a, Bérard 2015 and Huybrechts 2014a.
[149] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013.
[150] Bérard 2015.
[151] The Bérard 2015 study used an insured population and as such the prevalence of septal malformations in this study (1.83%) is not likely to be representative of the general population with depression. To estimate the prevalence, it is assumed that 50% of cardiac malformations are septal, resulting in an estimate of 0.3%.

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[143]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[144]**	**Risk with intervention[145]**
Miscarriage: see Section AppD4.1.1.7.2											
4,868 (1 – OBS)[152]	None	NA	None	Serious (b)	None	●○○○ Very low	**Unexposed** NA	**Sertraline (up to 20 weeks)** NA	OR 1.33 (0.85, 2.08)	81 per 1000[153]	Not estimable
Evidence Statement: *Maternal use of sertraline during the first 20 weeks of pregnancy does not appear to be associated with an increased risk of miscarriage (very low certainty evidence)*											
Autism spectrum disorder: see Section AppD4.1.1.15.2											
143,008 (1 – OBS)[154]	Very serious(c)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** 142,716	**Sertraline (second or third trimester)** 292	RR 0.45 (0.05, 4.05)	9 per 1000[155]	4 per 1000 (<1, 36)
Evidence Statement: Due to the inadequate certainty of the evidence, any association between maternal use of sertraline during the second or third trimester of pregnancy and autism spectrum disorder in the child is uncertain.											
Footnotes: a. Downgraded one level due to moderate risk of bias; potential for selection bias due to exclusion of planned abortions, miscarriages and stillborn from the analysis. b. Downgraded one level due to imprecision; 95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25, no measure of precision available, or no events. c. Downgraded two levels due to high risk of bias; lack of adjustment for potential confounding by maternal disease severity in the antenatal and postnatal period.											

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: AD, antidepressant; CI, confidence interval; NA, not available; NE, not estimable; NR, not reported; OBS, observational studies; OR, odds ratio; RR, relative risk; SSRI, selective serotonin reuptake inhibitor; TCA, tricyclic antidepressant.

[152] Nakhai-Pour 2010; population likely overlaps with that of Almeida 2016.
[153] Almeida 2016 and Ban 2012.
[154] Bérard 2016.
[155] Based on the pooled prevalence from Sørensen 2013 and Malm 2016.

Table D3-6 Evidence Profile table: citalopram harms

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias**[156]	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control**[157]	**Risk with intervention**[158]
Major malformations: see Section AppD4.1.1.3.2											
25,779 (1 – OBS)[159]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed** NA	**Citalopram (first trimester)** NA	RR 0.97 (0.71, 1.31)	28 per 1000[160]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of citalopram during the first trimester of pregnancy and major malformation in the newborn, is uncertain.*											
Cardiac malformations: see Section AppD4.1.1.4.2											
25,779 (1 – OBS)[161]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed** NA	**Citalopram (first trimester)** NA	RR 1.02 (0.61, 1.71)	6 per 1000[162]	-
Evidence Statement: *Due to the inadequate certainty of the evidence any association between maternal use of citalopram during the first trimester of pregnancy and cardiac malformation in the newborn, is uncertain.*											
Miscarriage: see Section AppD4.1.1.7.2											
4,859 (1 – OBS)[163]	None	NA	None	Serious (b)	None	●○○○ Very low	**Unexposed** NA	**Citalopram (up to 20 weeks)** NA	OR 1.55 (0.89, 2.69)	81 per 1000[164]	Not estimable
Evidence Statement: *Maternal use of citalopram during the first 20 weeks of pregnancy does not appear to be associated with an increased risk of miscarriage (very low certainty evidence)*											

[156] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns are noted and may result in further downgrading of the quality of the evidence.
[157] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[158] Calculated by multiplying relative effect by control risk.
[159] Ban 2014a.
[160] Ban 2014a.
[161] Ban 2014a.
[162] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013.
[163] Nakhai-Pour 2010; population likely overlaps with that of Almeida 2016.
[164] Almeida 2016 and Ban 2012.

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[156]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[157]**	**Risk with intervention[158]**
Autism spectrum disorder: see Section AppD4.1.1.15.2											
143,137 (1 – OBS)[165]	Very serious(c)	NA	None	None	None	○○○○ Inadequate	**Unexposed** NA	**Citalopram (second or third trimester)** NA	**RR 2.23 (1.01, 4.92)**	9 per 1000[166]	20 per 1000 (9, 44)
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of citalopram during the second or third trimester of pregnancy and autism spectrum disorder in the child is uncertain.*											
Footnotes: a. Downgraded one level due to moderate risk of bias; potential for selection bias due to exclusion of planned abortions, miscarriages and still born from the analysis. b. Downgraded one level due to imprecision; 95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25, no measure of precision available, or no events. c. Downgraded two levels due to high risk of bias; lack of adjustment for potential confounding by maternal disease severity in the antenatal and postnatal period.										

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: CI, confidence interval; NA, not available; OBS, observational studies; OR, odds ratio; RR, relative risk.

[165] Bérard 2016.

[166] Based on the pooled prevalence from Sørensen 2013 and Malm 2016.

Table D3-7 Evidence Profile table: escitalopram harms

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[167]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[168]**	**Risk with intervention[169]**
Major malformations: see Section AppD4.1.1.3.2											
24,166 (1 – OBS)[170]	Serious(a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed** NA	**Escitalopram (first trimester)** NA	RR 0.77 (0.36, 1.66)	28 per 1000[171]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of escitalopram during the first trimester of pregnancy and major malformation in the newborn, is uncertain.*											
Cardiac malformations: see Section AppD4.1.1.4.2											
24,166 (1 – OBS)[172]	Serious(a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed** NA	**Escitalopram (first trimester)** NA	RR 1.09 (0.34, 3.50)	6 per 1000[173]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of escitalopram during the first trimester of pregnancy and cardiac malformation in the newborn, is uncertain.*											
Footnotes: a. Downgraded one level due to moderate risk of bias; potential for selection bias due to exclusion of planned abortions, miscarriages and still born from the analysis. b. Downgraded one level due to imprecision; 95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25, no measure of precision available, or no events.											

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: CI, confidence interval; NA, not available; OBS, observational studies; RR, relative risk.

[167] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns are noted and may result in further downgrading of the quality of the evidence.

[168] Based on average risk from unexposed, depressed control groups of population-based cohort studies providing data for that outcome.

[169] Calculated by multiplying relative effect by control risk.

[170] Ban 2014a.

[171] Ban 2014a.

[172] Ban 2014a.

[173] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013.

Table D3-8 Evidence Profile table: fluvoxamine harms

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* (No. studies)	Additional risk of bias[174]	Inconsistency	Indirectness	Imprecision	Publication bias	Overall certainty of evidence	Population (N)		Risk estimate (95% CI)	Anticipated absolute effects	
							Unexposed	Exposed		Risk with control[175]	Risk with intervention[176]
Major malformations: see Section AppD4.1.1.3.2											
107,439 (1 – OBS)[177]	Serious(a)	NA	None	None	None	●○○○ Very low	**Unexposed** NA	**Fluvoxamine (first trimester)** NA	RD -0.0152 (-0.0402, 0.0098)	28 per 1000[178]	28 per 1000 (27, 28)
Evidence Statement: *Maternal use of fluvoxamine during the first trimester of pregnancy does not appear to be associated with an increased risk of major malformation in the newborn (very low certainty evidence)*											
Cardiac malformations: see Section AppD4.1.1.4.2											
107,439 (1 – OBS)[179]	Serious(a)	NA	None	None	None	●○○○ Very low	**Unexposed** NA	**Fluvoxamine (first trimester)** NA	RD -0.0055 (-0.0145, 0.0036)	6 per 1000[180]	6 per 1000 (6, 6)
Evidence Statement: *Maternal use of fluvoxamine during the first trimester of pregnancy does not appear to be associated with an increased risk of cardiac malformation in the newborn (very low certainty evidence)*											
Miscarriage: see Section AppD4.1.1.7.2											
4,845 (1 – OBS)[181]	None	NA	None	Serious (b)	None	●○○○ Very low	**Unexposed** NA	**Fluvoxamine (up to 20 weeks)** NA	OR 2.19 (0.79, 6.08)	81 per 1000[182]	Not estimable
Evidence Statement: *Maternal use of fluvoxamine during the first 20 weeks of pregnancy does not appear to be associated with an increased risk of miscarriage (very low certainty evidence)*											

[174] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns are noted and may result in further downgrading of the quality of the evidence.
[175] Based on average risk from unexposed, depressed control groups of population-based cohort studies providing data for that outcome.
[176] Calculated by multiplying relative effect by control risk.
[177] Oberlander 2008a.
[178] Ban 2014a.
[179] Oberlander 2008a.
[180] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013.
[181] Nakhai-Pour 2010; population likely overlaps with that of Almeida 2016.
[182] Almeida 2016 and Ban 2012.

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[174]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[175]**	**Risk with intervention[176]**
Autism spectrum disorder: see Section AppD4.1.1.15.2											
142,751 (1 – OBS)[183]	Very serious(d)	NA	None	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Fluvoxamine (second or third trimester)** NA	RR 7.30 (0.30, 178)	10 per 1000[184]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of fluvoxamine during the second or third trimester of pregnancy and autism spectrum disorder in the child is uncertain.*											
Footnotes: a. Downgraded one level due to moderate risk of bias; potential for selection bias due to exclusion of planned abortions, miscarriages and still born from the analysis. b. Downgraded one level due to imprecision; 95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25, no measure of precision available, or no events. c. Downgraded two levels due to high risk of bias; lack of adjustment for potential confounding by maternal disease severity in the antenatal and postnatal period.										

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: AD, antidepressant; CI, confidence interval; NA, not available; NE, not estimable; NR, not reported; OBS, observational studies; OR, odds ratio; RR, relative risk; SSRI, selective serotonin reuptake inhibitor; TCA, tricyclic antidepressant.

[183] Bérard 2016.

[184] Based on the pooled prevalence from Sørensen 2013 and Malm 2016.

Table D3-9 Evidence Profile table: SNRI/venlafaxine harms

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[185]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[186]**	**Risk with intervention[187]**
Major malformations: see Section AppD4.1.1.3.2											
107,570 (1 – OBS)[188]	Serious(a)	NA	None	None	None	●○○○ Very low	**Unexposed** NA	**Venlafaxine (first trimester)** NA	RD -0.0118 (-0.0320, 0.0084)	28 per 1000[189]	28 per 1000 (27, 28)
Evidence Statement: *Maternal use of venlafaxine during the first trimester of pregnancy does not appear to be associated with an increased risk of major malformation in the newborn (very low certainty evidence)*											
Cardiac malformations: see Section AppD4.1.1.4.2											
186,574 (1 – OBS)[190]	Serious(a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed** NA	**SNRIs (first trimester)** NA	RR 1.20 (0.91, 1.57)	6 per 1000[191]	-
107,570 (1 – OBS)[192]	Serious(a)	NA	None	None	None	●○○○ Very low	**Unexposed** NA	**Venlafaxine (first trimester)** NA	RD 0.0001 (-0.0077, 0.0079)	6 per 1000[193]	6 per 1000 (6, 6)
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of SNRIs during the first trimester of pregnancy and cardiac malformation in the newborn, is uncertain.*											
Miscarriage: see Section AppD4.1.1.7.2											
9,014 (1 – OBS)[194]	None	NA	None	None	None	●●○○ Low	**Unexposed** NA	**SNRIs (first trimester)** NA	**RR 1.7 (1.2, 2.6)**	81 per 1000[195]	138 per 1000 (97, 211)

[185] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns are noted and may result in further downgrading of the quality of the evidence.
[186] Based on average risk from unexposed, depressed control groups of population-based cohort studies providing data for that outcome.
[187] Calculated by multiplying relative effect by control risk.
[188] Oberlander 2008a.
[189] Ban 2014a.
[190] Huybrechts 2014a.
[191] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013.
[192] Oberlander 2008a.
[193] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013.
[194] Almeida 2016.
[195] Almeida 2016 and Ban 2012.

Certainty assessment							**Summary of findings**				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[185]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[186]**	**Risk with intervention[187]**
4,873 (1 – OBS)[196]	None	NA	None	None	None	●●○○ Low	**Unexposed** NA	**SNRIs (up to 20 weeks)** NA	**OR 2.11 (1.34, 3.30)**	81 per 1000[197]	Not estimable
Evidence Statement: *Maternal use of SNRIs during the first 20 weeks of pregnancy is associated with an increased risk of miscarriage, from an absolute risk of 8% to 14% (low certainty evidence)*											
Autism spectrum disorder: see Section AppD4.1.1.15.2											
143,371 (1 – OBS)[198]	Very serious(c)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**SNRIs (second or third trimester)** NA	RR 1.04 (0.20, 5.46)	9 per 1000[199]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of SNRIs during the second or third trimester of pregnancy and autism spectrum disorder in the child is uncertain.*											
ADHD: see Section AppD4.1.1.16.2											
863,533 (1 – OBS)[200]	Very serious(c)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**SNRIs (any time)** NA	RR 1.0 (0.4, 2.5)	10 per 1000[201]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of SNRIs at any time during pregnancy and attention deficit hyperactivity disorder in the child is uncertain.*											

Footnotes:
a. Downgraded one level due to moderate risk of bias; potential for selection bias due to exclusion of planned abortions, miscarriages and still born from the analysis.
b. Downgraded one level due to imprecision; 95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25, no measure of precision available, or no events.
c. Downgraded two levels due to high risk of bias; lack of adjustment for potential confounding by maternal disease severity in the antenatal and postnatal period.

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: AD, antidepressant; CI, confidence interval; NA, not available; NE, not estimable; NR, not reported; OBS, observational studies; OR, odds ratio; RR, relative risk; SSRI, selective serotonin reuptake inhibitor; TCA, tricyclic antidepressant.

[196] Nakhai-Pour 2010; population likely overlaps with that of Almeida 2016.
[197] Almeida 2016 and Ban 2012.
[198] Boukhris 2016.
[199] Based on the pooled prevalence from Sørensen 2013 and Malm 2016.
[200] Laugesen 2013.
[201] Based on Malm 2016.

Table D3-10 Evidence Profile table: NaSSA/mirtazapine harms

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[202]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[203]**	**Risk with intervention[204]**
Major malformations: see Section AppD4.1.1.3.2											
208 (1 – OBS)[205]	Serious(a)	NA	None	Unknown(b)	None	○○○○ Inadequate	**Other ADs** 104	**Mirtazapine (any time)** 104	*P=0.50*	Unknown	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any additional risk of major malformation in the newborn that may be associated with maternal use of mirtazapine at any time during pregnancy, compared with maternal use of other antidepressants at any time during pregnancy, is uncertain.*											
Stillbirth: see Section AppD4.1.1.6.2											
208 (1 – OBS)[206]	Serious(a)	NA	None	Unknown(b)	None	○○○○ Inadequate	**Other ADs** NA	**Mirtazapine (any time)** NA	*P=0.50*	Unknown	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any additional risk of stillbirth that may be associated with maternal use of mirtazapine at any time during pregnancy, compared with maternal use of other antidepressants at any time during pregnancy, is uncertain.*											
Miscarriage: see Section AppD4.1.1.7.2											
208 (1 – OBS)[207]	Serious(a)	NA	None	Unknown(b)	None	○○○○ Inadequate	**Other ADs** NA	**Mirtazapine (any time)** NA	*P=0.86*	Unknown	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any additional risk of miscarriage that may be associated with maternal use of mirtazapine at any time during pregnancy, compared with maternal use of other antidepressants at any time during pregnancy, is uncertain.*											

[202] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns are noted and may result in further downgrading of the quality of the evidence.
[203] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[204] Calculated by multiplying relative effect by control risk.
[205] Djulus 2006.
[206] Djulus 2006.
[207] Djulus 2006.

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[202]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[203]**	**Risk with intervention[204]**
Preterm birth: see Section AppD4.1.1.8.2											
208 (1 – OBS)[208]	Serious(a)	NA	None	Unknown(b)	None	○○○○ Inadequate	**Other ADs** NA	**Mirtazapine (any time)** NA	*P=0.61*	Unknown	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any additional risk of preterm birth in the newborn that may be associated with maternal use of mirtazapine at any time during pregnancy, compared with maternal use of other antidepressants at any time during pregnancy, is uncertain.*											
Footnotes: a. Downgraded one level due to moderate risk of bias; no information on extent of follow-up. b. Downgraded one level due to imprecision; 95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25, no measure of precision available, or no events.											

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: AD, antidepressant; CI, confidence interval; NA, not available; NE, not estimable; NR, not reported; OBS, observational studies; OR, odds ratio; RR, relative risk; SSRI, selective serotonin reuptake inhibitor; TCA, tricyclic antidepressant.

[208] Djulus 2006.

Table D3-11 Evidence Profile table: TCA harms

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias**[209]	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control**[210]	**Risk with intervention**[211]
Major malformations: see Section AppD4.1.1.3.2											
29,008 (3 – OBS)[212]	Serious(a)	None	None	None	None	●○○○ Very low	**Unexposed** NA	**TCAs (first trimester)** NA	RR 0.99 (0.78, 1.25)	28 per 1000[213]	28 per 1000 (22, 35)
Evidence Statement: *Maternal use of TCAs during the first trimester of pregnancy does not appear to be associated with major malformation in the newborn (very low certainty evidence)*											
Cardiac malformations: see Section AppD4.1.1.4.2											
210,555 (3 – OBS)[214]	Serious(a)	None	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**TCAs (any time)** NA	RR 0.81 (0.59, 1.10)	6 per 1000[215]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of TCAs during the first trimester of pregnancy and cardiac malformation in the newborn, is uncertain.*											
Neonatal mortality:[216] **see Section AppD4.1.1.6.2**											
NR (1 – OBS)[217]	None	NA	None	Serious(b)	None	●○○○ Very low	**Unexposed** NA	**TCAs (first trimester)** NA	RR 1.2 (0.5, 2.7)	5 per 1000[218]	6 per 1000 (3, 14)
Evidence Statement: *Maternal use of TCAs during the first trimester of pregnancy does not appear to be associated with neonatal mortality (very low certainty evidence)*											

[209] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns are noted and may result in further downgrading of the quality of the evidence.
[210] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[211] Calculated by multiplying relative effect by control risk.
[212] Based on a de novo meta-analysis of data from Ban 2014a, Ramos 2008 and Simon 2002.
[213] Ban 2014a.
[214] Based on a de novo meta-analysis of data from Ban 2014a, Huybrechts 2014a and Simon 2002.
[215] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013.
[216] Includes stillbirth and neonatal death up to 28 days.
[217] Ban 2012.
[218] Ban 2012.

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[209]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[210]**	**Risk with intervention[211]**
Miscarriage: see Section AppD4.1.1.7.2											
NR (2 – OBS)[219]	None	None	None	None	None	●●○○ Low	**Unexposed** NA	**TCAs (first trimester)** NA	**RR 1.32 (1.13, 1.55)**	81 per 1000[220]	**107 per 1000 (92, 126)**
4,876 (1 – OBS)[221]	None	NA	None	Serious(b)	None	●○○○ Very low	**Unexposed** NA	**TCAs (up to 20 weeks)** NA	OR 1.27 (0.85, 1.91)	81 per 1000[222]	Not estimable
Evidence Statement: *Maternal use of TCAs during the first trimester of pregnancy may be associated with an increased risk of miscarriage, from an absolute risk of 8% to 11% (low certainty evidence)*											
Autism spectrum disorder: see Section AppD4.1.1.15.2											
18,524 (1 – OBS)[223]	Very serious(c)	NA	None	None	None	○○○○ Inadequate	**Unexposed** NA	**TCAs (any time)** NA	**RR 2.69 (1.04, 6.96)**	9 per 1000[224]	24 per 1000 (9, 63)
143,153 (1 – OBS)[225]	Very serious(c)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed** NA	**TCAs (second or third trimester)** NA	RR 1.03 (0.23, 4.61)	9 per 1000[224]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of TCAs at any time during pregnancy and autism spectrum disorder in the child is uncertain.*											

[219] Based on a de novo meta-analysis of data from Almeida 2016 and Ban 2012.
[220] Almeida 2016 and Ban 2012.
[221] Nakhai-Pour 2010; population likely overlaps with that of Almeida 2016.
[222] Almeida 2016 and Ban 2012.
[223] Rai 2013.
[224] Based on the pooled prevalence from Sørensen 2013 and Malm 2016.
[225] Boukhris 2016.

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[209]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[210]**	**Risk with intervention[211]**
ADHD: see Section AppD4.1.1.16.2											
863,533 (1 – OBS)[226]	Very serious(d)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**TCAs (any time)** NA	RR 1.1 (0.6, 2.0)	10 per 1000[227]	-
Evidence Statement: Due to the inadequate certainty of the evidence, any association between maternal use of TCAs at any time during pregnancy and attention deficit hyperactivity disorder in the child is uncertain.											

Footnotes:

a. Downgraded one level due to moderate risk of bias; potential for selection bias due to exclusion of planned abortions, miscarriages and stillborns from the analysis.

b. Downgraded one level due to imprecision; 95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25, no measure of precision available, or no events.

c. Downgraded two levels due to high risk of bias; lack of adjustment for potential confounding by maternal disease severity in the antenatal and postnatal period.

d. Downgraded two levels due to high risk of bias; inadequate adjustment for potential confounding by indication and lack of adjustment for potential confounding by maternal disease severity in the antenatal and postnatal period.

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: AD, antidepressant; CI, confidence interval; NA, not available; NE, not estimable; NR, not reported; OBS, observational studies; OR, odds ratio; RR, relative risk; SSRI, selective serotonin reuptake inhibitor; TCA, tricyclic antidepressant.

[226] Laugesen 2013.

[227] Based on Malm 2016.

Table D3-12 Evidence Profile table: bupropion harms

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[228]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[229]**	**Risk with intervention[230]**
Cardiac malformations: see Section AppD4.1.1.4.2											
187,254 (1 – OBS)[231]	Serious(a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed** NA	**Bupropion (first trimester)** NA	RR 0.92 (0.69, 1.22)	6 per 1000[232]	6 per 1000 (4, 7)
5,381 (1 – OBS)[233]	Serious(a)	NA	None	Serious(b)	None	○○○○ Inadequate	**Bupropion (first trimester)** NA	**Other AD (first trimester)** NA	RR 0.54 (0.19, 1.51)[234]	Unknown	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of bupropion during the first trimester of pregnancy and cardiac malformation in the newborn, is uncertain.* *Due to the inadequate certainty of the evidence, any additional risk of cardiac malformation associated with maternal use of bupropion during the first trimester of pregnancy, compared with maternal used of other antidepressants during the same period, is uncertain.*											
ADHD: see Section AppD4.1.1.16.2											
38,074 (1 – OBS)[235]	Very serious(c)	NA	None	None	None	○○○○ Inadequate	**Unexposed** 37,960	**Bupropion (any time)** 114	**RR 3.63 (1.20, 11.0)**	10 per 1000[236]	36 per 1000 (12, 110)
38,074 (1 – OBS)[237]	Very serious(c)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** 37,995	**Bupropion (first trimester)** 79	RR 2.06 (0.35, 12.2)	10 per 1000[236]	21 per 1000 (4, 122)

[228] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns are noted and may result in further downgrading of the quality of the evidence.
[229] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[230] Calculated by multiplying relative effect by control risk.
[231] Huybrechts 2014a.
[232] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013.
[233] Cole 2007a.
[234] In the analysis, bupropion is used as the reference group.
[235] Figueroa 2010.
[236] Based on Malm 2016.
[237] Figueroa 2010.

Certainty assessment							**Summary of findings**				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[228]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[229]**	**Risk with intervention[230]**
38,074 (1 – OBS)[238]	Very serious(c)	NA	None	None	None	○○○○ Inadequate	**Unexposed** 38,036	**Bupropion (second trimester)** 46	**RR 14.7 (3.27, 65.7)**	10 per 1000[236]	147 per 1000 (33, 657)
38,074 (1 – OBS)[239]	Very serious(c)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** 38,037	**Bupropion (third trimester)** 37	NE[240]	10 per 1000[236]	-
38,074 (1 – OBS)[241]	Very serious(c)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** 37,889	**Bupropion (after pregnancy)** 185	RR 0.90 (0.32, 2.53)	10 per 1000[236]	9 per 1000 (3, 25)
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of bupropion at any time during or after pregnancy and attention deficit hyperactivity disorder in the child is uncertain.*											
Footnotes: a. Downgraded one level due to moderate risk of bias; potential for selection bias due to exclusion of planned abortions, miscarriages and still born from the analysis. b. Downgraded one level due to imprecision; 95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25, no measure of precision available, or no events. c. Downgraded two levels due to high risk of bias; lack of adjustment for potential confounding by maternal disease severity in the antenatal and postnatal period.											

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: AD, antidepressant; CI, confidence interval; NA, not available; NE, not estimable; NR, not reported; OBS, observational study/studies; RR, relative risk.

[238] Figueroa 2010.
[239] Figueroa 2010.
[240] No events.
[241] Figueroa 2010.

D3.1.2 Antipsychotics

The following section presents the Evidence Profile tables for any antipsychotics use, the specific antipsychotics classes, and individual medications examined. The quantity of evidence available for the assessment of antipsychotics was sufficient to limit the evidence to studies that adjusted risk estimates for confounding. A summary of the characteristics of the individual included studies can be found in **Table AppD2-11** in **Appendix D2.1.2.2**. A detailed discussion of the evidence for each group or individual intervention type and outcome can be found in **Appendix D4.1.2**.

The following observations were made regarding the body of evidence for antipsychotic harms:

- No meta-analyses were feasible for any outcome, so the body of evidence for each outcome comprised collections of studies or single studies.
- Three studies included an unexposed comparator group with a mental health diagnosis[242].

As the evidence was based on data from cohort and case-control studies, in many cases the results were presented as odds ratios instead of relative risks. Where the baseline risk was < 7%, it was assumed that the odds ratio approximates the relative risk and the results were interpreted as relative risks.

Table D 3-13 presents a summary of the results of the Evidence Review of antipsychotics and the location of the detailed assessment of the certainty of evidence in the evidence profile tables. Unlike antidepressants, groupings of antipsychotics were not pharmacologically-based, but instead grouped as any antipsychotics, second-generation antipsychotics (SGAs) and first-generation antipsychotics (FGAs). These groupings have been included here, although it is unclear how useful the grouped findings are, with the increased risks of harm associated with a number of the individual antipsychotics examined suggesting these may be masked when they are grouped together.

Maternal exposure to any antipsychotics and SGAs as a group appear to not be associated with the majority of outcomes assessed, although the available evidence for malformations, and neurodevelopmental outcomes were uncertain for any antipsychotics as a group. Exposure to the SGAs risperidone and quetiapine, FGAs as a group, and the long-acting agent flupentixol, are or may be associated with an increased risk of harm, including major and cardiac malformations (risperidone), miscarriage (quetiapine and flupentixol) and preterm birth (FGAs). For most outcomes assessed for individual antipsychotics, the certainty of the evidence was inadequate.

[242] Huybrechts 2016, Sorensen 2015, Lin 2010.

Table D 3-13 Summary of results of the Evidence Review for antipsychotics

Intervention	**Increased/may be increased risk of harm** Outcome Certainty of evidence	**Appears to be no increased risk of harm** Outcome Certainty of evidence	**Decreased/may be decreased risk of harm** Outcome Certainty of evidence	**Uncertain** Outcome ○○○○	**Evidence profile table**
Any antipsychotics		Neonatal mortality ●○○○ Stillbirth ●○○○ Miscarriage ●●○○ Preterm birth ●○○○ SFGA ●○○○ LFGA ●○○○ Seizures ●○○○ Respiratory distress ●○○○ PNAS ●○○○		Major malformation Cardiac malformation Neurodevelopment/ behavioural disorders Neuromotor performance	**Table D3-14**
SGAs		Major malformation ●○○○ Cardiac malformation ●○○○ Preterm birth ●○○○ SFGA ●○○○ LFGA ●○○○		Major malformations (vs FGAs)	**Table D3-15**
Aripiprazole		Major malformation ●○○○		Cardiac malformation	**Table D3-17**
Risperidone	Major malformation ●●○○ Cardiac malformation ●●○○				**Table D3-23**
Ziprasidone				Major malformation Cardiac malformation	**Table D3-24**
Olanzapine				Major malformation Cardiac malformation Miscarriage	**Table D3-20**
Quetiapine	Miscarriage ●○○○	Major malformation ●○○○		Cardiac malformation	**Table D3-22**
FGAs	Preterm birth ●●○○	SFGA ●○○○ LFGA ●○○○		Major malformation Cardiac malformation	**Table D3-16**
Haloperidol				Major malformation	**Table D3-19**
Perphenazine				Miscarriage	**Table D3-21**
Zuclopenthixol				Miscarriage	**Table D3-25**
Flupenthixol (long-acting)	Miscarriage ●○○○			Major malformation	**Table D3-18**

Abbreviations: FGA, first-generation antipsychotic; LFGA, large for gestational age; PNAS, poor neonatal adaptation syndrome; SFGA, small for gestational age; SGA, second-generation antipsychotics.

Note: All comparisons are against non-exposure, unless otherwise stated. Certainty of evidence gradings are as follows: ●●●● – high certainty; ●●●○ – moderate certainty; ●●○○ – low certainty; ●○○○ – very low certainty; ○○○○ – inadequate certainty.

Table D3-14 Evidence Profile table: any antipsychotics

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[243]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[244]**	**Risk with intervention[245]**
Major malformations: see Section AppD4.1.2.3.2											
(2 – OBS)[246]	Serious (a)	None	Serious (b)	None	None	○○○○ Inadequate	**Unexposed – no adjustment for indication** (N = 1,184733)	**Any antipsychotics (early pregnancy)** (N = 848)	**RR 1.49 (1.07, 2.06)**	41 per 1000[247]	-
(1 – OBS)[248]	Serious (a)	NA	None	Serious (c)	None	○○○○ Inadequate	**Discontinued ≥4 months before pregnancy – no further adjustment for indication** (N = 492)	**Any antipsychotics (early pregnancy)** (N = 290)	RR 1.79 (0.72, 4.47)	41 per 1000[249]	-
Evidence Statement: *Maternal use of any antipsychotic medication during early pregnancy may be associated with an increased risk of major malformation in the newborn, but due to the inadequate certainty of the evidence any such association is uncertain.*											
Cardiac malformations: see Section AppD4.1.2.4.2											
(1 – OBS)[250]	Serious (a)	NA	Serious (b)	Serious (c)	None	○○○○ Inadequate	**Unexposed – no adjustment for indication** (N = 1,575,847)	**Any antipsychotics or lithium[251] (early pregnancy)** (N = ~1,344)	OR 0.83 (0.48, 1.41)	15 per 1000[252]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of any antipsychotic medication during early pregnancy and cardiac malformation in the newborn is uncertain.*											

[243] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns will be noted and may result in further downgrading of the quality of the evidence.
[244] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[245] Calculated by multiplying relative effect by control risk.
[246] Petersen 2016a, Reis 2008.
[247] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.
[248] Petersen 2016a
[249] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.
[250] Källén 2013
[251] Lithium is the most commonly used (17% of neuroleptic-exposed women in the database), confounding the data for antipsychotics.
[252] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.

Certainty assessment							**Summary of findings**				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[243]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[244]**	**Risk with intervention[245]**
Neonatal mortality: see Section AppD4.1.2.5.2											
(1 – OBS)[254]	None	NA	None	Serious (c)	None	●○○○ Very low	**Unexposed – hdPS-matched (includes indication)** (N = 1,021)	**Any antipsychotics (≥2 consecutive scripts, one in 1st or 2nd trimester)** (N = 1,021)	RR 1.50 (0.53, 4.21)	6 per 1000[253]	9 per 1000 (3, 25)
Evidence Statement: *Maternal use of any antipsychotics during pregnancy does not appear to be associated with an increased risk of neonatal mortality (very low certainty evidence).*											
Stillbirth: see Section AppD4.1.2.5.2											
(1 – OBS)[254]	None	NA	None	Serious (c)	None	●○○○ Very low	**Unexposed – hdPS-matched (includes indication)** (N = 1,021)	**Any antipsychotics (≥2 consecutive scripts, one in 1st or 2nd trimester)** (N = 1,021)	RR 0.56 (0.25, 1.27)	16 per 1000[255]	9 per 1000 (4, 20)
Evidence Statement: *Maternal use of any antipsychotics during pregnancy does not appear to be associated with an increased risk of stillbirth (very low certainty evidence).*											
Miscarriage: see Section AppD4.1.2.6.2											
(1 – OBS)[256]	None	NA	None	None	None	●●○○ Low	**Discontinued ≥30 days before pregnancy – no further adjustment for indication** (N = 2,745)	**Any antipsychotics (any time from 30 days before, to end of pregnancy)** (N = 1,181)	RR 1.04 (0.93, 1.17)	197 per 1000[257]	205 per 1000 (183, 230)

[253] From hdPS-matched, unexposed cohort, Vigod 2015.
[254] Vigod 2015
[255] Vigod 2015 hdPS-matched, unexposed cohort.
[256] Sorensen 2015
[257] Sorensen 2015 unexposed patients with hospital diagnosis of severe mental disorder.

Certainty assessment							**Summary of findings**				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[243]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[244]**	**Risk with intervention[245]**
(1 – OBS)[256]	None	NA	None	Serious (c)	None	●○○○ Very low	**Unexposed patients with hospital diagnosis of severe mental disorder – no further adjustment for indication** (N = 1,337)	**Any antipsychotics, in patients with hospital diagnosis of severe mental disorder (any time from 30 days before, to end of pregnancy)** (N = 461)	RR 1.14 (0.94, 1.39)	197 per 1000[257]	225 per 1000 (185, 274)
Evidence Statement: *Maternal use of any antipsychotics during pregnancy does not appear to be associated with an increased risk of miscarriage (low certainty evidence).*											
Preterm birth: see Section AppD4.1.2.7.2											
(1 – OBS)[258]	None	NA	None	Serious (c)	None	●○○○ Very low	**Unexposed – hdPS-matched (includes indication) and adjusted for other psychotropic medication** (N = 1,021)	**Any antipsychotics (≥2 consecutive scripts, one in 1st or 2nd trimester)** (N = 1,021)	RR 0.99 (0.78, 1.26)	82 per 1000[259]	81 per 1000 (64, 103)
(1 – OBS)[258]	None	NA	None	Serious (c)	None	●○○○ Very low	**Unexposed – hdPS-matched (includes indication) and adjusted for other psychotropic medication** (N = 893)	**Any antipsychotics (1st trimester)** (N = 893)	RR 0.99 (0.77, 1.27)	82 per 1000[259]	81 per 1000 (63, 104)
(1 – OBS)[258]	None	NA	None	Serious (c)	None	●○○○ Very low	**Unexposed – hdPS-matched (includes indication) and adjusted for other psychotropic medication** (N = 758)	**Any antipsychotics (2nd trimester)** (N = 758)	RR 1.00 (0.75, 1.35)	82 per 1000[259]	82 per 1000 (62, 111)
(1 – OBS)[258]	None	NA	None	Serious (c)	None	●○○○ Very low	**Unexposed – hdPS-matched (includes indication) and adjusted for other psychotropic medication** (N = 614)	**Any antipsychotics (3rd trimester)** (N = 614)	RR 0.83 (0.59, 1.16)	82 per 1000[259]	68 per 1000 (48, 95)
Evidence Statement: *Maternal use of any antipsychotics during pregnancy (either first, second or third trimester) does not appear to be associated with an increased risk of preterm birth (very low certainty evidence).*											

[258] Vigod 2015
[259] Lin 2010 unexposed patients with schizophrenia.

Certainty assessment							**Summary of findings**				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[243]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[244]**	**Risk with intervention[245]**
Small for gestational age (<3rd centile): see Section AppD4.1.2.8.2											
(1 – OBS)[260]	None	NA	None	Serious (c)	None	●○○○ Very low	**Unexposed – hdPS-matched (includes indication) and adjusted for other psychotropic medication** (N = 1,021)	**Any antipsychotics (≥2 consecutive scripts, one in 1st or 2nd trimester)** (N = 1,021)	RR 1.21 (0.81, 1.82)	203 per 1000[261]	246 per 1000 (164, 369)
(1 – OBS)[260]	None	NA	None	Serious (c)	None	●○○○ Very low	**Unexposed – hdPS-matched (includes indication) and adjusted for other psychotropic medication** (N = 893)	**Any antipsychotics (1st trimester)** (N = 893)	RR 1.33 (0.88, 2.02)	203 per 1000[261]	270 per 1000 (179, 410)
(1 – OBS)[260]	None	NA	None	Serious (c)	None	●○○○ Very low	**Unexposed – hdPS-matched (includes indication) and adjusted for other psychotropic medication** (N = 758)	**Any antipsychotics (2nd trimester)** (N = 758)	RR 1.21 (0.74, 1.96)	203 per 1000[261]	246 per 1000 (150, 398)
(1 – OBS)[260]	None	NA	None	Serious (c)	None	●○○○ Very low	**Unexposed – hdPS-matched (includes indication) and adjusted for other psychotropic medication** (N = 614)	**Any antipsychotics (3rd trimester)** (N = 614)	RR 1.24 (0.73, 2.10)	203 per 1000[261]	252 per 1000 (148, 426)
Evidence Statement: *Maternal use of any antipsychotics during pregnancy (either first, second or third trimester) does not appear to be associated with an increased risk of the newborn being small for gestational age (very low certainty evidence).*											
Large for gestational age (>97th centile): see Section AppD4.1.2.9.2											
(1 – OBS)[262]	None	NA	None	Serious (c)	None	●○○○ Very low	**Unexposed – hdPS-matched (includes indication) and adjusted for other psychotropic medication** (N = 1,021)	**Any antipsychotics (≥2 consecutive scripts, one in 1st or 2nd trimester)** (N = 1,021)	RR 1.26 (0.69, 2.29)	97 per 1000[263]	122 per 1000 (67, 222)

[260] Vigod 2015
[261] Lin 2010 unexposed patients with schizophrenia.
[262] Vigod 2015
[263] Lin 2010 unexposed patients with schizophrenia.

Certainty assessment							**Summary of findings**				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[243]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[244]**	**Risk with intervention[245]**
(1 – OBS)[262]	None	NA	None	Serious (c)	None	●○○○ Very low	**Unexposed – hdPS-matched (includes indication) and adjusted for other psychotropic medication** (N = 893)	**Any antipsychotics (1st trimester)** (N = 893)	RR 0.94 (0.46, 1.93)	97 per 1000[263]	91 per 1000 (45, 187)
(1 – OBS)[262]	None	NA	None	Serious (c)	None	●○○○ Very low	**Unexposed – hdPS-matched (includes indication) and adjusted for other psychotropic medication** (N = 758)	**Any antipsychotics (2nd trimester)** (N = 758)	RR 1.83 (0.89, 3.77)	97 per 1000[263]	178 per 1000 (86, 366)
Evidence Statement: *Maternal use of any antipsychotics during pregnancy (either first or second trimester) does not appear to be associated with an increased risk of the newborn being large for gestational age (very low certainty evidence).*											
(1 – OBS)[262]	None	NA	None	None	None	●●○○ Low	**Unexposed – hdPS-matched (includes indication) and adjusted for other psychotropic medication** (N = 614)	**Any antipsychotics (3rd trimester)** (N = 614)	**RR 2.39 (1.00, 5.75)**	97 per 1000[263]	232 per 1000 (97, 558)
Evidence Statement: *Maternal use of any antipsychotics during the third trimester may be associated with an increased risk of the newborn being large for gestational age, from an absolute risk of 10% to 23% (low certainty evidence).*											
Seizures: see Section AppD4.1.2.10.2											
(1 – OBS)[262]	None	NA	None	Serious (c)	None	●○○○ Very low	**Unexposed – hdPS-matched (includes indication)** (N = 1,021)	**Any antipsychotics (≥2 consecutive scripts, one in 1st or 2nd trimester)** (N = 1,021)	RR 1.29 (0.48, 3.45)	7 per 1000[264]	9 per 1000 (3, 24)
Evidence Statement: *Maternal use of any antipsychotics during pregnancy does not appear to be associated with an increased risk of seizures in the newborn (very low certainty evidence).*											

[264] Vigod 2015 hdPS-matched, unexposed cohort.

Certainty assessment							**Summary of findings**				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[243]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[244]**	**Risk with intervention[245]**
Respiratory distress: see Section AppD4.1.2.11.2											
(1 – OBS)[265]	None	NA	None	Serious (c)	None	●○○○ Very low	**Unexposed – hdPS-matched (includes indication) and adjusted for other psychotropic medication** (N = 1,021)	**Any antipsychotics (≥2 consecutive scripts, one in 1st or 2nd trimester)** (N = 1,021)	RR 0.82 (0.46, 1.43)	29 per 1000[266]	24 per 1000 (13, 41)
Evidence Statement: *Maternal use of any antipsychotics during pregnancy does not appear to be associated with an increased risk of respiratory distress in newborns (very low certainty evidence).*											
Poor neonatal adaptation syndrome: see Section AppD4.1.2.12.2											
(1 – OBS)[265]	None	NA	None	Serious (c)	None	●○○○ Very low	**Unexposed – hdPS-matched (includes indication) and adjusted for other psychotropic medication** (N = 1,021)	**Any antipsychotics (≥2 consecutive scripts, one in 1st or 2nd trimester)** (N = 1,021)	RR 1.15 (0.88, 1.50)	109 per 1000[266]	125 per 1000 (96, 164)
(1 – OBS)[265]	None	NA	None	Serious (c)	None	●○○○ Very low	**Unexposed – hdPS-matched (includes indication)** (N = 151)	**Any antipsychotics (only in 1st or 2nd trimester)** (N = 151)	RR 1.50 (0.72, 3.11)	109 per 1000[266]	164 per 1000 (78, 339)
(1 – OBS)[265]	None	NA	None	Serious (c)	None	●○○○ Very low	**Unexposed – hdPS-matched (includes indication) and adjusted for other psychotropic medication** (N = 614)	**Any antipsychotics (only in 3rd trimester)** (N = 614)	RR 1.31 (0.91, 1.90)	109 per 1000[266]	164 per 1000 (78, 339)
Evidence Statement: *Maternal use of any antipsychotics during pregnancy does not appear to be associated with an increased risk of PNAS in newborns (very low certainty evidence).*											
Neurodevelopmental outcomes: see Section AppD4.1.2.13.2											
Neurodevelopment/behavioural disorders											
(1 – OBS)[267]	Very serious (d)	NA	Serious (b)	Serious (c)	None	○○○○ Inadequate	**Unexposed – no adjustment for indication** (N = 210,966)	**Any antipsychotics (early; 31-105 days)** (N = 290)	RR 1.22 (0.80, 1.84)	102 per 1000[268]	-

[265] Vigod 2015
[266] Vigod 2015 hdPS-matched, unexposed cohort.
[267] Petersen 2016a
[268] Petersen 2016a women who discontinued antipsychotics.

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[243]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[244]**	**Risk with intervention[245]**
(1 – OBS)[269]	Very serious (d)	NA	None	Serious (c)	None	○○○○ Inadequate	**Discontinued ≥4 months before pregnancy – no further accounting for indication** (N = 492)	**Any antipsychotics (early; 31-105 days)** (N = 290)	RR 0.83 (0.49, 1.39)	102 per 1000[268]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of any antipsychotics during pregnancy and an increased risk of neurodevelopment or behavioural disorders in the child is uncertain.*											
Neuromotor performance (INFANIB)											
(1 – OBS)[270]	Very serious (e)	NA	None	None	None	○○○○ Inadequate	**Unexposed –adjusted for lifetime history of psychiatric illness[271]** (N = 85)	**Any antipsychotic (pregnancy)** (N = 22)	**OR 5.41[272] (1.22, 24.09)**	unknown	-
(1 – OBS)[270]	Very serious (e)	NA	None	None	None	○○○○ Inadequate	**Antidepressants –adjusted for lifetime history of psychiatric illness[273]** (N = 202)	**Any antipsychotic (pregnancy)** (N = 22)	**OR 4.11[272] (1.05, 15.99)**	unknown	-
Evidence Statement: *Maternal use of any antipsychotics during pregnancy may be associated with an increased risk of poor neuromotor performance in the child, but due to the inadequate certainty of the evidence any such association is uncertain.*											

Footnotes:
a. Downgraded one level due to moderate risk of bias; potential selection bias due to not capturing potential excess malformations coinciding with miscarriage, abortion or stillbirth.
b. Downgraded one level due to indirectness caused by use of control group without a mental health disorder diagnosis, with no adjustment for indication.
c. Downgraded one level due to imprecision (95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25).
d. Downgraded two levels due to high risk of bias: neurodevelopmental outcome without adjustment for maternal disease severity and use of a non-validated outcome assessment tool.
e. Downgraded two levels due to high risk of bias: neurodevelopmental outcome without adjustment for maternal disease severity and use of a non-validated output from an outcome assessment tool.

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.
Abbreviations: CI, confidence interval; NA, not available; OBS, observational studies; OR, odds ratio; RR, risk ratio.

[269] Petersen 2016a
[270] Johnson 2012
[271] No data reported regarding psychiatric status at pregnancy or at infant assessment psychiatric status, but 62% of comparator group had experienced psychiatric illness in their lifetime.
[272] This is the adjusted odds ratio for the likelihood of a normal score in the unexposed group. This indicates a significantly higher likelihood of a 'not normal' score in the exposed group (categories condensed into two for increased power in statistical analysis).
[273] No data reported regarding psychiatric status at pregnancy or at infant assessment psychiatric status, but 62% of comparator group had experienced psychiatric illness in their lifetime.

Table D3-15 Evidence Profile table: SGAs

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[274]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk Unexposed[275]**	**Risk Exposed[276]**
Major malformations: see Section AppD4.1.2.3.2											
(1 – OBS)[277]	Serious (a)	None	None	None	None	●○○○ Very low	**Unexposed – fully adjusted (indication, medication, propensity score)** (N = 1,289,826)	**SGAs (1st trimester)** (N = 9,237)	RR 1.05 (0.96, 1.16)	41 per 1000[278]	43 per 1000 (39, 48)
(1 – OBS)[277]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed, restricted to schizophrenia, bipolar disorder, psychosis – fully adjusted (indication, medication, propensity score)** (N = 11,606)	**SGAs, restricted to schizophrenia, bipolar disorder, psychosis (1st trimester)** (N = 3,995)	RR 1.16 (0.99, 1.35)	41 per 1000[278]	-
(1 – OBS)[279]	Serious (a)	N/A	None	Serious (b)	None	○○○○ Inadequate	**FGAs – no further adjustment for indication (pregnancy)** (N = 284)	**SGAs (pregnancy)** (N = 561)	OR 1.27 (0.57, 2.82)	41 per 1000[278]	-
Evidence Statement: *Maternal use of SGAs during the first trimester of pregnancy does not appear to be associated with an increased risk of major malformation in the newborn (very low certainty evidence).* *Due to the inadequate certainty of the evidence, any additional risk of major malformations in the newborn associated with maternal use of SGAs at any time during pregnancy, compared with maternal use of FGAs during the same period, is uncertain.*											
Cardiac malformations: See AppD4.1.2.4.2											
(1 – OBS)[277]	Serious (a)	NA	None	None	None	●○○○ Very low	**Unexposed – fully adjusted (indication, medication, propensity score)** (N = 1,289,826)	**SGAs (1st trimester)** (N = 9,237)	RR 1.06 (0.90, 1.24)	15 per 1000[280]	16 per 1000 (14, 19)

[274] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns will be noted and may result in further downgrading of the quality of the evidence.
[275] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[276] Calculated by multiplying relative effect by control risk.
[277] Huybrechts 2016
[278] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.
[279] Habermann 2013
[280] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[274]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk Unexposed[275]**	**Risk Exposed[276]**
(1 – OBS)[277]	Serious (a)	N/A	None	Serious (b)	None	○○○○ Inadequate	**Unexposed, restricted to schizophrenia, bipolar disorder, psychosis – fully adjusted (indication, medication, propensity score)** (N = 11,606)	**SGA s, restricted to schizophrenia, bipolar disorder, psychosis (1st trimester)** (N = 3,995)	RR 1.21 (0.93, 1.57)	15 per 1000[280]	-
Evidence Statement: *Maternal use of SGAs during the first trimester of pregnancy does not appear to be associated with an increased risk of cardiac malformation in the newborn (very low certainty evidence).*											
Preterm birth (<37 weeks): see Section AppD4.1.2.7.2											
(1 – OBS)[281]	None	NA	None	Serious (b)	None	●○○○ Very low	**Unexposed to FGAs or SGAs, schizophrenia[282]** (N = 454)	**SGAs, schizophrenia (pregnancy)** (N = 48)	OR 1.61 (0.63, 4.12)	82 per 1000[283]	not estimable (OR ≠ RR when assumed risk >5%)
Evidence Statement: *Maternal use of SGAs during pregnancy does not appear to be associated with an increased risk of preterm birth (very low certainty evidence).*											
Small for gestation age (<10rd centile): see Section AppD4.1.2.8.2											
(1 – OBS)[281]	None	NA	None	Serious (b)	None	●○○○ Very low	**Unexposed to FGAs or SGAs, schizophrenia[282]** (N = 454)	**SGAs, schizophrenia (pregnancy)** (N = 48)	OR 1.15 (0.55, 2.41)	203 per 1000[284]	not estimable (OR ≠ RR when assumed risk >5%)
Evidence Statement: *Maternal use of SGAs during pregnancy does not appear to be associated with an increased risk of the newborn being small for gestational age (very low certainty evidence).*											
Large for gestation age (>90th centile): see Section AppD4.1.2.9.2											
(1 – OBS)[281]	None	NA	None	Serious (b)	None	●○○○ Very low	**Unexposed to FGAs or SGAs, schizophrenia[282]** (N = 454)	**SGAs, schizophrenia (pregnancy)** (N = 48)	OR 0.55 (0.16, 1.85)	97 per 1000[285]	not estimable (OR ≠ RR when assumed risk >5%)
Evidence Statement: *Maternal use of SGAs during pregnancy does not appear to be associated with an increased risk of the newborn being large for gestational age (very low certainty evidence).*											

[281] Lin 2010
[282] Smoking was not adjusted for in this study, but this confounder will be largely accounted for by the use of a comparator group of women with schizophrenia.
[283] Lin 2010 unexposed patients with schizophrenia.
[284] Lin 2010 unexposed patients with schizophrenia.
[285] Lin 2010 unexposed patients with schizophrenia.

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[274]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk Unexposed[275]**	**Risk Exposed[276]**
Footnotes: a. Downgraded one level due to moderate risk of bias; potential selection bias due to not capturing potential excess malformations coinciding with miscarriage, abortion or stillbirth. b. Downgraded one level due to imprecision (95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25).											

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: CI, confidence interval; FGA, first generation antipsychotics; NA, not available; OBS, observational studies; OR, odds ratio; RR, risk ratio; SGA, second generation antipsychotic.

Table D3-16 Evidence Profile table: FGAs

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[286]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk Unexposed[287]**	**Risk Exposed[288]**
Major malformations: see Section AppD4.1.2.3.2											
(1 – OBS)[289]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed – fully adjusted (indication, medication, propensity score)** (N = 1,297,638)	**FGAs (1st trimester)** (N = 727)	RR 0.90 (0.62, 1.31)	41 per 1000[290]	-
(1 – OBS)[289]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed, restricted to psychosis, schizophrenia, bipolar disorder– fully adjusted (indication, meds, propensity score)** (N = 10,418)	**FGAs, restricted to schizophrenia, bipolar disorder, psychosis (1st trimester)** (N = 381)	RR 0.93 (0.57, 1.51)	41 per 1000[291]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of FGAs during the first trimester of pregnancy and an increased risk of major malformation in the newborn is uncertain.*											
Cardiac malformations: see Section AppD4.1.2.4.2											
(1 – OBS)[289]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed – fully adjusted (indication, medication, propensity score)** (N = 1,297,638)	**FGAs (1st trimester)** (N = 727)	RR 0.75 (0.39, 1.43)	15 per 1000[292]	-

[286] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns will be noted and may result in further downgrading of the quality of the evidence.
[287] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[288] Calculated by multiplying relative effect by control risk.
[289] Huybrechts 2016
[290] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.
[291] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.
[292] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.

Certainty assessment							**Summary of findings**				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[286]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk Unexposed[287]**	**Risk Exposed[288]**
(1 – OBS)[289]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed, restricted to psychosis, schizophrenia, bipolar disorder – fully adjusted (indication, medication, propensity score)** (N = 10,418)	**FGAs, restricted to schizophrenia, bipolar disorder, psychosis (1st trimester)** (N = 381)	RR 0.91 (0.43, 1.91)	15 per 1000[293]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of FGAs during the first trimester of pregnancy and an increased risk of cardiac malformation in the newborn is uncertain*											
Preterm birth (<37 weeks): see Section AppD4.1.2.7.2											
(1 – OBS)[294]	None	NA	None	None	None	●●○○ Low	**Unexposed to FGAs or SGAs, schizophrenia[295]** (N = 454)	**FGAs, schizophrenia (pregnancy)** (N = 194)	**OR 2.46 (1.50, 4.11)**	82 per 1000[296]	not estimable (OR ≠ RR when assumed risk >5%)
Evidence Statement: *Maternal use of FGAs during pregnancy is associated with an increased risk of preterm birth, with a 2.5-fold increase of an absolute risk of 8% (low certainty evidence).*											
Small for gestational age (<10th centile): see Section AppD4.1.2.8.2											
(1 – OBS)[294]	None	NA	None	Serious (b)	None	●○○○ Very low	**Unexposed to FGAs or SGAs, schizophrenia[295]** (N = 454)	**FGAs, schizophrenia (pregnancy)** (N = 194)	OR 1.39 (0.93, 2.08)	203 per 1000[297]	not estimable (OR ≠ RR when assumed risk >5%)
Evidence Statement: *Maternal use of FGAs during pregnancy does not appear to be associated with an increased risk of the newborn being small for gestational age (very low certainty evidence).*											
Large for gestational age (>90th centile): see Section AppD4.1.2.9.2											
(1 – OBS)[294]	None	NA	None	Serious (b)	None	●○○○ Very low	**Unexposed to FGAs or SGAs, schizophrenia[295]** (N = 454)	**FGAs, schizophrenia (pregnancy)** (N = 194)	OR 0.72 (0.39, 1.34)	97 per 1000[298]	not estimable (OR ≠ RR when assumed risk >5%)
Evidence Statement: *Maternal use of FGAs during pregnancy does not appear to be associated with an increased risk of the newborn being large for gestational age (very low certainty evidence).*											

[293] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.
[294] Lin 2010
[295] Smoking was not adjusted for in this study, but this confounder will be largely accounted for by the use of a comparator group of women with schizophrenia.
[296] Lin 2010 unexposed patients with schizophrenia.
[297] Lin 2010 unexposed patients with schizophrenia.
[298] Lin 2010 unexposed patients with schizophrenia.

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[286]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk Unexposed[287]**	**Risk Exposed[288]**
Footnotes: a. Downgraded one level due to moderate risk of bias; potential selection bias due to not capturing potential excess malformations coinciding with miscarriage, abortion or stillbirth. b. Downgraded one level due to imprecision (95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25).											

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: CI, confidence interval; FGA, first generation antipsychotics; meds, medication; NA, not available; OBS, observational studies; OR, odds ratio; RR, risk ratio; SGA, second generation antipsychotic.

Table D3-17 Evidence Profile table: aripiprazole

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias**[299]	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk Unexposed**[300]	**Risk Exposed**[301]
Major malformations: see Section AppD4.1.2.3.2											
(1 – OBS)[302]	Serious (a)	NA	None	None	None	●○○○ Very low	**Unexposed – fully adjusted (indication, medication, propensity score)** (N = 957,012)	**Aripiprazole (1st trimester)** (N = 1,752)	RR 0.95 (0.76, 1.19)	41 per 1000[303]	39 per 1000 (31, 49)
(1 – OBS)[302]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed, restricted to psychosis, schizophrenia, bipolar disorder – fully adjusted (indication, meds, propensity score)** (N = 10,174)	**Aripiprazole, restricted to schizophrenia, bipolar disorder, psychosis (1st trimester)** (N = 949)	RR 1.13 (0.86, 1.50)	41 per 1000[303]	-
Evidence Statement: *Maternal use of aripiprazole during the first trimester of pregnancy does not appear to be associated with an increased risk of major malformation in the newborn (very low certainty evidence).*											
Cardiac malformations: see Section AppD4.1.2.4.2											
(1 – OBS)[302]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed – fully adjusted (indication, medication, propensity score)** (N = 957,012)	**Aripiprazole (1st trimester)** (N = 1,752)	RR 0.93 (0.64, 1.37)	15 per 1000[304]	-
(1 – OBS)[302]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed, restricted to psychosis, schizophrenia, bipolar disorder – fully adjusted (indication, meds, propensity score)** (N = 10,174)	**Aripiprazole, restricted to schizophrenia, bipolar disorder, psychosis (1st trimester)** (N = 949)	RR 1.13 (0.71, 1.80)	15 per 1000[305]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of aripiprazole during the first trimester of pregnancy and an increased risk of cardiac malformation in the newborn is uncertain.*											

[299] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns will be noted and may result in further downgrading of the quality of the evidence.
[300] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[301] Calculated by multiplying relative effect by control risk.
[302] Huybrechts 2016
[303] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.
[304] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.
[305] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[299]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk Unexposed[300]**	**Risk Exposed[301]**
Footnotes: a. Downgraded one level due to moderate risk of bias; potential selection bias due to not capturing potential excess malformations coinciding with miscarriage, abortion or stillbirth. b. Downgraded one level due to imprecision (95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25).											

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: CI, confidence interval; NA, not available; OBS, observational studies; RR, risk ratio.

Table D3-18 Evidence Profile table: flupenthixol

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias**[306]	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk Unexposed**[307]	**Risk Exposed**[308]
Major malformations: see Section AppD4.1.2.3.2											
(1 – OBS)[309]	Serious (a)	NA	Serious (b)	Serious (c)	None	○○○○ Inadequate	**Unexposed – no adjustment for indication** (N = 1,575,847)	**Flupenthixol (early pregnancy)** (N = 154)	RR 1.94 (1.00, 3.40)[310]	41 per 1000[311]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of flupenthixol during early pregnancy and an increased risk of major malformation in the newborn is uncertain.*											
Miscarriage: see Section AppD4.1.2.6.2											
(1 – OBS)[312]	None	NA	Serious (b)	None	None	●○○○ Very low	**Unexposed – no adjustment for indication** (N = 841,183)	**Flupenthixol (any time from 30 days before, to the end of pregnancy)** (N = 233)	**RR 1.55 (1.22, 1.97)**	197 per 1000[313]	305 per 1000 (240, 388)
Evidence Statement: *Maternal use of flupenthixol during or just prior to pregnancy may be associated with an increased risk of miscarriage, from an absolute risk of 20% to 30% (very low certainty evidence).*											

Footnotes:
a. Downgraded one level due to moderate risk of bias; potential selection bias due to not capturing potential excess malformations coinciding with miscarriage, abortion or stillbirth.
b. Downgraded one level due to indirectness caused by use of control group without a mental health disorder diagnosis, with no adjustment for indication.
c. Downgraded one level due to imprecision (95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25).

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.
No distinction was made between long-acting versus oral flupenthixol for either of the included studies.
Abbreviations: CI, confidence interval; NA, not available; OBS, observational studies; RR, risk ratio.

[306] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns will be noted and may result in further downgrading of the quality of the evidence.
[307] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[308] Calculated by multiplying relative effect by control risk.
[309] Källén 2013
[310] As the expected number of events in the exposed group was less than 10, a RR was calculated instead of OR, using the observed over expected number with 95% CI from exact Poisson distributions.
[311] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.
[312] Sorensen 2015
[313] Sorensen 2015 unexposed patients with hospital diagnosis of severe mental disorder.

Table D3-19 Evidence Profile table: haloperidol, infant harms

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias**[314]	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk Unexposed**[315]	**Risk Exposed**[316]
Major malformations: see Section AppD4.1.2.3.2											
(1 – OBS)[317]	Serious (a)	NA	Serious (b)	Serious (c)	None	○○○○ Inadequate	**Unexposed – no adjustment for indication** (N = 1,575,847)	**Haloperidol (early pregnancy)** (N = 115)	RR 1.21 (0.39, 2.83)[318]	41 per 1000[319]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of haloperidol during early pregnancy and an increased risk of major malformation in the newborn is uncertain.*											
Footnotes: a. Downgraded one level due to moderate risk of bias; potential selection bias due to not capturing potential excess malformations coinciding with miscarriage, abortion or stillbirth b. Downgraded one level due to indirectness caused by use of control group without a mental health disorder diagnosis, with no adjustment for indication. c. Downgraded one level due to imprecision (95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25).											

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: CI, confidence interval; NA, not available; OBS, observational studies; RR, risk ratio.

[314] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns will be noted and may result in further downgrading of the quality of the evidence.

[315] Based on average risk from unexposed, depressed control groups of population-based cohort studies.

[316] Calculated by multiplying relative effect by control risk.

[317] Källén 2013

[318] As the expected number of events in the exposed group was less than 10, a RR was calculated instead of OR, using the observed over expected number with 95% CI from exact Poisson distributions

[319] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.

Table D3-20 Evidence Profile table: olanzapine, infant harms

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias**[320]	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk Unexposed**[321]	**Risk Exposed**[322]
Major malformations: see Section AppD4.1.2.3.2											
(1 – OBS)[323]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed – fully adjusted (indication, medication, propensity score)** (N = 1,231,441)	**Olanzapine (1st trimester)** (N = 1,392)	RR 1.09 (0.85, 1.41)	41 per 1000[324]	-
(1 – OBS)[323]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed, restricted to psychosis, schizophrenia, bipolar disorder – fully adjusted (indication, medication, propensity score)** (N = 10,949)	**Olanzapine, restricted to schizophrenia, bipolar disorder, psychosis (1st trimester)** (N = 648)	RR 1.19 (0.84, 1.67)	41 per 1000[325]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of olanzapine during the first trimester of pregnancy and an increased risk of major malformation in the newborn is uncertain.*											
Cardiac malformations: see Section AppD4.1.2.4.2											
(1 – OBS)[323]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed – fully adjusted (indication, medication, propensity score)** (N = 1,231,441)	**Olanzapine (1st trimester)** (N = 1,392)	RR 0.99 (0.64, 1.53)	15 per 1000[326]	-

[320] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns will be noted and may result in further downgrading of the quality of the evidence.
[321] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[322] Calculated by multiplying relative effect by control risk.
[323] Huybrechts 2016
[324] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.
[325] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.
[326] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.

Certainty assessment							**Summary of findings**				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[320]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk Unexposed[321]**	**Risk Exposed[322]**
(1 – OBS)[323]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed, restricted to psychosis, schizophrenia, bipolar disorder – fully adjusted (indication, medication, propensity score)** (N = 10,949)	**Olanzapine, restricted to schizophrenia, bipolar disorder, psychosis (1st trimester)** (N = 648)	RR 1.23 (0.69, 2.19)	15 per 1000[327]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of olanzapine during the first trimester of pregnancy and an increased risk of cardiac malformation in the newborn is uncertain.*											
Miscarriage: see Section AppD4.1.2.6.2											
(1 – OBS)[328]	None	NA	Serious (c)	Serious (b)	None	○○○○ Inadequate	**Unexposed – no adjustment for indication** (N = 841,183)	**Olanzapine (any time from 30 days before, to the end of pregnancy)** (N = 223)	RR 1.10 (0.83, 1.46)	197 per 1000[329]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of olanzapine during or just prior to pregnancy and an increased risk of miscarriage is uncertain*											
Footnotes: a. Downgraded one level due to moderate risk of bias; potential selection bias due to not capturing potential excess malformations coinciding with miscarriage, abortion or stillbirth. b. Downgraded one level due to imprecision (95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25). c. Downgraded one level due to indirectness caused by use of control group without a mental health disorder diagnosis, with no adjustment for indication.											

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: CI, confidence interval; NA, not available; OBS, observational studies; RR, risk ratio.

[327] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.

[328] Sorensen 2015

[329] Sorensen 2015 unexposed patients with hospital diagnosis of severe mental disorder.

Table D3-21 Evidence Profile table: perphenazine, infant harms

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias**[330]	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk Unexposed**[331]	**Risk Exposed**[332]
Miscarriage: see Section AppD4.1.2.6.2											
(1 – OBS)[333]	None	NA	Serious (a)	Serious (b)	None	○○○○ Inadequate	**Unexposed – no adjustment for indication** (N = 841,183)	**Perphenazine (any time from 30 days before, to end of pregnancy)** (N = 229)	RR 1.25 (0.95 1.64)	197 per 1000[334]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of perphenazine during or just prior to pregnancy and an increased risk of miscarriage is uncertain.*											
Footnotes: a. Downgraded one level due to high risk of bias; not limiting control group to women with a mental health disorder and not controlling for indication. b. Downgraded one level due to imprecision (95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25).											

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: CI, confidence interval; NA, not available; OBS, observational studies; RR, risk ratio.

[330] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns will be noted and may result in further downgrading of the quality of the evidence.

[331] Based on average risk from unexposed, depressed control groups of population-based cohort studies.

[332] Calculated by multiplying relative effect by control risk.

[333] Sorensen 2015

[334] Sorensen 2015 unexposed patients with hospital diagnosis of severe mental disorder.

Table D3-22 Evidence Profile table: quetiapine, infant harms

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[335]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk Unexposed[336]**	**Risk Exposed[337]**
Major malformations: See AppD4.1.2.3.2											
Exposed: 4,213 Unexposed: 1,161,955 (1 – OBS)[338]	Serious (a)	NA	None	None	None	●○○○ Very low	**Unexposed – fully adjusted (indication, medication, propensity score)** (N = 1,161,955)	**Quetiapine (1st trimester)** (N = 4,213)	RR 1.01 (0.88, 1.17)	41 per 1000[339]	41 per 1000 (36, 48)
Exposed: 1,747 Unexposed: 11,440 (1 – OBS)[338]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed, restricted to psychosis, schizophrenia, bipolar disorder – fully adjusted (indication, medication, propensity score)** (N = 11,440)	**Quetiapine, restricted to schizophrenia, bipolar disorder, psychosis (1st trimester)** (N = 1,747)	RR 1.13 (0.92, 1.41)	41 per 1000[340]	-
Evidence Statement: *Maternal use of quetiapine during the first trimester of pregnancy does not appear to be associated with an increased risk of major malformation in the newborn (very low certainty evidence).*											
Cardiac malformations: See AppD4.1.2.4.2											
Exposed: 4,213 Unexposed: 1,161,955 (1 – OBS)[338]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed – fully adjusted (indication, medication, propensity score)** (N = 1,161,955)	**Quetiapine (1st trimester)** (N = 4,213)	RR 1.07 (0.85, 1.35)	15 per 1000[341]	-

[335] [335] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns will be noted and may result in further downgrading of the quality of the evidence.
[336] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[337] Calculated by multiplying relative effect by control risk.
[338] Huybrechts 2016
[339] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.
[340] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.
[341] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.

Certainty assessment							**Summary of findings**				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[335]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk Unexposed[336]**	**Risk Exposed[337]**
Exposed: 1,747 Unexposed: 11,440 (1 – OBS)[338]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed, restricted to psychosis, schizophrenia, bipolar disorder – fully adjusted (indication, medication, propensity score)** (N = 11,440)	**Quetiapine, restricted to schizophrenia, bipolar disorder, psychosis (1st trimester)** (N = 1,747)	RR 1.17 (0.81, 1.67)	15 per 1000[342]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of quetiapine during the first trimester of pregnancy and an increased risk of cardiac malformation in the newborn is uncertain.*											
Miscarriage: see Section AppD4.1.2.6.2											
Exposed: 174 Unexposed: 841,183 (1 – OBS)[343]	None	NA	Serious (c)	No serious	None	●○○○ Very low	**Unexposed – no adjustment for indication** (N = 841,183)	**Quetiapine (any time from 30 days before, to end of pregnancy)** (N = 174)	**RR 1.65 (1.28, 2.15)**	197 per 1000[344]	325 per 1000 (252, 424)
Evidence Statement: *Maternal use of quetiapine during or just prior to pregnancy may be associated with an increased risk of miscarriage, from an absolute risk of 20% to 33% (very low certainty evidence).*											
Footnotes: a. Downgraded one level due to moderate risk of bias; potential selection bias due to not capturing potential excess malformations coinciding with miscarriage, abortion or stillbirth. b. Downgraded one level due to imprecision (95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25). c. Downgraded one level due to high risk of bias; not limiting control group to women with a mental health disorder and not controlling for indication.											

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: CI, confidence interval; NA, not available; OBS, observational studies; RR, risk ratio.

[342] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.

[343] Sorensen 2015

[344] Sorensen 2015 unexposed patients with hospital diagnosis of severe mental disorder.

Table D3-23 Evidence Profile table: risperidone, infant harms

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias**[345]	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk Unexposed**[346]	**Risk Exposed**[347]
Major malformations: See AppD4.1.2.3.2											
(1 – OBS)[348]	None[349]	NA	None	None	None	●●○○ Low	**Unexposed – fully adjusted (indication, medication, propensity score)** (N = 1,290,485)	**Risperidone (1st trimester)** (N = 1,565)	**RR 1.26 (1.02, 1.56)**	41 per 1000[350]	52 per 1000 (42, 64)
(1 – OBS)[348]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed, restricted to psychosis, schizophrenia, bipolar disorder – fully adjusted (indication, medication, propensity score)** (N = 11,497)	**Risperidone, restricted to schizophrenia, bipolar disorder, psychosis (1st trimester)** (N = 740)	RR 1.19 (0.86, 1.64)	41 per 1000[351]	-
(1 – OBS)[348]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Discontinued: no Rx from 8 weeks before pregnancy – no further adjustment for indication (PS adjusted)** (N = 496)	**Risperidone, continued use from 3 months prior (1st trimester)** (N = 866)	RR 1.64 (0.90, 2.98)	41 per 1000[352]	-
Evidence Statement: *Maternal use of risperidone during the first trimester of pregnancy may be associated with an increased risk of major malformation in the newborn, from an absolute risk of 4% to 5% (low certainty evidence).*											

[345] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns will be noted and may result in further downgrading of the quality of the evidence.
[346] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[347] Calculated by multiplying relative effect by control risk.
[348] Huybrechts 2016
[349] This outcome normally carries an increased risk of bias is due to the possibility of missing malformations in the exposed group and thereby not detecting any increased risk associated with exposure. As a statistically significant increase in risk is reported, the only remaining risk of bias associated with this risk estimate is an underestimation of magnitude. Therefore, in this instance, it seems reasonable not to apportion additional risk of bias to the major malformations outcome in this analysis.
[350] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.
[351] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.
[352] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[345]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk Unexposed[346]**	**Risk Exposed[347]**
Cardiac malformations: see Section AppD4.1.2.4.2											
(1 – OBS)[348]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed – fully adjusted (indication, medication, propensity score)** (N = 1,290,485)	**Risperidone (1st trimester)** (N = 1,565)	RR 1.26 (0.88, 1.81)	15 per 1000[353]	-
(1 – OBS)[348]	None[354]	NA	None	None	None	●●○○ Low	**Unexposed, restricted to psychosis, schizophrenia, bipolar disorder – fully adjusted (indication, medication, propensity score)** (N = 11,497)	**Risperidone, restricted to schizophrenia, bipolar disorder, psychosis (1st trimester)** (N = 740)	**RR 1.64 (1.03, 2.62)**	15 per 1000[353]	25 per 1000 (15, 39)
(1 – OBS)[348]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Discontinued: no Rx from 8 weeks before pregnancy – no further adjustment for indication (PS adjusted)** (N = 496)	**Risperidone, continued use from 3 months prior (1st trimester)** (N = 866)	RR 2.46 (0.77, 7.87)	15 per 1000[353]	-
(1 – OBS)[348]	None[354]	NA	None	None	None	●●○○ Low	**Unexposed – fully adjusted (indication, medication, propensity score)** (N = 1,094,959)	**Risperidone, ≥2mg/day[355] (1st trimester)** (N = 609)	**RR 2.08 (1.32, 3.28)**	15 per 1000[353]	19 per 1000 (13, 27)
Evidence Statement: *Maternal use of risperidone during the first trimester of pregnancy may be associated with an increased risk of cardiac malformation in the newborn, from an absolute risk of 1.5% to 2.5% (low certainty evidence).*											
Footnotes: a. Downgraded one level due to moderate risk of bias; potential selection bias due to not capturing potential excess malformations coinciding with miscarriage, abortion or stillbirth. b. Downgraded one level due to imprecision (95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25).											

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: CI, confidence interval; NA, not available; OBS, observational studies; PS, propensity score; RR, risk ratio; Rx, prescription.

[353] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.

[354] This outcome normally carries and increased risk of bias is due to the possibility of missing malformations in the exposed group and thereby not detecting any increased risk associated with exposure. As a statistically significant increase in risk is reported, the only remaining risk of bias associated with this risk estimate is an underestimation of magnitude. Therefore, in this instance, it seems reasonable not to apportion additional risk of bias to the cardiac malformations outcome in this analysis.

[355] Doses less than 1 mg and doses from 1-2 mg were also analysed, and neither group showed a statistically significant increase in risk of cardiac malformations.

Table D3-24 Evidence Profile table: ziprasidone, infant harms

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[356]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk Unexposed[357]**	**Risk Exposed[358]**
Major malformations: see Section AppD4.1.2.3.2											
(1 – OBS)[359]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed – fully adjusted (indication, medication, propensity score)** (N = 979,614)	**Ziprasidone (1st trimester)** (N = 696)	RR 0.88 (0.60, 1.28)	41 per 1000[360]	-
(1 – OBS)[359]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed, restricted to psychosis, schizophrenia, bipolar disorder – fully adjusted (indication, meds, propensity score)** (N = 10,971)	**Ziprasidone, restricted to schizophrenia, bipolar disorder, psychosis (1st trimester)** (N = 425)	RR 0.84 (0.51, 1.39)	41 per 1000[361]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of ziprasidone during the first trimester of pregnancy and an increased risk of major malformation in the newborn is uncertain.*											
Cardiac malformations: see Section AppD4.1.2.4.2											
(1 – OBS)[359]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed – fully adjusted (indication, medication, propensity score)** (N = 979,614)	**Ziprasidone (1st trimester)** (N = 696)	RR 0.85 (0.44, 1.63)	15 per 1000[353]	-
(1 – OBS)[359]	Serious (a)	NA	None	Serious (b)	None	○○○○ Inadequate	**Unexposed, restricted to psychosis, schizophrenia, bipolar disorder – fully adjusted (indication, meds, propensity score)** (N = 10,971)	**Ziprasidone, restricted to schizophrenia, bipolar disorder, psychosis (1st trimester)** (N = 425)	RR 0.75 (0.31, 1.81)	15 per 1000[353]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of ziprasidone during the first trimester of pregnancy and increased risk of cardiac malformation in the newborn is uncertain.*											

[356] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns will be noted and may result in further downgrading of the quality of the evidence.
[357] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[358] Calculated by multiplying relative effect by control risk.
[359] Huybrechts 2016
[360] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.
[361] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[356]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk Unexposed[357]**	**Risk Exposed[358]**
Footnotes: a. Downgraded one level due to moderate risk of bias; potential selection bias due to not capturing potential excess malformations coinciding with miscarriage, abortion or stillbirth. b. Downgraded one level due to imprecision (95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25).											

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control.
Abbreviations: CI, confidence interval; NA, not available; OBS, observational studies; RR, risk ratio.

Table D3-25 Evidence Profile table: zuclopenthixol, infant harms

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[362]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk Unexposed[363]**	**Risk Exposed[364]**
Miscarriage: see Section AppD4.1.2.6.2											
(1 – OBS)[365]	None	NA	Serious (a)	Serious (b)	None	○○○○ Inadequate	**Unexposed – no adjustment for indication** (N = 841,183)	**Zuclopenthixol (any time from 30 days before, to end of pregnancy)** (N = 229)	RR 1.26 (0.95, 1.66)	41 per 1000[366]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of zuclopenthixol during or just prior to pregnancy and an increased risk of miscarriage is uncertain.*											
Footnotes: a. Downgraded one level due to high risk of bias; not limiting control group to women with a mental health disorder and not controlling for indication. b. Downgraded one level due to imprecision (95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25).											

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: CI, confidence interval; NA, not available; OBS, observational studies; RR, risk ratio.

[362] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns will be noted and may result in further downgrading of the quality of the evidence.

[363] Based on average risk from unexposed, depressed control groups of population-based cohort studies.

[364] Calculated by multiplying relative effect by control risk.

[365] Sorensen 2015

[366] Huybrechts 2016 unexposed patients with psychosis, schizophrenia, bipolar disorder.

D3.1.3 Anticonvulsants

The following section presents the Evidence Profile tables for the specific anticonvulsant medications examined. The consideration of these medications was limited to those used as mood stabilisers in women with mood disorders, and included sodium valproate, carbamazepine and lamotrigine. This is in line with the consideration of anticonvulsants by NICE 2015.

Extensive research on the effects of maternal use of anticonvulsants during pregnancy on infants has been carried out, and for this reason, the consideration of anticonvulsants has been limited to an examination of existing SRs only. However, all of this evidence has been conducted in a population with epilepsy, rather than a population with a mental health disorder. Where a comparison was made between an exposed population with epilepsy, and an unexposed population with epilepsy, no downgrading for indirectness was applied.

Regarding downgrading for risk of bias, one particular concern for the evidence available for anticonvulsants was that all included meta-analyses analysed the raw data from the included studies; thus, potential confounding was not minimised. However, a decision was made to not downgrade due to risk of bias due to lack of adjustment for confounding where there was a large magnitude of effect; i.e. where the RR lower 95% CI was > 1.25, which is the limit of appreciable harm used by NICE 2015. The rationale for this decision is that while not adjusted for potential confounders, the evidence for anticonvulsants is based on a large number of studies, is highly homogenous, and, being based on patients with epilepsy rather than a psychiatric disorder, is not likely to be subject to the same level of confounding by indication.

As baseline risk was not available in a pregnant unexposed population with a mental health disorder, where available the baseline risks identified for a depressed population were used as a proxy.

A summary of the characteristics of the individual included studies can be found in **Table AppD2-12** in **Appendix D2.1.3.1**. A detailed discussion of the evidence for each intervention and outcome can be found in **Appendix D4.1.3**.

Table D 3-26 presents a summary of the results of the Evidence Review of anticonvulsants and the location of the detailed assessment of the certainty of evidence in the evidence profile tables. Although the certainty of the evidence was very low to low, the results for sodium valproate strongly suggest that maternal exposure during pregnancy is associated with major and cardiac malformations and a reduction in IQ. In addition, the risk was greater following exposure to sodium valproate compared with carbamazepine and lamotrigine. Carbamazepine was also associated with major malformation, and the risk was greater for carbamazepine compared with lamotrigine; there appeared to be no increased risk associated with IQ. The evidence was inadequate for the assessment of maternal exposure to lamotrigine and all outcomes assessed, although as noted above, it was shown to present a lower risk than sodium valproate for major and cardiac malformations, and reduction in IQ, and a lower risk than carbamazepine for major malformation.

Table D 3-26 Summary of results of the Evidence Review for anticonvulsants

Intervention	**Increased/may be increased risk of harm** Outcome Certainty of evidence	**Appears to be no increased risk of harm** Outcome Certainty of evidence	**Decreased/may be decreased risk of harm** Outcome Certainty of evidence	**Uncertain** Outcome ○○○○	**Evidence profile table**
Sodium valproate	Major malformation ●●○○ Major malformation (vs carbamazepine) ●●○○ Major malformation (vs lamotrigine) ●●○○ Cardiac malformation ●●○○ Cardiac malformation (vs carbamazepine) ●●○○ Cardiac malformation (vs lamotrigine) ●●○○ IQ ●●○○ IQ (vs carbamazepine) ●○○○ IQ (vs lamotrigine) ●○○○			Neonatal mortality Preterm birth ASD	**Table D3-27**
Carbamazepine	Major malformation ●○○○ Major malformation (vs lamotrigine) ●○○○	IQ ●○○○		Cardiac malformation Cardiac malformation (vs lamotrigine) Neonatal mortality Preterm birth ASD IQ (vs lamotrigine)	**Table D3-28**
Lamotrigine				Major malformation Cardiac malformation Neonatal mortality Preterm birth ASD IQ	**Table D3-29**

Abbreviations: ASD, autism spectrum disorder; IQ, intelligence quotient.

Note: All comparisons are against non-exposure, unless otherwise stated. Certainty of evidence gradings are as follows: ●●●● – high certainty; ●●●○ – moderate certainty; ●●○○ – low certainty; ●○○○ – very low certainty; ○○○○ – inadequate certainty.

Table D3-27 Evidence Profile table: sodium valproate harms

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* (No. studies)	Additional risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall certainty of evidence	Study event rates		Risk estimate (95% CI)	Anticipated absolute effects	
							With control	With intervention		Risk with control[367]	Risk with intervention[368]
Major malformations: see Section AppD4.1.3.1.1											
3,182 (14 – OBS)[369]	None[370]	None	None	None	None	●●○○ Low	**Unexposed** NA	**Sodium valproate** NA	**RR 3.13** **(2.16, 4.54)**	28 per 1000[371]	88 per 1000 (73, 127)
7,078 (25 – OBS)[372]	None[370]	None	None	None	None	●●○○ Low	**Carbamazepine** NA	**Sodium valproate** NA	**RR 2.44** **(2.00, 2.94)[373]**	42 per 1000[374]	102 per 1000 (84, 123)
6,185 (7–OBS)[375]	None[370]	None	None	None	None	●●○○ Low	**Lamotrigine** NA	**Sodium valproate** NA	**RR 3.56** **(2.77, 4.58)**	Unknown[376]	Not estimable
Evidence Statements: *Maternal use of sodium valproate during pregnancy is associated with an increased risk of major malformation in the newborn, from an absolute risk of 3% to 9% (very low certainty evidence)* *Maternal use of sodium valproate during pregnancy is associated with an increased risk of major malformation in the newborn, when compared with maternal use of carbamazepine during pregnancy, from an absolute risk of 4% to 10% (very low certainty evidence)* *Maternal use of sodium valproate during pregnancy is associated with an increased risk of major malformation in the newborn, when compared with maternal use of lamotrigine during pregnancy (absolute increase in risk not estimable) (very low certainty evidence)*											
Cardiac malformations: see Section AppD4.1.3.2.1											
768 (6 – OBS)[377]	None[378]	None	None	None	None	●●○○ Low	**Unexposed** NA	**Sodium valproate** NA	**RR 4.85** **(1.28, 18.47)**	6 per 1000[379]	29 per 1000 (8, 111)

[367] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[368] Calculated by multiplying relative effect by control risk.
[369] Weston 2016 SR (includes Al Bunyan 1999, Campbell 2014, Canger 1999, Fairgrieve 2000, Garza-Morales 1996, Kaaja 2003, Kaneko 1999, Kelly 1984, Koch 1992, Lindhout 1992, Mawer 2010, Oguni 1992, Thomas 2008 and Vajda 2012).
[370] Based on the large magnitude of effect upper 95% CI > RR 1.25), the lack of downgrading for any other reason and the minimal effect of confounding by indication, the evidence will not be downgraded one level due to lack of adjustment for confounding. Also, not downgraded due to consideration of malformations in live births only (not explicitly stated in Weston 2016 but likely to be the case) because inclusion of live births would result in underreporting and likely underestimating the effect, and there is already a strong risk shown here.
[371] Ban 2014a (baseline risk from a population with depression/anxiety).
[372] Weston 2016 (includes Al Bunyan 1999, Arulmozhi 2006, Campbell 2014, Canger 1999, Cassina 2013, Eroglu 2008, Fairgrieve 2000, Froscher 1991, Garza=Morales 1996, Hernandez-Diaz 2012, Kaaja 2003, Kaneko 1999, Koch 1992, Lindhout 1992, Martinez Ferri 2009, Mawer 2010, Meador 2006, Meischenguiser 2004, Ogani 1992, Omtzigt 1992, Pardi 1982, Steegers-Theunissen 1994, Tanganelli 1992, Thomas 2008 and Vajda 2012.
[372] Weston 2016 (includes Campbell 2013, Mawer 2010 and Vajda 2012).
[373] Calculated from the analysis of carbamazepine versus sodium valproate (RR 0.41; 0.34, 0.50).
[374] Calculated from baseline risk with carbamazepine; see **Table D3-28**.
[375] Weston 2016 SR (includes Campbell 2014, Cassina 2013, Hernandez-Diaz 2012, Matrinez Ferri 2009, Mawer 2010, Meador 2006 and Vajda 2012).
[376] Not estimable; see **Table D3-29**
[377] Weston 2016 SR (includes Canger 1999, Fairgrieve 2000, Garza-Morales 1996, Koch 1992, Mawer 2010 and Vajda 2012).
[378] Based on the large magnitude of effect (upper 95% CI > RR 1.25), the lack of downgrading for any other reason and the minimal effect of confounding by indication, the evidence will not be downgraded one level due to lack of adjustment for confounding. Also, not downgraded due to consideration of malformations in live births only (not explicitly stated in Weston 2016 but likely to be the case) because inclusion of live births would result in underreporting and likely underestimating the effect, and there is already a strong risk shown here.
[379] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013 (baseline risk from a population with depression/anxiety).

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Study event rates**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							With control	**With intervention**		**Risk with control[367]**	**Risk with intervention[368]**
6,476 (16 – OBS)[380]	None[378]	None	None	None	None	●●○○ Low	**Carbamazepine** NA	**Sodium valproate** NA	**RR 2.22 (1.47, 3.03)[381]**	Unknown[382]	-
6,151 (6–OBS)[383]	None[378]	None	None	None	None	●●○○ Low	**Lamotrigine** NA	**Sodium valproate** NA	**RR 4.07 (2.33, 7.09)**	Unknown[384]	-
Evidence Statements: *Maternal use of sodium valproate during pregnancy is associated with an increased risk of cardiac malformation in the newborn, from an absolute risk of 0.6% to 3.0% (very low certainty evidence)* *Maternal use of sodium valproate during pregnancy is associated with an increased risk of cardiac malformation in the newborn, when compared with maternal use of carbamazepine during pregnancy (absolute increase in risk not estimable) (very low certainty evidence)* *Maternal use of sodium valproate during pregnancy is associated with an increased risk of cardiac malformation in the newborn, when compared with maternal use of lamotrigine during pregnancy (absolute increase in risk not estimable) (very low certainty evidence)*											
Neonatal mortality: see Section AppD4.1.3.3.1											
3,975 (2 – OBS)[385]	Serious(a)	None	Serious(b)	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Sodium valproate** NA	OR 1.93 (0.79, 4.7)	5 per 1000[386]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of sodium valproate during pregnancy and neonatal mortality is uncertain.*											
Preterm birth: see Section AppD4.1.3.4.1											
3,804 (2 – OBS)[387]	Serious(a)	None	Serious(b)	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Sodium valproate** NA	RR 1.31 (0.94, 1.83)	60 per 1000[388]	-
Evidence Statement: *Due to the inadequate quality of the evidence, any association between maternal use of sodium valproate during pregnancy and preterm birth is uncertain.*											

[380] Weston 2016 SR (includes Campbell 2014, Canger 1999, Cassina 2013, Eroglu 2008, Fairgrieve 2000, Froscher 1991, Hernandez-Diaz 2012, Kaaja 2003, Koch 1992, Martinez Ferri 2009, Meador 2006, Meischenguiser 2004, Omtzigt 1992, Pardi 1982, Thomas 2008 and Vajda 2012).
[381] Calculated from the analysis of carbamazepine versus sodium valproate (RR 0.45; 0.31, 0.68).
[382] Not calculable; see **Table D3-28**.
[383] Weston 2016 SR (includes Campbell 2014, Cassina 2013, Hernandez-Diaz 2012, Matrinez Ferri 2009, Meador 2006 and Vajda 2012).
[384] Not calculable, see **Table D3-29.**
[385] NICE 2015 SR (includes Artama 2013 and Diav-Citrin 2001).
[386] Ban 2012 (baseline risk from a population with depression/anxiety).
[387] NICE 215 SR (includes Artama 2013 and Diav-Citrin 2001).
[388] Malm 2015 (baseline risk from a population with depression/anxiety).

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Study event rates**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							With control	**With intervention**		**Risk with control[367]**	**Risk with intervention[368]**
Autism spectrum disorder: see Section AppD4.1.3.6.1											
655,495 (1 – OBS)[389]	Serious(a)	NA	Serious(b)	None	None	○○○○ Inadequate	**Unexposed** NA	**Sodium valproate** NA	**RR3.82 (2.15, 6.80)**	9 per 1000[390]	34 per 1000 (19,
Due to the inadequate certainty of the evidence, any association between maternal use of sodium valproate during pregnancy and autism spectrum disorder is uncertain.											
Autism checklist: see Section AppD4.1.3.6.1											
246 (1 – OBS)[391]	Serious(a)	NA	Serious(b)	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Sodium valproate** NA	RR 0.87 (0.19, 3.98)	Unknown	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of sodium valproate during pregnancy and autism (as measured by the Modified Checklist for Autism in Toddlers) is uncertain.*											
IQ: see Section AppD4.1.3.7.1											
Full scale IQ - < 1 SD 76 (2 – OBS)[392]	None[393]	None	None	None	None	●●○○ Low	**Unexposed** NA	**Sodium valproate** NA	**RR 10.33 (2.05, 52.01)**	Unknown	-
Full scale IQ 176 (4 – OBS)[394]	Serious(a)	Serious(d)	None	None	None	○○○○ Inadequate	**Unexposed** NA	**Sodium valproate** NA	**MD -8.17 (-12.80, -3.55)**	Unknown	-
Verbal IQ 160 (3 – OBS)[395]	Serious(a)	None	None	None	None	●○○○ Very low	**Unexposed** NA	**Sodium valproate** NA	**-MD -8.81 (-13.32, -4.30)[396]**	Unknown	-
Performance IQ 160 (3 – OBS)[397]	Serious(a)	None	None	None	None	●○○○ Very low	**Unexposed** NA	**Sodium valproate** NA	**MD -7.20 (-12.44, -1.96)[398]**	Unknown	-

[389] NICE 2015 SR (includes Christensen 2013).
[390] Sørensen 2013 and Malm 2016 (baseline risk from a population with depression/anxiety).
[391] NICE 2015 SR (includes Veiby 2013).
[392] Bromley 2014 SR (includes Bromley 2010 and Eriksson 2005).
[393] Based on the large magnitude of effect (lower 95% CI > RR 1.25 or upper 95% CI < 0.5), the lack of downgrading for any other reason and the minimal effect of confounding by indication, the evidence will not be downgraded one level due to lack of adjustment for confounding.
[394] Bromley 2014 SR (includes Bromley 2010, Thomas 2007, Eriksson 2005 and Gaily 2004).
[395] Bromley 2014 SR (includes Bromley 2010, Eriksson 2005 and Gaily 2004).
[396] Corresponds to a SMD -0.64 (-0.98, -0.29).
[397] Bromley 2014 SR (includes Bromley 2010, Eriksson 2005 and Gaily 2004).
[398] Corresponds to a SMD -0.46 (-0.81, -0.12).

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Study event rates**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							With control	**With intervention**		**Risk with control[367]**	**Risk with intervention[368]**
Full scale IQ - > 1 SD 178 (3 – OBS)[399]	Serious(a)	None	None	None	None	●○○○ Very low	**Carbamazepine** NA	**Sodium valproate** NA	**RR 2.5** **(1.20, 5.26)[400]**	Unknown	-
Full scale IQ 303 (5 – OBS)[401]	Serious(a)	Serious(d)	None	None	None	○○○○ Inadequate	**Carbamazepine** NA	**Sodium valproate** NA	**MD -8.69** **(-11.87, -5.51)[402]**	Unknown	-
Verbal IQ 226 (3 – OBS)[403]	Serious(a)	None	None	None	None	●○○○ Very low	**Carbamazepine** NA	**Sodium valproate** NA	**MD -8.44** **(-12.66, -4.21)[404]**	Unknown	-
Performance IQ 226 (3 – OBS)[405]	Serious(a)	None	None	None	None	●○○○ Very low	**Carbamazepine** NA	**Sodium valproate** NA	**MD -10.48** **(-14.94, -6.02)[406]**	Unknown	-
Full scale IQ - > 1 SD 157 (2 – OBS)[407]	None[408]	None	None	None	None	●●○○ Low	**Lamotrigine** NA	**Sodium valproate** NA	**RR 4.87** **(1.50, 15.78)**	Unknown	-
Full scale IQ 158 (2 – OBS)[409]	None[410]	None	None	None	None	●●○○ Low	**Lamotrigine** NA	**Sodium valproate** NA	**MD -10.80** **(-14.42, -7.17)[411]**	Unknown	-

Evidence Statements:

Maternal use of sodium valproate during pregnancy is associated with an increased risk of below average IQ (full-scale IQ score at 1 SD level) in the child (low certainty evidence)

Due to the inadequate certainty of the evidence, any association between maternal use of sodium valproate during pregnancy and full-scale IQ score in the child is uncertain.

[399] Bromley 2014 SR (includes Bromley 2010, Eriksson 2005 and Meador 2013).
[400] Calculated from the analysis of carbamazepine versus sodium valproate (RR 0.40; 0.19, 0.83).
[401] Bromley 2014 SR (includes Bromley 2010, Eriksson 2005, Gaily 2014, Meador 2013 and Thomas 2007).
[402] Calculated from the analysis of carbamazepine versus sodium valproate (MD 8.69; 5.51, 11.87).
[403] Bromley 2014 SR (includes Bromley 2010, Eriksson 2005 and Gaily 2004).
[404] Calculated from the analysis of carbamazepine versus sodium valproate (MD 8.44; 4.21, 12.66). Corresponds to a SMD -0.56 (-0.86, -0.26).
[405] Bromley 2014 SR (includes Bromley 2010, Eriksson 2005 and Gaily 2004).
[406] Calculated from the analysis of carbamazepine versus sodium valproate (MD 10.48; 6.02, 14.94). Corresponds to a SMD -0.71 (-1.02, -0.40).
[407] Bromley 2014 SR (includes Bromley 2010 and Meador 2013).
[408] Based on the large magnitude of effect (upper 95% CI > RR 1.25), the lack of downgrading for any other reason and the minimal effect of confounding by indication, the evidence will not be downgraded one level due to lack of adjustment for confounding.
[409] Bromley 2014 SR (includes Bromley 2010 and Meador 2013).
[410] Based on the large magnitude of effect (lower 95% CI < SMD -0.5), the lack of downgrading for any other reason and the minimal effect of confounding by indication, the evidence will not be downgraded one level due to lack of adjustment for confounding.
[411] Corresponds to SMD -0.92 (-1.26, -0.58).

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Study event rates**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							With control	**With intervention**		**Risk with control[367]**	**Risk with intervention[368]**
Maternal use of sodium valproate during pregnancy may be associated with a reduction in mean verbal IQ score in the child (very low certainty evidence)											
Maternal use of sodium valproate during pregnancy may be associated with a reduction in mean performance IQ score in the child (very low certainty evidence)											
Maternal use of sodium valproate during pregnancy may be associated with an increased risk of below average IQ (at 1 SD level in the child), compared with maternal use of carbamazepine during pregnancy (very low certainty evidence)											
Due to the inadequate certainty of the evidence, any additional reduction in full-scale IQ score in the child that may be associated with maternal use of sodium valproate during pregnancy, compared with maternal use of carbamazepine during pregnancy, is uncertain.											
Maternal use of sodium valproate during pregnancy may be associated with a reduction in mean verbal IQ score in the child, compared with maternal use of carbamazepine during pregnancy (very low certainty evidence)											
Maternal use of sodium valproate during pregnancy may be associated with a reduction in mean performance IQ score in the child, compared with maternal use of carbamazepine during pregnancy (very low certainty evidence)											
Maternal use of sodium valproate during pregnancy is associated with an increased risk of below average IQ (full-scale IQ score at 1 SD level) in the child, compared with maternal use of lamotrigine during pregnancy (low certainty evidence)											
Maternal use of sodium valproate during pregnancy is associated with a reduction in mean full-scale IQ score in the child, compared with maternal use of lamotrigine during pregnancy (low certainty evidence)											

Footnotes:

a. Downgraded one level due to a moderate risk of bias; analysis of raw data from observational studies.

b. Downgraded one level due to serious risk of indirectness; comparison with a general population.

c. Downgraded one level due to imprecision; 95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25 or SMD –0.5/0.5, no measure of precision available, or no events.

d. Downgraded one level due to serious heterogeneity; I^2 between 25% and 59%.

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: CI, confidence interval; IQ, intelligence quotient; MD, mean difference; NA, not available; OBS, observational studies; OR, odds ratio; RR, relative risk

Table D3-28 Evidence Profile table: carbamazepine harms

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* (No. studies)	Additional risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall certainty of evidence	Study event rates		Risk estimate (95% CI)	Anticipated absolute effects	
							With control	With intervention		Risk with control[412]	Risk with intervention[413]
Major malformations: see Section AppD4.1.3.1.1											
4,345 (17– OBS)[414]	Serious(a)[415]	None	None	None	None	●○○○ Very low	**Unexposed** NA	**Carbamazepine** NA	**RR 1.50 (1.03, 2.19)**	28 per 1000[416]	42 per 1000 (29, 61)
7,549 (7–OBS)[417]	Serious(a)[415]	None	None	None	None	●○○○ Very low	**Lamotrigine** NA	**Carbamazepine** NA	**RR 1.34 (1.01, 1.76)**	Unknown[418]	40 per 1000 (30, 53)
Evidence Statements: *Maternal use of carbamazepine during pregnancy may be associated with an increased risk of major malformation in the newborn, from an absolute risk of 3% to 4% (very low certainty evidence)* *Maternal use of carbamazepine during pregnancy may be associated with an increased risk of major malformation in the newborn, compared with maternal use of lamotrigine during pregnancy, from an absolute risk of 3.0% to 4.0% (very low certainty evidence)*											
Cardiac malformations: see Section AppD4.1.3.2.1											
1,026 (7 – OBS)[419]	Serious(a)[420]	None	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**Carbamazepine** NA	RR 1.84 (0.32, 10.71)	6 per 1000[421]	-
7,509 (6–OBS)[422]	Serious(a)[420]	None	None	Serious(b)	None	○○○○ Inadequate	**Lamotrigine** NA	**Carbamazepine** NA	RR 1.57 (0.85, 2.89)	Unknown[423]	-
Evidence Statements: *Due to the inadequate certainty of the evidence, any association between maternal use of carbamazepine during pregnancy and cardiac malformation in the newborn is uncertain.* *Due to the inadequate certainty of the evidence, any additional risk of cardiac malformation in the newborn that may be associated with maternal use of carbamazepine during pregnancy, compared with maternal use of lamotrigine during pregnancy, is uncertain.*											

[412] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[413] Calculated by multiplying relative effect by control risk.
[414] Weston 2016 SR (includes Al Bunyan 1999, Campbell 2014, Canger 1999, D'Souza 1990, Delmis 1991, Fairgrieve 2000, Garza-Morales 1996, Kaaja 2003, Kaneko 1999, Koch 1992, Lindhout 1992, Mawer 2010, Oguni 1992, Thomas 2008, Vajda 2012, Waters 1994).
[415] Not downgraded due to consideration of malformations in live births only (not explicitly stated in Weston 2016 but likely to be the case) because inclusion of live births would result in underreporting and likely underestimating the effect, and there was already a statistically significant risk shown.
[416] Ban 2014a (baseline risk from a population with depression/anxiety).
[417] Weston 2016 SR (includes Campbell 2014, Cassina 2013, Hernandez-Diaz 2012, Martinez Ferri 2009, Mawer 2010, Meador 2006 and Vajda 2012).
[418] Not calculable; see **Table D3-29**
[419] Weston 2016 SR (includes Al Bunyan 1999, Barqawi 2005, Canger 1999, Fairgrieve 2000, Koch 1992 and Mawer 2010 and Vajda 2012).
[420] Not downgraded due to consideration of malformations in live births only (not explicitly stated in Weston 2016 but likely to be the case) because inclusion of live births would result in underreporting and likely underestimating the effect, and there was already a statistically significant risk shown.
[421] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013 (baseline risk from a population with depression/anxiety).
[422] Weston 2016 SR (includes Campbell 2014, Cassina 2013, Hernandez-Diaz 2012, Martinez Ferri 2009, Meador 2006, and Vajda 2012).
[423] Not calculable; see **Table D3-29**

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Study event rates**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							With control	**With intervention**		**Risk with control[412]**	**Risk with intervention[413]**
Neonatal mortality: see Section AppD4.1.3.3.1											
3,202 (2 – OBS)[424]	Serious(a)	Very serious(c)	Serious(d)	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**Carbamazepine** NA	OR 0.79 (0.12, 5.31)	5 per 1000[425]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of carbamazepine during pregnancy and neonatal mortality is uncertain.*											
Preterm birth: see Section AppD4.1.3.4.1											
3,202 (2 – OBS)[426]	Serious(a)	None	Serious(d)	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**Carbamazepine** NA	OR 1.65 (0.64, 4.22)	60 per 1000[427]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of carbamazepine during pregnancy and preterm birth is uncertain.*											
Autism spectrum disorder: see Section AppD4.1.3.5.1											
655,539 (1 – OBS)[428]	Serious(a)	NA	Serious(d)	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**Carbamazepine** NA	OR 1.25 (0.47, 3.35)	9 per 1000[429]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of carbamazepine during pregnancy and autism spectrum disorder is uncertain.*											
Autism checklist: see Section AppD4.1.3.6.1											
262 (1 – OBS)[430]	Serious(a)	NA	Serious(d)	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**Carbamazepine** NA	OR 0.79 (0.22, 2.8)	Unknown	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of carbamazepine during pregnancy and autism (as measured by the Modified Checklist for Autism in Toddlers) is uncertain.*											
IQ: see Section AppD4.1.3.7.1											
Full scale IQ 250 (4 – OBS)[431]	Serious(a)	None	None	None[432]	None	●○○○ Very low	**Unexposed** NA	**Carbamazepine** NA	MD 1.84 (-2.13, 5.80)	Unknown	-

[424] NICE 2015 SR (includes Artama 2013 and Diav-Citrin 2001).
[425] Ban 2012 (baseline risk from a population with depression/anxiety).
[426] NICE 215 SR (includes Artama 2013 and Diav-Citrin 2001).
[427] Malm 2015 (baseline risk from a population with depression/anxiety).
[428] NICE 2015 SR (includes Christensen 2013).
[429] Sørensen 2013 and Malm 2016 (baseline risk from a population with depression/anxiety).
[430] NICE 2015 SR (includes Veiby 2013).
[431] Bromley 2014 SR (includes Bromley 2010, Thomas 2007, Eriksson 2005 and Gaily 2004).
[432] Based on analysis conducted for this review; SMD 0.15 (95% CI -0.11, 0.41).

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Study event rates**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							With control	**With intervention**		**Risk with control[412]**	**Risk with intervention[413]**
Verbal IQ 232 (3 – OBS)[433]	Serious(a)	None	None	None[434]	None	●○○○ Very low	**Unexposed** NA	**Carbamazepine** NA	MD 0.13 (-3.98, 4.23)	Unknown	-
Performance IQ 232 (3 – OBS)[435]	Serious(a)	None	None	Serious(b)[436]	None	○○○○ Inadequate	**Unexposed** NA	**Carbamazepine** NA	MD 3.65 (-0.60, 7.90)	Unknown	-
Full scale IQ - > 1 SD 159 (2 – OBS)[437]	Serious(a)	None	None	Serious(b)	None	○○○○ Inadequate	**Lamotrigine** NA	**Carbamazepine** NA	RR 2.28 (0.63, 8.22)	Unknown	-
Full scale IQ 162 (2 – OBS)[438]	Serious(a)	None	None	None[439]	None	●○○○ Very low	**Lamotrigine** NA	**Carbamazepine** NA	MD -1.62 (-5.44, 2.21)	Unknown	-

Evidence Statements:

Maternal use of carbamazepine during pregnancy does not appear to be associated with a reduction in mean full-scale IQ score in the child (very low certainty evidence)

Maternal use of carbamazepine during pregnancy does not appear to be associated with a reduction in mean verbal IQ score in the child (very low certainty evidence)

Due to the inadequate certainty of the evidence, any association between maternal use of carbamazepine during pregnancy and mean performance IQ is uncertain.

Due to the inadequate certainty of the evidence, any additional risk of below average IQ (full-scale IQ score at 1 SD level) in the child that may be associated with maternal use of carbamazepine during pregnancy, compared with maternal use of lamotrigine during pregnancy, is uncertain.

Maternal use of carbamazepine during pregnancy does not appear to be associated with a reduction in mean full-scale IQ score in the child, compared with maternal use of lamotrigine during pregnancy (very low certainty evidence)

Footnotes:

a. Downgraded one level due to a moderate risk of bias; analysis of raw data from observational studies.

b. Downgraded one level due to imprecision; 95% CI crosses line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25 or SMD –0.5/0.5, no measure of precision available, or no events.

c. Downgraded two levels due to very serious heterogeneity; $I^2 \geq 60\%$.

d. Downgraded one due to serious risk of indirectness; comparison with a general population

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Those shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: CI, confidence interval; IQ, intelligence quotient; MD, mean difference; NA, not available; OBS, observational studies; OR, odds ratio; RR, relative risk

[433] Bromley 2014 SR (includes Bromley 2001, Eriksson 2005 and Gaily 2004).

[434] Based on analysis conducted for this review; SMD 0.02 (95% CI -0.25, 0.29).

[435] Bromley 2014 SR (includes Bromley 2010, Eriksson 2005 and Gaily 2004).

[436] Based on analysis conducted for this review; SMD 0.25 (95% CI -0.02, 0.52).

[437] Bromley 2014 SR (includes Bromley 2010 and Meador 2013).

[438] Bromley 2014 SR (includes Bromley 2010 and Meador 2013).

[439] Based on analysis conducted for this review; SMD -0.13 (95% CI -0.44, 0.18).

Table D3-29 Evidence Profile table: lamotrigine harms

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* (No. studies)	Additional risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall certainty of evidence	Study event rates		Risk estimate (95% CI)	Anticipated absolute effects	
							With control	With intervention		Risk with control[440]	Risk with intervention[441]
Major malformations: see Section AppD4.1.3.1.1											
3,181 (3– OBS)[442]	Serious(a,b)	None	None	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Lamotrigine** NA	RR 1.07 (0.64, 1.77)	28 per 1000[443]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of lamotrigine during pregnancy and major malformation in the newborn is uncertain.*											
Cardiac malformations: see Section AppD4.1.3.2.1											
542 (2 – OBS)[444]	Serious(a,b)	NA	None	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Lamotrigine** NA	RR 1.40 (0.15, 13.35)	6 per 1000[445]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of lamotrigine during pregnancy and cardiac malformation in the newborn is uncertain.*											
Neonatal mortality: see Section AppD4.1.3.3.1											
1,973 (1 – OBS)[446]	Serious(a)	NA	Serious(d)	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Lamotrigine** NA	RR 0.49 (0.03, 8.42)	5 per 1000[447]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of lamotrigine during pregnancy and neonatal mortality is uncertain.*											
Preterm birth: see Section AppD4.1.3.4.1											
1,973 (1 – OBS)[448]	Serious(a)	None	Serious(d)	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Lamotrigine** NA	RR 0.98 (0.47, 2.05)	60 per 1000[449]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of lamotrigine during pregnancy and preterm birth is uncertain.*											

[440] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[441] Calculated by multiplying relative effect by control risk.
[442] Weston 2016 SR (includes Campbell 2013, Mawer 2010 and Vajda 2012).
[443] Ban 2014a (baseline risk from a population with depression/anxiety).
[444] Weston 2016 SR (includes Mawer 2010 (no events) and Vajda 2012).
[445] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013 (baseline risk from a population with depression/anxiety).
[446] NICE 2015 SR (includes Artama 2013).
[447] Ban 2012 (baseline risk from a population with depression/anxiety).
[448] NICE 215 SR (includes Artama 2013 and Diav-Citrin 2001).
[449] Malm 2015 (baseline risk from a population with depression/anxiety).

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Study event rates**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							With control	**With intervention**		**Risk with control[440]**	**Risk with intervention[441]**
Autism spectrum disorder: see Section AppD4.1.3.5.1											
655,394 (1 – OBS)[450]	Serious(a)	NA	Serious(d)	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Lamotrigine** NA	RR 1.5 (0.75, 3.01)	Unknown	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of lamotrigine during pregnancy autism spectrum disorder in the child is uncertain.*											
Autism checklist: see Section AppD4.1.3.6.1											
286 (1 – OBS)[451]	Serious(a)	NA	Serious(d)	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Lamotrigine** NA	RR 1.83 (0.81, 4.13)	Unknown	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of lamotrigine during pregnancy and autism (as measured by the Modified Checklist for Autism in Toddlers) is uncertain.*											
IQ: see Section AppD4.1.3.7.1											
Full scale IQ 54 (1 – OBS)[452]	Serious(a)	None	None	Serious(c)[453]	None	○○○○ Inadequate	**Unexposed** NA	**Lamotrigine** NA	MD -1.0 (-7.48, 5.48)	Unknown	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of lamotrigine during pregnancy and reduction in full-scale IQ score in the child is uncertain.*											
Footnotes: a. Downgraded one level due to a moderate risk of bias; analysis of raw data from observational studies. b. Downgraded one level due to serious risk of bias; selection bias due to exclusion of planned abortions, miscarriages and stillborn from the analysis. c. Downgraded one level due to imprecision; 95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25 or SMD –0.5/0.5, no measure of precision available, or no events. d. Downgraded one level due to serious risk of indirectness; comparison with a general population.											

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: CI, confidence interval; IQ, intelligence quotient; MD, mean difference; NA, not available; OBS, observational studies; OR, odds ratio; RR, relative risk.

[450] NICE 2015 SR (includes Christensen 2013).
[451] NICE 2015 SR (includes Veiby 2013).
[452] Bromley 2014 SR (includes Bromley 2010).
[453] Based on analysis conducted for this review; SMD -0.08 (95% CI -0.62, 0.45).

D3.1.4 Benzodiazepines and z-drugs

The following section presents the Evidence Profile tables for benzodiazepine and z-drugs. No intervention-specific rules were required for downgrading of the certainty of this body of evidence.

As the evidence was based on data from cohort and case-control studies, in many cases the results were presented as odds ratios instead of relative risks. Where the baseline risk was < 7%, it was assumed that the odds ratio approximates the relative risk and the results were interpreted as relative risks. Where baseline risk was not available in a pregnant unexposed population with a mental health disorder, the baseline risks identified for a depressed population were used as a proxy.

A summary of the characteristics of the individual included studies can be found in **Table AppD2-18** in **Appendix D2.1.4.2**. A detailed discussion of the evidence for each group or individual intervention type and outcome can be found in **Appendix D4.1.4**.

Table D 3-30 presents a summary of the results of the Evidence Review of benzodiazepines and z-drugs and the location of the detailed assessment of the certainty of evidence in the evidence profile tables. The majority of the evidence assessed was of inadequate certainty, so the findings for most outcomes were considered uncertain. Exceptions to this were the associations between exposure in late pregnancy to benzodiazepines and respiratory difficult, and zolpidem and preterm birth and the infant being small for gestational age. The evidence suggests maternal exposure to benzodiazepines may not be associated with major malformation, and zolpidem may not be associated with respiratory difficulty.

Table D 3-30 Summary of results of the Evidence Review for benzodiazepines and z-drugs

Intervention	**Increased/may be increased risk of harm** Outcome Certainty of evidence	**Appears to be no increased risk of harm** Outcome Certainty of evidence	**Decreased/may be decreased risk of harm** Outcome Certainty of evidence	**Uncertain** Outcome ○○○○	**Evidence profile table**
Benzodiazepines ± z drugs	Respiratory difficulty[454] ●○○○	Major malformation ●○○○		Cardiac malformation Septal malformation Miscarriage Preterm birth SFGA Convulsions Language competence	**Table D3-31**
Diazepam				Major malformation Cardiac malformation	**Table D3-32**
Temazepam				Major malformation Cardiac malformation	**Table D3-33**
Z-drugs				Major malformation Cardiac malformation	**Table D3-34**
Zolpidem	Preterm birth ●●○○ SFGA ●●○○	Respiratory difficulty ●○○○		Major malformation	**Table D3-35**
Zopiclone				Major malformation Cardiac malformation Miscarriage Preterm birth SFGA	**Table D3-36**

Abbreviations: SFGA, small for gestational age.

Note: All comparisons are against non-exposure, unless otherwise stated. Certainty of evidence gradings are as follows: ●●●● – high certainty; ●●●○ – moderate certainty; ●●○○ – low certainty; ●○○○ – very low certainty; ○○○○ – inadequate certainty.

[454] Late exposure only.

Table D3-31 Evidence Profile table: benzodiazepines ± z-drugs

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* (No. studies)	Additional risk of bias[455]	Inconsistency	Indirectness	Imprecision	Publication bias	Overall certainty of evidence	Population (N) Unexposed	Population (N) Exposed	Risk estimate (95% CI)	Anticipated absolute effects Risk with control[456]	Anticipated absolute effects Risk with intervention[457]
Major malformations: see Section AppD4.1.4.1.2											
108,288 (1 – OBS)[458]	Serious(a)	NA	None	None	None	●○○○ Very low	**Unexposed** NA	**Benzodiazepines[459] (first trimester)** NA	RD -0.0041 (-0.015, 0.0069)	28 per 1000[460]	28 per 1000 (28, 28)
NR (1 – OBS)[461]	Serious(a)	NA	Serious(b)	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Benzodiazepines and z-drugs[462] –excluding anticonvulsants (any time)** NA	RR 1.22 (0.97, 1.52)	28 per 1000[463]	-
Evidence Statement: *Maternal use of benzodiazepines during the first trimester of pregnancy does not appear to be associated with an increased risk of major malformation in the newborn (very low certainty evidence)*											
Cardiac malformations: see Section AppD4.1.4.2.2											
4,467 (1 – OBS)[464]	Serious(d)	NA	None	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Benzodiazepines[465] (any time)** NA	RR 1.6 (0.9, 2.8)	6 per 1000[466]	-
4,467 (1 – OBS)[467]	Serious(d)	NA	None	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Benzodiazepines[468] (Month 1)** NA	RR 1.6 (0.7, 3.7)	6 per 1000[469]	-

[455] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns will be noted and may result in further downgrading of the quality of the evidence.
[456] Based on average risk from unexposed, depressed control groups of population-based cohort studies providing data for that outcome.
[457] Calculated by multiplying relative effect by control risk.
[458] Oberlander 2008a.
[459] Includes lorazepam (44.0%), clonazepam (21.4%), oxazepam (15.0%), alprazolam (6.8%), temazepam (5.1%), diazepam (5.0%) and others.
[460] Ban 2014a (depressed/anxious population).
[461] Wikner 2007.
[462] Of the 2,169 infant exposures in early pregnancy, 72.3% were to benzodiazepines and 27.7% were to z-drugs. Of the 415 infant exposures in late pregnancy, 82.2% were to benzodiazepines and 17.8% were to z-drugs.
[463] Ban 2014a (depressed/anxious population).
[464] Eros 2002.
[465] Includes nitrazepam, medazepam, tofisopam, alprazolam and clonazepam.
[466] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013 (depressed/anxious population).
[467] Eros 2002.
[468] Includes nitrazepam, medazepam, tofisopam, alprazolam and clonazepam.
[469] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013 (depressed/anxious population).

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[455]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[456]**	**Risk with intervention[457]**
4,467 (1 – OBS)[470]	Serious(d)	NA	None	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Benzodiazepines[471] (Months 2-3)** NA	RR 1.0 (0.2, 4.6)	6 per 1000[472]	-
4,467 (1 – OBS)[473]	Serious(d)	NA	None	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Benzodiazepines[474] (Months 4-9)** NA	RR 1.9 (0.8, 4.6)	6 per 1000[475]	-
4,467 (1 – OBS)[476]	Serious(d)	NA	None	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Benzodiazepines[477] (any time)** NA	OR 1.6 (0.7, 3.6)[478]	6 per 1000[479]	-
4,467 (1 – OBS)[480]	Serious(d)	NA	None	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Benzodiazepines[481] (Months 2-3)** NA	OR 5.0 (0.2, 104)[478]	6 per 1000[482]	-
108,288 (1 – OBS)[483]	Serious(d)	NA	None	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Benzodiazepine (any time)** NA	RD –0.0013 (-0.0055, 0.0029)	6 per 1000[484]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of benzodiazepines and an increased risk of cardiac malformation in the newborn is uncertain.*											

[470] Eros 2002.
[471] Includes nitrazepam, medazepam, tofisopam, alprazolam and clonazepam.
[472] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013 (depressed/anxious population).
[473] Eros 2002.
[474] Includes nitrazepam, medazepam, tofisopam, alprazolam and clonazepam.
[475] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013 (depressed/anxious population).
[476] Eros 2002.
[477] Includes nitrazepam, medazepam, tofisopam, alprazolam and clonazepam.
[478] McNemar analysis.
[479] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013 (depressed/anxious population).
[480] Eros 2002.
[481] Includes nitrazepam, medazepam, tofisopam, alprazolam and clonazepam.
[482] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013 (depressed/anxious population).
[483] Oberlander 2008a.
[484] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013 (depressed/anxious population).

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[455]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[456]**	**Risk with intervention[457]**
Septal malformations: see Section AppD4.1.4.3.2											
108,288 (1 – OBS)[485]	Very serious(e)	NA	Serious(b)	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Benzodiazepines** NA	RR 1.48 (0.21, 10.65)	3 per 1000[486]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of benzodiazepines and an increased risk of septal malformation in the newborn is uncertain.*											
Miscarriage: see Section AppD4.1.4.4.1											
1,204 (3 – OBS)[487]	Serious(f)	None	Serious(b)	None	None	○○○○ Inadequate	**Unexposed** NA	**Benzodiazepines** NA	**OR 1.83 (1.19, 2.82)**	81 per 1000[488]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of benzodiazepines and an increased risk of miscarriage is uncertain.*											
Preterm birth: see Section AppD4.1.4.5.2											
42,875 (1 – OBS)[489]	Serious(g)	NA	Serious(b)	None	None	○○○○ Inadequate	**Unexposed** NA	**Benzodiazepines or z-drugs[490] (early exposure)** NA	**RR 1.48 (1.26, 1.75)**	60 per 1000[491]	-
42,875 (1 – OBS)[492]	Serious(g)	NA	Serious(b)	None	None	○○○○ Inadequate	**Unexposed** NA	**Benzodiazepines or z-drugs[493] (late exposure)** NA	**RR 2.57 (1.92, 3.43)**	60 per 1000[491]	-
42,875 (1 – OBS)[494]	None	NA	Serious(b)	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Benzodiazepines or z-drugs – excluding antidepressants (any time)** NA	RR 1.20 (0.97, 1.50)	6 per 1000[491]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of benzodiazepines or z-drugs during pregnancy and an increased risk of preterm birth is uncertain.*											

[485] Based on results presented in NICE 2015 (includes Oberlander 2008a).
[486] The Bérard 2015 (examining antidepressants) study used an insured population and as such the prevalence of septal malformations in this study (1.83%) is not likely to be representative of the general population with depression. To estimate the prevalence, it is assumed that 50% of cardiac malformations are septal, resulting in an estimate of 0.3%.
[487] Based on results presented in NICE 2015 (includes Laegreid 1992, Ornoy 1998 and Pastuszak 1996).
[488] Almeida 2016 and Ban 2012 (depressed/anxious population).
[489] Wikner 2007.
[490] Of the 2,169 infant exposures in early pregnancy, 72.3% were to benzodiazepines and 27.7% were to z-drugs.
[491] Malm 2015 (depressed population).
[492] Wikner 2007.
[493] Of the 415 infants exposed in late pregnancy, 82.2% were exposed to benzodiazepines and 17.8% were exposed to z-drugs.
[494] Wikner 2007.

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[455]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[456]**	**Risk with intervention[457]**
Small for gestational age: see Section AppD4.1.4.6.2											
18,260 (1 – OBS)[495]	Serious(g)	NA	Serious(b)	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Benzodiazepines or z-drugs[496] (early exposure)** NA	OR 1.12 (0.87, 1.44)	Unknown	-
18,260 (1 – OBS)[497]	Serious(g)	NA	Serious(b)	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Benzodiazepines or z-drugs[498] (late exposure)** NA	OR 1.39 (0.80, 2.40)	Unknown	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of benzodiazepines or z-drugs during pregnancy and an increased risk of the newborn being small for gestational age is uncertain.*											
Respiratory difficulty: see Section AppD4.1.4.7.2											
38,638 (1 – OBS)[499]	None	NA	Serious(b)	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Benzodiazepines or z-drugs[500] (early exposure)** NA	RR 1.19 (0.98, 1.45)	32 per 1000[501]	-
38,638 (1 – OBS)[502]	None	NA	Serious(b)	None	None	●○○○ Very low	**Unexposed** NA	**Benzodiazepines or z-drugs[503] (late exposure)** NA	**RR 2.21 (1.62, 3.02)**	32 per 1000[504]	71 per 1000 (52, 97)
NR (1 – OBS)[505]	None	NA	Serious(b)	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Benzodiazepines or z-drugs[506] - excluding antidepressants (any time)** NA	RR 1.12 (0.88, 1.43)	32 per 1000[507]	-
Evidence Statement: *Maternal use of benzodiazepines during late pregnancy may be associated with an increased risk of respiratory difficulty in the newborn, from an absolute risk of 3.2% to 7% (very low certainty evidence)*											

[495] Wikner 2007.
[496] Of the 2,169 infant exposures in early pregnancy, 72.3% were to benzodiazepines and 27.7% were to z-drugs.
[497] Wikner 2007.
[498] Of the 415 infants exposed in late pregnancy, 82.2% were exposed to benzodiazepines and 17.8% were exposed to z-drugs.
[499] Wikner 2007
[500] Of the 2,169 infant exposures in early pregnancy, 72.3% were to benzodiazepines and 27.7% were to z-drugs.
[501] Malm 2015.
[502] Wikner 2007
[503] Of the 415 infant exposures in late pregnancy, 82.2% were to benzodiazepines and 17.8% were to z-drugs.
[504] Malm 2015.
[505] Wikner 2007
[506] Of the 2,169 infant exposures in early pregnancy, 72.3% were to benzodiazepines and 27.7% were to z-drugs. Of the 415 exposures in late pregnancy, 82.2% were to benzodiazepines and 17.8% were to z-drugs.
[507] Malm 2015.

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[455]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[456]**	**Risk with intervention[457]**
Convulsions: see Section AppD4.1.4.8.2											
1,386 (1 – OBS)[508]	Serious(g)	NA	Serious(b)	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Benzo or z-drug (early exposure)** NA	RR 1.35 (0.44, 3.15)	Unknown	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of benzodiazepines or z-drugs during early pregnancy and an increased risk of convulsions in the newborn is uncertain.*											
Language competence: see Section AppD4.1.4.9.2											
51,411 (1 – OBS)[509]	Serious(h)	NA	None	Unknown(c)	None	○○○○ Inadequate	**Unexposed** NA	**Benzo or z-drug (short-term use)[510]** NA	OR 1.0 (0.7, 1.3)	Unknown	-
51,174 (1 – OBS)[511]	Serious(h)	NA	None	Unknown(c)	None	○○○○ Inadequate	**Unexposed** NA	**Benzo or z-drug (long-term use)[512]** NA	OR 1.3 (0.8, 2.3)	Unknown	-
Evidence Statement: Due to the inadequate certainty of the evidence, any association between maternal use of benzodiazepines or z-drugs at any time during pregnancy and an increased risk of decreased language competence in the child is uncertain.											

Footnotes:
a. Downgraded one level due to serious risk of bias; selection bias due to exclusion of planned abortions, miscarriages and stillborn from the analysis.
b. Downgraded one level due to serious indirectness; compared to a general population with no adjustment for potential confounding by indication.
c. Downgraded one level due to imprecision; 95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25 or SMD –0.5/0.5, no measure of precision available, or no events.
d. Downgraded one level due to serious risk of bias; selection bias due to exclusion of miscarriages from the analysis.
e. Downgraded two levels due to very serious risk of bias; analysis based on raw data and potential for selection bias due to exclusion of planned abortions, miscarriages and stillborn from the analysis.
f. Downgraded one level due to serious risk of bias; analysis based on raw data.
g. Downgraded one level due to moderate risk of bias; no adjustment for/consideration of other treatments.
h. Downgraded one level due to moderate risk of bias; self-reported exposure and outcome.

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.
Abbreviations: Benzo, benzodiazepine; CI, confidence interval; NA, not available; NR, not reported; OBS, observational studies; OR, odds ratio; RD, risk difference; RE, risk estimate; RR, relative risk; SMD, standardised mean difference.

[508] Wikner 2007.
[509] Odsbu 2015.
[510] Woman reported use on one questionnaire during pregnancy only. Women answered three questionnaires during pregnancy.
[511] Odsbu 2015.
[512] Woman reported use on more than one questionnaire during pregnancy. Women answered three questionnaires during pregnancy.

Table D3-32 Evidence Profile table: diazepam

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[513]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[514]**	**Risk with intervention[515]**
Major malformation: see Section AppD4.1.4.1.2											
Heart anomalies *20,352* (1 – OBS)[516]	Serious(a)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**Diazepam (first trimester)** NA	RR 0.99 (0.61, 1.61)	28 per 1000[517]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of diazepam during the first trimester of pregnancy and major malformation in the newborn is uncertain.*											
Cardiac malformation: see Section AppD4.1.4.2.2											
Heart anomalies *20,532* (1 – OBS)[518]	Serious(a)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**Diazepam (first trimester)** NA	OR 1.29 (0.60, 2.80)	6 per 1000[519]	-
Cardiovascular congenital anomalies 42,630 (1 – OBS)[520]	Serious(c)	NA	None	Serious(b)	None	○○○○ Inadequate	**Diazepam (Months 5-6)** NA	**Diazepam (Months 2-3)** NA	OR 1.0 (0.8, 1.4)	6 per 1000[521]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of diazepam during the first trimester of pregnancy and cardiac malformation in the newborn is uncertain.*											

Footnotes:
a. Downgraded one level due to moderate risk of bias; potential for selection bias due to exclusion of planned abortions, miscarriages and stillborn from the analysis.
b. Downgraded one level due to imprecision; 95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25 or SMD –0.5/0.5, no measure of precision available, or no events.
c. Downgraded one level due to moderate risk of bias; potential for selection bias due to exclusion of miscarriages and stillborn from the analysis.

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.
Abbreviations: CI, confidence interval; NA, not available; NR, not reported; OBS, observational studies; OR, odds ratio; RE, risk estimate; RR, relative risk; SMD, standardised mean difference.

[513] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns will be noted and may result in further downgrading of the quality of the evidence.
[514] Based on average risk from unexposed, depressed control groups of population-based cohort studies providing data for that outcome.
[515] Calculated by multiplying relative effect by control risk.
[516] Ban 2014b.
[517] Ban 2014a (depressed/anxious population).
[518] Ban 2014b.
[519] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013 (depressed/anxious population).
[520] Kjaer 2007.
[521] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013 (depressed/anxious population).

Table D3-33 Evidence Profile table: temazepam

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias**[522]	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control**[523]	**Risk with intervention**[524]
Major malformation: see Section AppD4.1.4.1.2											
Heart anomalies *19,572* (1 – OBS)[525]	Serious(a)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**Temazepam (first trimester)** NA	OR 1.04 (0.47, 2.32)	28 per 1000[526]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of temazepam during the first trimester of pregnancy and major malformation in the newborn is uncertain.*											
Cardiac malformation: see Section AppD4.1.4.2.2											
Heart anomalies *19,572* (1 – OBS)[527]	Serious(a)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**Temazepam (first trimester)** NA	OR 1.31 (0.35, 4.92)	6 per 1000[528]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of temazepam during the first trimester of pregnancy and cardiac malformation in the newborn is uncertain.*											

Footnotes:
a. Downgraded one level due to moderate risk of bias; potential for selection bias due to exclusion of planned abortions, miscarriages and stillborn from the analysis.
b. Downgraded one level due to imprecision; 95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25 or SMD –0.5/0.5, no measure of precision available, or no events.

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.
Abbreviations: CI, confidence interval; NA, not available; NR, not reported; OBS, observational studies; OR, odds ratio; RE, risk estimate; RR, relative risk; SMD, standardised mean difference.

[522] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns will be noted and may result in further downgrading of the quality of the evidence.
[523] Based on average risk from unexposed, depressed control groups of population-based cohort studies providing data for that outcome.
[524] Calculated by multiplying relative effect by control risk.
[525] Ban 2014b.
[526] Ban 2014a (depressed/anxious population).
[527] Ban 2014b.
[528] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013 (depressed/anxious population).

Table D3-34 Evidence Profile table: z-drugs

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[529]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[530]**	**Risk with intervention[531]**
Major malformation: see Section AppD4.1.4.1.2											
1,127,075 (1 – OBS)[532]	Very serious(a)	NA	Serious(b)	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Z-drugs (any time)** NA	RR 0.95 (0.69, 1.30)	28 per 1000[533]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of z-drugs at any time during pregnancy and relatively severe malformation[534] in the newborn is uncertain.*											
Cardiac malformation: see Section AppD4.1.4.2.2											
1,127,075 (1 – OBS)[535]	Very serious(a)	NA	Serious(b)	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Z-drugs (any time)** NA	RR 0.55 (0.27, 1.09)	6 per 1000[536]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of z-drugs at any time during pregnancy and cardiac malformation in the newborn is uncertain.*											

Footnotes:
a. Downgraded two levels due to high risk of bias; potential for selection bias due to exclusion of planned abortions, miscarriages and stillborn from the analysis and lack of adjustment for use of other treatments.
b. Downgraded one level due to serious indirectness; compared to a general population with no adjustment for potential confounding by indication.
c. Downgraded one level due to imprecision; 95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25 or SMD –0.5/0.5, no measure of precision available, or no events.

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.
Abbreviations: CI, confidence interval; NA, not available; OBS, observational studies; RE, risk estimate; RR, relative risk; SMD, standardised mean difference.

[529] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns will be noted and may result in further downgrading of the quality of the evidence.
[530] Based on average risk from unexposed, depressed control groups of population-based cohort studies providing data for that outcome.
[531] Calculated by multiplying relative effect by control risk.
[532] Wikner 2011.
[533] Ban 2014a (depressed/anxious population).
[534] Excludes preauricular appendix, undescended testicle, unstable hip, patent ductus arteriosus in preterm infants, single umbilical artery, tongue tie and nevus.
[535] Wikner 2011.
[536] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013 (depressed/anxious population).

Table D3-35 Evidence Profile table: zolpidem

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[537]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[538]**	**Risk with intervention[539]**
Major malformation: see Section AppD4.1.4.1.2											
14,982 (1 – OBS)[540]	Serious(a)	NA	Serious(b)	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Zolpidem (any time)** NA	RR 0.70 (0.38, 1.28)	28 per 1000[541]	-
14,447 *(1 – OBS)[542]*	Serious(a)	NA	Serious(b)	Serious(c)	None	○○○○ Inadequate	**Unexposed** NA	**Zolpidem (second or third trimester)** NA	RR 0.74 (0.38, 1.44)	28 per 1000[543]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of zolpidem at any time during pregnancy and major malformation[544] in the neonate is uncertain.*											
Preterm birth: see Section AppD4.1.4.5.2											
14,982 (1 - OBS)[545]	None	NA	None[546]	None	None	●●○○ Low	**Unexposed** NA	**Zolpidem (any time)** NA	**RR 1.49 (1.28, 1.74)**	60 per 1000[547]	89 per 1000 (77, 104)
13,020 (1 – OBS)[548]	None	NA	None[546]	None	None	●●○○ Low	**Unexposed** NA	**Zolpidem (first trimester)** NA	**RR 1.48 (1.10, 1.98)**	60 per 1000[547]	89 per 1000 (66, 119)

[537] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns will be noted and may result in further downgrading of the quality of the evidence.
[538] Based on average risk from unexposed, depressed control groups of population-based cohort studies providing data for that outcome.
[539] Calculated by multiplying relative effect by control risk.
[540] Wang 2010.
[541] Ban 2014a (in a depressed/anxious population).
[542] Wang 2010.
[543] Ban 2014a (in a depressed/anxious population).
[544] Limited to hydrocephaly, anencephaly, microcephaly, meningomyelocele, encephalocele and spina bifida.
[545] Wang 2010.
[546] Compared exposure in a non-mental health disorder population with non-exposure in a non-mental health disorder population.
[547] Malm 2015 (depressed population).
[548] Wang 2010.

Certainty assessment							**Summary of findings**				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[537]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)** *P value*	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[538]**	**Risk with intervention[539]**
14,447 (1 – OBS)[549]	None	NA	None[546]	None	None	●●○○ Low	**Unexposed** NA	**Zolpidem (second or third trimester)** NA	**OR 1.49 (1.26, 1.77)**	Unknown	-
Evidence Statement: *Maternal use of zolpidem at any time during pregnancy is associated with an increased risk of preterm birth, from an absolute risk of 6% to 9% (low certainty evidence)*											
Small for gestational age: see Section AppD4.1.4.6.2											
14,982 (1 - OBS)[550]	None	NA	None[551]	None	None	●●○○ Low	**Unexposed** NA	**Zolpidem (any time)** NA	**OR 1.34 (1.20, 1.49)**	Unknown	-
13,020 (1 – OBS)[552]	None	NA	None[551]	None	None	●●○○ Low	**Unexposed** NA	**Zolpidem (first trimester)** NA	**OR 1.36 (1.09, 1.69)**	Unknown	-
14,447 (1 – OBS)[553]	None	NA	None[551]	None	None	●●○○ Low	**Unexposed** NA	**Zolpidem (second or third trimester)** NA	**OR 1.33 (1.18, 1.50)**	Unknown	-
Evidence Statement: *Maternal use of zolpidem at any time during pregnancy may be associated with an increased risk of the newborn being small for gestational age (low certainty evidence)*											

[549] Wang 2010.
[550] Wang 2010.
[551] Compared exposure in a non-mental health disorder population with non-exposure in a non-mental health disorder population.
[552] Wang 2010.
[553] Wang 2010.

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[537]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)** *P value*	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[538]**	**Risk with intervention[539]**
Respiratory difficulty: see Section AppD4.1.4.7.2											
90 (1 – OBS)[554]	None	NA	None	Unknown(a)	None	●○○○ Very low	**Unexposed – exposed to other psychotropic drugs** NA	**Zolpidem and other psychotropic drugs (any time)** NA	NR P=0.49	32 per 1000[555]	Not estimable
Evidence Statement: *Maternal use of zolpidem at any time during pregnancy does not appear to be associated with an increased risk of respiratory difficulty (very low certainty evidence)*											

Footnotes:
a. Downgraded one level due to moderate risk of bias; potential for selection bias due to exclusion of planned abortions, miscarriages and stillborn from the analysis.
b. Downgraded one level due to indirectness; compared with a general population.
c. Downgraded one level due to imprecision; 95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25 or SMD –0.5/0.5, no measure of precision available, or no events.

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.
Abbreviations: CI, confidence interval; NA, not available; NR, not reported; OBS, observational studies; OR, odds ratio; RE, risk estimate; RR, relative risk; SMD, standardised mean difference.

[554] Juric 2009.
[555] Malm 2015.

Table D3-36 Evidence Profile table: zopiclone

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias**[556]	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control**[557]	**Risk with intervention**[558]
Major malformation: see Section AppD4.1.4.1.2											
Heart anomalies *19,599* (1 – OBS)[559]	Serious(a)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**Zopiclone (first trimester)** NA	OR 0.93 (0.40, 2.15)	28 per 1000[560]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of zopiclone during the first trimester of pregnancy and major malformation in the newborn is uncertain.*											
Cardiac malformation: see Section AppD4.1.4.2.2											
Heart anomalies *19,599* (1 – OBS)[561]	Serious(a)	NA	None	Serious(b)	None	○○○○ Inadequate	**Unexposed** NA	**Zopiclone (first trimester)** NA	OR 2.03 (0.69, 6.02)	6 per 1000[562]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of zopiclone during the first trimester of pregnancy and cardiac malformation in the newborn is uncertain.*											
Miscarriage: see Section AppD4.1.4.4.2											
80 (1 – OBS)[563]	None	NA	Serious(b)	Unknown(c)	None	○○○○ Inadequate	**Unexposed** NA	**Zopiclone (any time)** NA	NR 17.5% vs. 7.5% *NR*	81 per 1000[564]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of zopiclone at any time during pregnancy and miscarriage is uncertain.*											

[556] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns will be noted and may result in further downgrading of the quality of the evidence.
[557] Based on average risk from unexposed, depressed control groups of population-based cohort studies providing data for that outcome.
[558] Calculated by multiplying relative effect by control risk.
[559] Ban 2014b.
[560] Ban 2014a (depressed/anxious population).
[561] Ban 2014b.
[562] Ban 2014a, Huybrechts 2014a, Petersen 2016, Margulis 2013 (depressed/anxious population).
[563] Diav-Citrin 1999.
[564] Almeida 2016 and Ban 2012.

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Additional risk of bias[556]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects**	
							Unexposed	**Exposed**		**Risk with control[557]**	**Risk with intervention[558]**
Preterm birth: see Section AppD4.1.4.5.2											
69 (1 – OBS)[565]	Serious(d)	NA	Serious(b)	Unknown(c)	None	○○○○ Inadequate	**Unexposed** NA	**Zopiclone (any time)** NA	NR 21.9% vs. 5.4% *0.07*	60 per 1000[566]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of zopiclone at any time during pregnancy and preterm birth is uncertain.*											
Small for gestational age: see Section AppD4.1.4.6.2											
68 (1 – OBS)[567]	Serious(d)	NA	Serious(b)	Unknown(c)	None	○○○○ Inadequate	**Unexposed** NA	**Zopiclone (any time)** NA	NR 6.3% vs. 5.6% *NR*	Unknown	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of zopiclone at any time during pregnancy and being small for gestational age is uncertain.*											

Footnotes:
a. Downgraded one level due to moderate risk of bias; potential for selection bias due to exclusion of planned abortions, miscarriages and stillborn from the analysis.
b. Downgraded one level due to imprecision; 95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25 or SMD –0.5/0.5, no measure of precision available, or no events.
c. Downgraded one level due to moderate risk of bias; potential for selection bias due to exclusion of miscarriages and stillborn from the analysis.
d. Downgraded one level due to moderate risk of bias; substantial number of exposures excluded from analysis.

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: CI, confidence interval; NA, not available; NR, not reported; OBS, observational studies; OR, odds ratio; RE, risk estimate; RR, relative risk; SMD, standardised mean difference.

[565] Diav-Citrin 1999.
[566] Malm 2015 (depressed population).
[567] Diav-Citrin 1999.

D3.1.5 Lithium

The following section presents the Evidence Profile tables for lithium use. The quantity of evidence available for the assessment of lithium was limited compared with other pharmacological agents. However, there was sufficient evidence available to limit the final analyses to those that adjusted risk estimates for confounding, or included a comparator population with a psychiatric diagnosis.

It should be noted that the Expert Working Group (EWG) and Harms Expert Subcommittee identified Ebstein's anomaly, a congenital heart defect, as an additional outcome of interest that may be associated with maternal exposure to lithium during pregnancy. As such, additional data relating to this specific outcome was also assessed.

As the evidence was based on data from cohort and case-control studies, in many cases the results were presented as odds ratios instead of relative risks. Where the baseline risk was < 7%, it was assumed that the odds ratio approximates the relative risk and the results were interpreted as relative risks. Where baseline risk was not available in a pregnant unexposed population with a mental health disorder, the baseline risks identified for a depressed population were used as a proxy.

The following observations were made regarding the body of evidence for lithium harms:

- No meta-analyses were feasible for any outcome, so the body of evidence for each outcome comprised single studies.
- Only two of the included studies adjusted for potential confounding in their analyses (for select outcomes only), and only one of those studies adjusted for mental health indication.
- Three studies included an unexposed comparator group with a mental health diagnosis.

The scoping search identified two SRs relating to the assessment of infant harms associated with lithium use, one of which provides a quantitative assessment of the included studies (NICE 2015), while the other provides a narrative assessment (Galbally 2010). The NICE 2015 SR noted that there was limited evidence for lithium due to the small number of studies that provided extractable data.

As none of the pooled risk estimates reported in NICE 2015 exclusively used data adjusted for potential confounders, it was necessary to update the literature search and assess the evidence from original comparative studies. A total of eight comparative studies were identified, six from the NICE 2015 and Galbally 2010 SRs and a further two (Diav-Citrin 2014; Källén 2013) from the updated literature search. Where available, studies that adjusted for potential confounders, or used a comparator population with a psychiatric diagnosis, have been designated as primary evidence for the outcomes of interest and are included in the EP table in preference to unadjusted data. Data were available for outcomes relating to lithium exposure during pregnancy and major malformations, cardiac malformations, septal malformations, miscarriage, stillbirth, neonatal mortality and preterm birth.

A summary of the characteristics of the individual included studies can be found in **Table AppD2-21** in **Appendix D2.1.5.2**. A detailed discussion of the evidence can be found in **Appendix D4.1.5**.

Table D 3-37 presents a summary of the results of the Evidence Review of lithium and the location of the detailed assessment of the certainty of evidence in the evidence profile table. The findings suggest that maternal exposure to lithium during pregnancy may be associated with an increased risk of cardiac malformation, miscarriage and neonatal mortality, while the evidence was inadequate and the risk uncertain for major and septal malformations, Ebstein's anomaly, still birth and preterm birth.

Although several studies compared birthweights in babies exposed to lithium during pregnancy versus unexposed controls, only one study was identified that assessed the association between lithium use and being large for gestational age (Troyer 1993). The definition of large for gestational age was not provided in the publication and the study results were poorly reported (although the discussion implied that there was no difference between study arms). As such, this outcome is not presented in the Evidence Profile table.

As noted, comparative data are also shown for the association between lithium exposure and Ebstein's anomaly of the heart in the offspring. In the 1970's, a very strong association was suggested between lithium treatment during pregnancy and Ebstein's anomaly. A retrospective analysis of data from the Danish Register of Lithium Babies suggested a high risk of Ebstein's anomaly: 6 out of 225 (2.7%) exposed children versus an incidence of 1 in 20,000 (0.005%) in the general population (Weinstein et al, 1976). However, this is now understood to be a gross overestimation due to voluntary reporting bias. Several subsequent controlled epidemiologic studies found no association between lithium use and Ebstein's anomaly, and a 1994 review of epidemiological data concluded that the teratogenic risk of first trimester lithium exposure is lower than originally suggested (Cohen et al, 1994).

Four of the comparative studies cited in the two identified SRs did not provide data for the current review but are noteworthy as they specifically relate to Ebstein's anomaly. Correa-Villasenor 1994 reviewed 44 cases of Ebstein's anomaly and 3,572 controls without cardiovascular malformations from the Baltimore-Washington Infant Study (BWIS). None of the case mothers reported lithium use during pregnancy but there were two lithium exposures in the control group. Edmonds 1990 reviewed 34 cases of Ebstein's anomaly and 34 control children and identified no history of maternal use of lithium or manic depression in pregnancy for any of the children. Zalzstein 1990 reviewed 59 cases of patients born between 1971 and 1988 who were diagnosed with Ebstein's anomaly in a single hospital in Canada. No cases had a lithium exposure recorded. Likewise, Kallen 1988 found no instances of lithium exposure in a review of 69 cases of Ebstein's anomaly or tricuspid atresia from the International Clearinghouse for Birth Defects Monitoring Systems and a review of 15 Ebstein cases from the France Rhone-Alps-Auverge monitoring system.

Table D 3-37 Summary of results of the Evidence Review for lithium

Intervention	**Increased/may be increased risk of harm** Outcome Certainty of evidence	**Appears to be no increased risk of harm** Outcome Certainty of evidence	**Decreased/may be decreased risk of harm** Outcome Certainty of evidence	**Uncertain** Outcome ○○○○	**Evidence profile table**
Lithium	Cardiac malformation ●○○○ Miscarriage ●○○○ Neonatal mortality ●○○○			Major malformation Septal malformation Ebstein's anomaly Stillbirth Preterm birth	**Table D3-38**

Note: All comparisons are against non-exposure, unless otherwise stated. Certainty of evidence gradings are as follows: ●●●● – high certainty; ●●●○ – moderate certainty; ●●○○ – low certainty; ●○○○ – very low certainty; ○○○○ – inadequate certainty.

Table D3-38 Evidence Profile table: lithium harms

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias**[568]	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Study event rates**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							With control	**With intervention**		**Risk with control**[569]	**Risk with intervention**[570]
Major malformations: see Section AppD4.1.5.1.2											
(1 – OBS)[571]	Serious (a)	NA	Serious (b)	Serious (d)	None	○○○○ Inadequate	**Unexposed – not adjusted for indication**[572] (N = 1,575,613)	**Lithium (pregnancy)**[573] (N = 234)	ARR 1.09 (0.52, 2.00)	28 per 1000[574]	-
(1 – OBS)[575]	Serious (c)	NA	None	Serious (d)	None	○○○○ Inadequate	**Unexposed – bipolar disorder** (N = 61)	**Lithium (1st trimester)** (N = 123)	RR 1.98 (0.43, 9.06)[576]	28 per 1000[574]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of lithium during early pregnancy and major malformation in the newborn is uncertain.*											
Cardiac malformations: see Section AppD4.1.5.2.2											
(1 – OBS)[575]	Serious (e)	NA	None	None	None	●○○○ Very low	**Unexposed – adjusted for bipolar disorder**[577] NR[578]	**Lithium (1st trimester)** NR[578]	**ARR 4.75 (1.11, 20.36)**	6 per 1000[579]	29 per 1000 (7, 122)
(1 – OBS)[575]	Serious (c)	NA	None	Serious (d)	None	○○○○ Inadequate	**Unexposed – bipolar disorder** (N = 61)	**Lithium (1st trimester)** (N = 123)	RR 1.24 (0.25, 6.21)[576]	6 per 1000[579]	-
Evidence Statement: *Maternal use of lithium during the first trimester of pregnancy may be associated with cardiac malformation, from an absolute risk of 0.6% to 2.9% (very low certainty evidence).*											

[568] As the quality of the evidence starts at 'low' for observational studies, the main biases associated with observational study design have already been taken into account. Any additional outcome-specific or other methodological concerns are noted and may result in further downgrading of the quality of the evidence.
[569] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[570] Calculated by multiplying relative effect by control risk.
[571] Källén 2013. Outcome captured as 'relatively severe malformations'.
[572] Adjusted for year of birth, maternal age (5-year class), parity (1-4+), smoking in early pregnancy and BMI.
[573] The exposure was at least in the first trimester of pregnancy in 90.2% of this lithium-exposed group. The medication was taken throughout pregnancy in 58.5% of these pregnancies. Concurrent psychiatric medications were taken by 66.1% of women in this cohort.
[574] Ban 2014a (depressed/anxious population).
[575] Diav-Citrin 2014
[576] Unadjusted risk calculated post hoc from crude data using Review Manager 5.3
[577] Adjusted for pregnancy order, smoking 10 or more cigarettes a day, bipolar disorder.
[578] Cases in analysis: 822

[579] Petersen 2016, Ban 2014a, Huybrechts 2014a and Margulis 2013(depressed/anxious population).

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[568]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Study event rates**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							With control	**With intervention**		**Risk with control[569]**	**Risk with intervention[570]**
Septal malformations: see Section AppD4.1.5.3.2											
(1 – OBS)[575]	Serious (c)	NA	None	Serious (d)	None	○○○○ Inadequate	**Unexposed – bipolar disorder** (N = 61)	**Lithium (1st trimester)** (N = 123)	RR 1.49 (0.16, 14.01)[576]	3 per 1000[580]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of lithium during the first trimester of pregnancy and septal malformation in the newborn is uncertain.*											
Ebstein's anomaly: see Section AppD4.1.5.4.2											
(1 – OBS)[575]	Serious (c)	NA	None	Serious (d)	None	○○○○ Inadequate	**Unexposed – bipolar disorder** (N = 61)	**Lithium (1st trimester)** (N = 123)	RR 1.50 (0.06, 36.29)[576]	<1 per 1000[581]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of lithium during the first trimester of pregnancy and Ebstein's anomaly in the newborn is uncertain.*											
Miscarriage: see Section AppD4.1.5.5.2											
(1 – OBS)[575]	Serious (e)	NA	None	None	None	●○○○ Very low	**Unexposed – adjusted for bipolar disorder[582]** NR[583]	**Lithium (pregnancy)[573]** NR[583]	**AOR 1.94 (1.08, 3.48)**	81 per 1000[584]	NE
(1 – OBS)[575]	Serious (c)	NA	None	Serious (d)	None	○○○○ Inadequate	**Unexposed – bipolar disorder** (N = 72)	**Lithium (1st trimester)** (N = 183)	RR 1.97 (0.86, 4.53)[576]	81 per 1000[584]	-
Evidence Statement: *Maternal use of lithium during early pregnancy may be associated with miscarriage (very low certainty evidence).*											

[580] The Bérard 2015 study used an insured population and as such the prevalence of septal malformations in this study (1.83%) is not likely to be representative of the general population with depression. To estimate the prevalence, it is assumed that 50% of cardiac malformations are septal, resulting in an estimate of 0.3%.
[581] Refers to risk in the general population (0.005%) from Weinstein et al (1976).
[582] Adjusted for maternal age, previous miscarriage, smoking status, bipolar disorder, gestational age at initial contact with the information centre.
[583] Cases in analysis: 911
[584] Based on an unexposed/depressed population (Almeida 2016 and Ban 2012).

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* **(No. studies)**	**Additional risk of bias[568]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Study event rates**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							With control	**With intervention**		**Risk with control[569]**	**Risk with intervention[570]**
Stillbirth: see Section AppD4.1.5.6.2											
(1 – OBS)[575]	Serious (e)	NA	None	Serious (d)	None	○○○○ Inadequate	**Unexposed – bipolar disorder** (N = 72)	**Lithium (pregnancy)[573]** (N = 183)	RR 2.78 (0.15, 53.10) [585]	Unknown	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of lithium during early pregnancy and stillbirth is uncertain.*											
Neonatal mortality: see Section AppD4.1.5.6.2											
(1 – OBS)[586]	None	NA	None	Serious (d)	None	●○○○ Very low	**Unexposed – manic depression inpatients** (N = 80)	**Lithium – manic depression inpatients (1st trimester)** (N = 41)	RR 17.36 (0.96, 314.78)[585]	5 per 1000[587]	87 per 1000 (5, 1574)
Evidence Statement: *Maternal use of lithium for severe manic depression[588] during the first trimester of pregnancy may be associated with neonatal mortality (very low certainty evidence).*											
Preterm birth: see Section AppD4.1.5.7.2											
(1 – OBS)[575]	Serious (c)	NA	None	Serious (d)	None	○○○○ Inadequate	**Unexposed – bipolar disorder** (N = 59)	**Lithium (pregnancy)[573]** (N = 131)	RR 1.35 (0.57, 3.23)[585]	60 per 1000[589]	-
Evidence Statement: *Due to the inadequate certainty of the evidence, any association between maternal use of lithium during early pregnancy and preterm birth is uncertain.*											

Footnotes:
a. Downgraded one level due to moderate risk of bias; potential selection bias due to not capturing potential excess malformations coinciding with miscarriage, abortion or stillbirth.
b. Downgraded one level due to indirectness caused by use of control group without a mental health disorder diagnosis, with no adjustment for indication.
c. Downgraded one level due to moderate risk of bias: inadequate adjustment for indication – restricting comparator population to only bipolar disorder.
d. Downgraded one level due to imprecision (95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25).
e. Downgraded one level due to moderate risk of bias: inadequate adjustment for indication –adjusting for only bipolar disorder where 33% of exposure group had other diagnoses.

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: AD, antidepressant; AOR, adjusted odds ratio; ARR, adjusted relative risk; CI, confidence interval; NE, not estimable; NR, not reported; OBS, observational studies; OR, odds ratio; RR, relative risk; SSRI, selective serotonin reuptake inhibitor; TCA, tricyclic antidepressant.

[585] Unadjusted risk calculated post hoc from crude data using Review Manager 5.3
[586] Källén 1983
[587] Based on an unexposed/depressed population (Ban 2012).
[588] Women in the study had been treated as an inpatient for manic depression and were therefore likely to have severe disease.
[589] Based on an unexposed/depressed population (Malm 2015).

D3.2 COMPLEMENTARY

The following section presents the Evidence Profile tables for the complementary treatments examined: omega-3 fatty acids, St John's wort and Gingko biloba. These specific complementary agents were identified by the Harms Expert Subcommittee as being used by pregnant women with mental health issues. No intervention-specific rules were required for downgrading of the certainty of this body of evidence.

D3.2.1 Omega-3 fatty acids

A summary of the characteristics of the individual included studies can be found in **Table AppD2-22** in **Appendix D2.2.1.1**. A detailed discussion of the evidence can be found in **Appendix D4.2.1**.

Table D 3-39 presents a summary of the results of the Evidence Review of omega-3 fatty acids and the location of the detailed assessment of the certainty of evidence in the evidence profile table. All comparisons are for exposure versus non-exposure, unless otherwise stated. It should be noted that the assessment of the harms associated with omega-3 fatty acids was limited to SRs of RCTs, due to the large amount of RCT evidence available; all of this evidence has been conducted in a general, rather than a population with a mental health disorder. As this evidence is based on data from RCTs, for all outcomes, results were presented as RRs of MDs, and risks with control could be calculated directly from the study results.

Five SRs were identified, although the assessment of the evidence shown below was limited to the two most recent that reported pregnancy and birth outcomes (Kar 2016 and Saccone 2016b) and one SR reporting on neurodevelopmental outcomes (Gould 2013).

The findings of the Kar 2016 SR suggest that maternal use of omega-3 fatty acids during pregnancy provides some benefit in terms of reducing the rate of preterm birth, and may provide benefit in reducing the risk of the infant being small for gestational age. Interestingly, when Saccone 2016b limited the population to women with a previous preterm birth or small for gestational age infant, these benefits were not seen. Saccone 2016b also showed a reduction in neonatal mortality associated with use of omega-3 fatty acids from prior to 20 weeks' gestation. Finally, Gould 2013 showed no adverse impact of exposure to omega-3 fatty acids during pregnancy and cognitive, motor and language development assessed at various ages; a significant benefit of omega-3 fatty acids on cognitive development was seen as 2-5 years.

In summary, there is no evidence available to suggest that the use of omega-3 fatty acids during pregnancy has an adverse effect on the fetus, infant or child.

Table D 3-39 Summary of results of the Evidence Review for omega-3 fatty acids

Intervention	**Increased/may be increased risk of harm** Outcome Certainty of evidence	**Appears to be no increased risk of harm** Outcome Certainty of evidence	**Decreased/may be decreased risk of harm** Outcome Certainty of evidence	**Uncertain** Outcome ○○○○	**Evidence profile table**
		Cognitive development < 2 years and 5-12 years ●●●○/●●●● Motor development (any time) ●●●● Language development (< 5 years) ●●●○/●●●●	Preterm birth ●●●○ SFGA ●●●○ Neonatal mortality ●●●● Cognitive development (2-5 years) ●●●●		**Table D3-40**

Abbreviations: SFGA, small for gestational age.

Note: All comparisons are against non-exposure, unless otherwise stated. Certainty of evidence gradings are as follows: ●●●● – high certainty; ●●●○ – moderate certainty; ●●○○ – low certainty; ●○○○ – very low certainty; ○○○○ – inadequate certainty.

Table D3-40 Evidence Profile table: omega-3 fatty acids

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Study event rates**		**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects**	
							With control	**With intervention**		**Risk with control[590]**	**Risk with intervention[591]**
Preterm Birth: see Section AppD4.2.1.1.1											
Early preterm birth (< 34 weeks) *4,193* (6 – RCT)[592]	None	None	None	None	None	●●●● High	**Placebo** 3.2%	**Omega-3 fatty acids** 1.3%	**RR 0.42 (0.27, 0.66)**	30 per 1,000[593]	13 per 1,000 (8, 20)
Early preterm birth (< 34 weeks) – high risk *3,670* (3 – RCT)[594]	None	None[595]	None	None	None	●●●● High	**Placebo** NR	**Omega-3 fatty acids** NR	**RR 0.36 (0.18, 0.71)**	30 per 1,000[596]	11 per 1000 (5, 21)
Early preterm birth (< 34 weeks) – any risk *523* (3 – RCT)[597]	None	None[598]	None	Serious(a)	None	●●●○ Moderate	**Placebo** NR	**Omega-3 fatty acids** NR	RR 0.50 (0.24, 1.06)	30 per 1,000[599]	15 per 1000 (7, 32)
Preterm birth (< 37 weeks) 5,980 (9 – RCTs)[600]	None	None	None	None	None	●●●● High	**Placebo** 9.1%	**Omega-3 fatty acids** 7.4%	**RR 0.83 (0.70, 0.98)**	60 per 1,000[601]	50 per 1,000 (42, 59)
Preterm birth (< 37 weeks) – high risk 814 (4 – RCTs)[602]	None	None[603]	None	Serious(a)	None	●●●○ Moderate	**Placebo** NR	**Omega-3 fatty acids** NR	RR 0.83 (0.61, 1.11)	60 per 1,000[604]	50 per 1000 (37, 67)

[590] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[591] Calculated by multiplying relative effect by control risk; it is not considered appropriate to calculate the risk with intervention where the quality of the evidence is inadequate.
[592] Kar 2016 SR (Includes Carlson 2013, Makrides 2010, Mardones 2008, Onwude 1995, Olsen 2000, and Bulstra-Ramakers 1995).
[593] Estimated based on an untreated, depressed population (Malm 2015).
[594] Kar 2016 SR (included studies not reported).
[595] Heterogeneity not reported but largely consistent results across all available studies.
[596] Estimated based on an untreated, depressed population (Malm 2015).
[597] Kar 2016 SR (included studies not reported).
[598] Heterogeneity not reported but largely consistent results across all available studies.
[599] Estimated based on an untreated, depressed population (Malm 2015).
[600] Kar 2016 SR (Includes Carlson 2013, Makrides 2010, Mardones 2008, Onwude 1995, Olsen 2000, Bulstra-Ramakers 1995, Olsen 1992, Ramakrishnan 2010, and Smuts 2003).
[601] Estimated based on an untreated, depressed population (Malm 2015).
[602] Kar 2016 SR (included studies not reported).
[603] Heterogeneity not reported but largely consistent results across all available studies.
[604] Estimated based on an untreated, depressed population (Malm 2015).

Certainty assessment							**Summary of findings**				
Outcome subgroup ***No. participants*** **(No. studies)**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Study event rates**		**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects**	
							With control	**With intervention**		**Risk with control[590]**	**Risk with intervention[591]**
Preterm birth (< 37 weeks) – any risk 5,166 (5 – RCTs)[605]	None	None[606]	None	Serious(a)	None	●●●○ Moderate	**Placebo** NR	**Omega-3 fatty acids** NR	RR 0.83 (0.66, 1.05)	60 per 1,000[607]	50 per 1000 (40, 63)
Preterm birth (< 37 weeks) 5,689 (8 – RCTs)[608]	None	None[609]	None	None	None	●●●● High	**Placebo** NR	**Omega-3 fatty acids (> 400 mg)** NR	**RR 0.83 (0.69, 1.00)**	60 per 1,000[610]	50 per 1000 (41, 60)
Preterm birth (< 37 weeks) 291 (1 – RCT)[611]	None	NA	None	Serious(a)	None	●●●○ Moderate	**Placebo** NR	**Omega-3 fatty acids (< 400 mg)** NR	RR 0.86 (0.44, 1.69)	60 per 1,000[612]	52 per 1000 (26, 101)
Preterm birth (< 37 weeks) 5,156 (7 – RCT)[613]	None	None[614]	None	Serious(a)	None	●●●○ Moderate	**Placebo** NR	**Omega-3 fatty acids (< 24 weeks)** NR	RR 0.84 (0.69, 1.03)	60 per 1,000[615]	50 per 1000 (41, 62)
Preterm birth (< 37 weeks) 824 (2 – RCT)[616]	None	None[617]	None	Serious(a)	None	●●●○ Moderate	**Placebo** NR	**Omega-3 fatty acids (< 24 weeks)** NR	RR 0.75 (0.45, 1.25)	60 per 1,000[618]	45 per 1000 (27, 75)

[605] Kar 2016 SR (included studies not reported).
[606] Heterogeneity not reported but largely consistent results across all available studies.
[607] Estimated based on an untreated, depressed population (Malm 2015).
[608] Kar 2016 SR (included studies not reported).
[609] Heterogeneity not reported but largely consistent results across all available studies.
[610] Estimated based on an untreated, depressed population (Malm 2015).
[611] Kar 2016 SR (included studies not reported).
[612] Estimated based on an untreated, depressed population (Malm 2015).
[613] Kar 2016 SR (included studies not reported).
[614] Heterogeneity not reported but largely consistent results across all available studies.
[615] Estimated based on an untreated, depressed population (Malm 2015).
[616] Kar 2016 SR (included studies not reported).
[617] Heterogeneity not reported but largely consistent results across all available studies.
[618] Estimated based on an untreated, depressed population (Malm 2015).

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Study event rates**		**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects**	
							With control	**With intervention**		**Risk with control[590]**	**Risk with intervention[591]**
Women with no prior preterm birth 3493 (7 RCT)[619]	None	None	None	Serious(a)	None	●●●○ Moderate	**Placebo** 9.1%	**Omega-3 fatty acids** 7.7%	RR 0.90 (0.72, 1.11)	60 per 1,000[620]	54 per 1000 (43, 67)
Evidence Statements: *Maternal use of omega-3 fatty acids at any time during pregnancy is associated with a decreased risk of early preterm birth (< 34 weeks), from an absolute risk of 3% to 1.3% (high certainty evidence).* *Maternal use of omega-3 fatty acids at any time during pregnancy is associated with a decreased risk of preterm birth (< 37 weeks), from an absolute risk of 6% to 5% (high certainty evidence).* *Maternal use of omega-3 fatty acids at any time during pregnancy in women with no prior preterm birth is not associated with a decreased risk of preterm birth (moderate certainty evidence)*											
Small for gestational age: see Section AppD4.2.1.1.2											
5,469 (8 – RCTs)[621]	None	None	None	Serious(a)	None	●●●○ Moderate	**Placebo** NR	**Omega-3 fatty acids** NR	RR 0.82 (0.66, 1.03)	Unknown	Not estimable
History of previous SGA infant 575 (3 – RCTs)[622]	None	None	None	Serious(a)	None	●●●○ Moderate	**Placebo** NR	**Omega-3 fatty acids** NR	RR 1.13 (0.83, 1.54)	Unknown	Not estimable
Evidence Statements: *Maternal use of omega-3 fatty acids at any time during pregnancy may be associated with a decreased risk of the infant being small for gestational age; however, the finding was not statistically significant (moderate certainty evidence).* *Maternal use of omega-3 fatty acids at any time during pregnancy in women with a history of small for gestational age infants is not associated with an increased risk of the infant being small for gestational age (moderate certainty evidence).*											
Neonatal Deaths: see Section AppD4.2.1.1.3											
6,751 (7 – RCTs)[623]	None	None	None	None	None	●●●○ Moderate	**Placebo** NR	**Omega-3 fatty acids** NR	RR 0.51 (0.26, 1.01)	5 per 1000[624]	3 per 1000 (1, 5)

[619] Saccone 2016b SR (included Olsen 1992, Bulstra-Ramakers 1994, Onwude 1995, Malcolm 2003, Tofail 2006, Makrides 2010, Escolano-Margarit 2011).
[620] Estimated based on an untreated, depressed population (Malm 2015).
[621] Kar 2016 SR (Includes Makrides 2010, Mardones 2008, Onwude 1995, Olsen 2000, Bulstra-Ramakers 1995, Olsen 1992, Ramakrishnan 2010, and Smuts 2003).
[622] Saccone 2016b SR (Includes Onwude 1995, Olsen 2000, Bulstra-Ramakers 1995).
[623] Kar 2016 SR (Includes Makrides 2010, Olsen 2000, Bulstra-Ramakers 1995, Olsen 1992, Ramakrishnan 2010).
[624] Estimated based on an untreated, depressed population (Ban 2012).

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Study event rates**		**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects**	
							With control	**With intervention**		**Risk with control[590]**	**Risk with intervention[591]**
2,462 (2 – RCTs)[625]	None	None	None	None	None	●●●● High	**Placebo** 1.2%	**Omega-3 fatty acids (from ≤ 20 w gestation)** 0.3%	**RR 0.27 (0.09, 0.79)**	5 per 1000[626]	1 per 1000 (<1, 4)
Evidence Statement: *Maternal use of omega-3 fatty acids from ≤ 20 weeks gestation is associated with a decreased risk of neonatal mortality; from an absolute risk of 0.5% to 0.1% (high certainty evidence).*											
Cognitive development: see Section AppD4.2.1.1.4											
< 12 months[627] *249* (1 – RCT)[628]	Serious(b)	NA	None	None	None	●●●○ Moderate	**Placebo** NA	**Omega-3 LCPUFA (P & L)** NA	MD 1.00 (-0.96, 2.96)	NA	-
12-24 months[629] *801* (2 – RCT)[630]	None	None	None	None	None	●●●● High	**Placebo** NA	**Omega-3 LCPUFA (P & L)** NA	MD -0.08 (-1.72, 1.57)	NA	-
2-5 years[631] *156* (2 – RCT)[632]	None	None	None	None	None	●●●● High	**Placebo** NA	**Omega-3 LCPUFA (P & L)** NA	**MD 3.92 (0.77, 7.08)**	NA	-
5-12 years[633] *225* (2 – RCT)[634]	None	None	None	None	None	●●●● High	**Placebo** NA	**Omega-3 LCPUFA (P & L)** NA	MD 0.36 (-2.61, 3.32)	NA	-
12-24 months[635] *726* (1 – RCT)[636]	None	NA	None	None	None	●●●● High	**Placebo** NA	**Omega-3 LCPUFA (P only)** NA	MD 0.06 (-1.66, 1.78)	NA	-

[625] Saccone 2016b (includes Bulstra-Ramakers 1994 and Makrides 2010).
[626] Estimated based on an untreated, depressed population (Ban 2012).
[627] Cognitive development measured using the BSID-II.
[628] Gould 2013 SR (includes Tofail 2006).
[629] Cognitive development measured using the BSID-II and III.
[630] Gould 2013 SR (includes Van Goor 2011 and Makrides 2010).
[631] Cognitive development measured using the GMDS and K-ABC.
[632] Gould 2013 SR (includes Dunstan 2008 and Helland 2003).
[633] Cognitive development measured using the K-ABC.
[634] Gould 2013 SR (includes Campoy 2011 and Helland 2008).
[635] Cognitive development measured using BSID III.
[636] Gould 2013 SR (includes Makrides 2010).

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Study event rates**		**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects**	
							With control	**With intervention**		**Risk with control[590]**	**Risk with intervention[591]**
2-5 years[637] *72* (1 – RCT)[638]	Serious(b)	NA	None	Serious(a)	None	●●○○ Low	**Placebo** NA	**Omega-3 LCPUFA (P only)** NA	MD 3.70 (-1.02, 8.42)	NA	-
5-12 years[639] *82* (1 – RCT)[640]	Unknown (b)[641]	NA	None	None	None	●●●○ Moderate	**Placebo** NA	**Omega-3 LCPUFA (P only)** NA	MD 0.00 (-5.52, 5.52)	NA	-
Evidence Statements: *Maternal use of omega-3 fatty acids at any time during pregnancy or lactation is not associated with a reduction in cognitive development at < 12 months, 12-24 months and 5-12 years (moderate to high certainty evidence).* *Maternal use of omega-3 fatty acids at any time during pregnancy or lactation is associated with an improvement in cognitive development at 2-5 years (high certainty evidence).* *Maternal use of omega-3 fatty acids at any time during pregnancy only is not associated with a reduction in cognitive development at 2-5 years (low to high certainty evidence).*											
Motor development: see Section AppD4.2.1.1.5											
< 12 months[642] *249* (1 – RCT)[643]	Serious(b)	NA	None	None	None	●●●○ Moderate	**Placebo** NA	**Omega-3 LCPUFA (P & L)** NA	MD 1.20 (-1.41, 3.81)	NA	-
12-24 months[644] *801* (2 – RCT)[645]	None	Very serious(c)	None	Serious(a)	None	●○○○ Very low	**Placebo** NA	**Omega-3 LCPUFA (P & L)** NA	MD 1.52 (-2.29, 5.32)	NA	-
2-5 years[646] *72* (1 – RCT)[647]	None	NA	None	Serious(a)	None	●●●○ Moderate	**Placebo** NA	**Omega-3 LCPUFA (P & L)** NA	MD 4.60 (-1.14, 10.34)	NA	-

[637] Cognitive development measured using the GMDS.
[638] Gould 2013 SR (includes Dunstan 2008).
[639] Cognitive development measurement used not reported.
[640] Gould 2013 SR (includes Campoy 2011).
[641] Quality for Campoy 2011 not reported in Gould 2013. Assumed to have a moderate risk of bias and downgraded one level for serious risk of bias.
[642] Motor development measured using BSID II.
[643] Gould 2013 (includes Tofail 2006).
[644] Motor development measured using BSID II and III.
[645] Gould 2013 SR (includes Van Goor 2011 and Makrides 2010).
[646] Motor development measured using GMDS.
[647] Gould 2013 SR (includes Dunstan 2008).

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Study event rates**		**Risk estimate (95% CI)** ***P value***	**Anticipated absolute effects**	
							With control	**With intervention**		**Risk with control[590]**	**Risk with intervention[591]**
12-24 months[648] *726* (1 – RCT)[649]	None	NA	None	None	None	●●●● High	**Placebo** NA	**Omega-3 LCPUFA (P only)** NA	MD 0.06 (-1.52, 1.64)	NA	-
Evidence Statements: *Maternal use of omega-3 fatty acids at any time during pregnancy or lactation is not associated with a reduction in motor development at < 12 months, 12-24 months and 2-5 years (very low to moderate certainty evidence).* *Maternal use of omega-3 fatty acids at any time during pregnancy only is not associated with a reduction in motor development at 12-24 months (high certainty evidence).*											
Language development: see Section AppD4.2.1.1.6											
12-24 months[650] *726* (1 – RCT)[651]	None	NA	None	None	None	●●●● High	**Placebo** NA	**Omega-3 LCPUFA (P only)** NA	MD -1.47 (-3.58, 0.64)	NA	-
2-5 years[652] *70* (1 – RCT)[653]	None	NA	None	Serious(a)	None	●●●○ Moderate	**Placebo** NA	**Omega-3 LCPUFA (P only)** NA	MD 3.90 (-0.73, 8.53)	NA	-
Evidence Statement: *Maternal use of omega-3 fatty acids at any time during pregnancy only is not associated with a reduction in language development at 12-24 months and 2-5 years (moderate to high certainty evidence).*											
Footnotes: a. Downgraded one level due to imprecision; 95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25, no measure of precision available, or no events. b. Downgraded two levels due to high risk of bias; unknown random sequence generation and allocation concealment, and high risk of bias for follow-up and other bias. c. Downgraded two levels due to substantial heterogeneity (I^2 > 60%).											

Abbreviations: BSID, Bayley Scales of Infant Development; CI, confidence interval; GMDS, Griffiths Mental Development Scales; K-ABC, Kaufman Assessment Battery for Children; MD, mean difference; NA, not available; NR, not reported; P, pregnancy; P & L, pregnancy and lactation; PPVT, Peabody Picture Vocabulary Test; RCT, randomised controlled trial; RR, relative risk, w weeks.

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

[648] Motor development measured using BSID II.
[649] Gould 2013 SR (includes Makrides 2010).
[650] Language development measured using
[651] Gould 2013 (includes Makrides 2010).
[652] Language development measured using PPVT.
[653] Gould 2013 SR (includes Dunstan 2008).

D3.2.2 St John's wort

A summary of the characteristics of the individual included studies can be found in **Table AppD2-27** in **Appendix D2.2.2.2**. A detailed discussion of the evidence for each group or individual intervention type and outcome can be found in **Appendix D4.2.2**.

Table D 3-41 presents a summary of the results of the Evidence Review of St John's wort and the location of the detailed assessment of the certainty of evidence in the evidence profile table. All comparisons are for exposure versus non-exposure, unless otherwise stated. Three SRs were identified via the searches; however, these each included only one to two cohort studies and two case reports and described them narratively. Two cohort studies were identified; the one by Moretti 2009 (based on data from the Mother-risk program in Canada) was included preferentially because it adjusted for potential confounders. Due to the inadequate certainty of this study, it was determined that the effect of antenatal or post-natal exposure to St John's wort on fetal, infant or child harms is uncertain. Moretti 2009 note that "though further large scale studies are still needed, this first study on the effects of St John's wort in human pregnancy does provide some evidence of fetal safety."

Table D 3-41 Summary of results of the Evidence Review for St John's wort

Intervention	**Increased/may be increased risk of harm** Outcome Certainty of evidence	**Appears to be no increased risk of harm** Outcome Certainty of evidence	**Decreased/may be decreased risk of harm** Outcome Certainty of evidence	**Uncertain** Outcome ○○○○	**Evidence profile table**
St John's wort				Major malformation Major malformation (vs ADs) Preterm birth Preterm birth (vs ADs)	**Table D3-42**

Abbreviations: AD, antidepressant.

Note: All comparisons are against non-exposure, unless otherwise stated. Certainty of evidence gradings are as follows: ●●●● – high certainty; ●●●○ – moderate certainty; ●●○○ – low certainty; ●○○○ – very low certainty; ○○○○ – inadequate certainty.

Table D3-42 Evidence Profile table: St John's wort

Certainty assessment							Summary of findings				
Outcome subgroup (No. studies)	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall certainty of evidence	Population (N)		Risk estimate (95% CI) or % vs. %; P value	Anticipated absolute effects	
							Non-exposure	Exposure		Risk with control[654]	Risk with intervention[655]
Major malformation: see Section AppD4.2.2.2.1											
1 – OBS[656]	Very serious(a)	NA	Serious(b)	Serious(c)[657]	None	○○○○ Inadequate	**Unexposed** 56	**St John's wort (any time)** 38	5.3% vs. 0%; 0.20[658]	28 per 1000[659]	-
1 – OBS[656]	Very serious(a)	NA	None	Serious(c)[660]	None	○○○○ Inadequate	**Antidepressants[661] (any time)** 48	**St John's wort (any time)** 38	5.3% vs. 4.2%; 0.81[658]	42 per 1000[656]	-
Evidence Statements: *Due to the inadequate certainty of the evidence, any association between maternal use of St John's wort at any time during pregnancy and an increased risk of major malformation in the newborn is uncertain.* *Due to the inadequate certainty of the evidence, any additional risk of major malformation in the newborn associated with maternal use of St John's wort at any time during pregnancy, compared with maternal use of conventional pharmacologic treatment for depression during the same period, is uncertain.*											
Preterm birth: see Section AppD4.2.2.2.3											
1 - OBS[656]	Very serious(d)	NA	Serious(b)	Serious(c)[662]	None	○○○○ Inadequate	**Unexposed** 45	**St John's wort (any time)** 43	4.7% vs. 13.3%; 0.18[663]	60 per 1000[664]	-

[654] Based on average risk from unexposed, depressed control groups of population-based cohort studies.
[655] Calculated by multiplying relative effect by control risk; it is not considered appropriate to calculate the risk with intervention where the quality of the evidence is inadequate.
[656] Moretti 2009.
[657] Based on post hoc calculation of risk estimate using Review Manager; RR 7.31 (0.36, 148.09).
[658] Calculated post hoc using Review Manager.
[659] Ban 2014a (depressed/anxious population).
[660] Based on post hoc calculation of risk estimate using Review Manager; RR 1.26 (0.19, 8.56).
[661] Described as conventional pharmacological treatment.
[662] Based on post hoc calculation of risk estimate using Review Manager; RR 0.35 (0.07, 1.63).
[663] Calculated post hoc using Review Manager.
[664] Petersen 2016, Ban 2014a, Huybrechts 2014a and Margulis 2013(depressed/anxious population).

Certainty assessment							Summary of findings				
Outcome subgroup (No. studies)	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI) or % vs. %; P value**	**Anticipated absolute effects**	
							Non-exposure	**Exposure**		**Risk with control[654]**	**Risk with intervention[655]**
1 – OBS[656]	Very serious(d)	NA	None	None	None	○○○○ Inadequate	**Antidepressants[665] (any time)** 39	**St John's wort (any time)** 43	**4.7% vs. 20.5%; 0.05**	205 per 1000	-
Evidence Statements: *Due to the inadequate certainty of the evidence, any association between maternal use of St John's wort at any time during pregnancy and an increased risk of preterm birth newborn is uncertain.* *Due to the inadequate certainty of the evidence, any decreased risk of preterm birth in the newborn associated with maternal use of St John's wort at any time during pregnancy, compared with maternal use of conventional pharmacologic treatment for depression during the same period, is uncertain.*											
Footnotes: a. Downgraded two levels due to high risk of bias; potential for selection bias due to exclusion of planned abortions, miscarriages and still born from the analysis, self-report ascertainment of outcome and incomplete follow-up. b. Downgraded one level due to indirectness; general population comparator group. c. Downgraded one level due to imprecision; 95% CI crosses the line of no effect and includes a measure of appreciable benefit and/or harm – RR 0.75/1.25, no measure of precision available, or no events. d. Downgraded two levels due to high risk of bias; self-report ascertainment of outcome and incomplete follow-up.											

Abbreviations: BSID, Bayley Scales of Infant Development; CI, confidence interval; GMDS, Griffiths Mental Development Scales; K-ABC, Kaufman Assessment Battery for Children; NA, not available; NR, not reported; P, pregnancy; P & L, pregnancy and lactation; PPVT, Peabody Picture Vocabulary Test; RCT, randomised controlled trial; RR, relative risk.

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

[665] Described as conventional pharmacological treatment.

D3.2.3 Gingko biloba

No SRs or individual comparative studies were identified that assessed the effect of perinatal exposure to Gingko biloba on fetal, infant or child harms.

D3.3 PHYSICAL

The following section presents the Evidence Profile tables for the physical treatments examined: ECT and TMS. These specific physical therapies were identified by the Harms Expert Subcommittee as potentially impacting on the fetus. No intervention-specific rules were required for downgrading of the certainty of this body of evidence.

D3.3.1 Electroconvulsive therapy

A summary of the characteristics of the individual included studies can be found in **Table AppD2-30** in **Appendix D2.3.1.2**. A detailed discussion of the evidence for each group or individual intervention type and outcome can be found in **Appendix D4.3.1**.

The EP table reporting the results of the assessment of ECT is presented in **Table D3-43**. The available evidence was based primarily on SRs of case series/reports and one very low certainty prospective cohort study that suggested no harm to the infant following exposure to ECT via breastfeeding (Babu 2013). As such, there is insufficient evidence available to make an Evidence Statement on the fetal/infant/child harms associated with use of ECT during pregnancy or the postnatal period.

Table D3-43 Evidence Profile table: ECT harms

Certainty assessment							Summary of findings				
Outcome subgroup *No. participants* (No. studies)	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall certainty of evidence	Population (N)		Risk estimate (95% CI)	Anticipated absolute effects	
							Non-exposure	Exposed		Risk with control	Risk difference with intervention
ECT – antenatal exposure											
There was no higher certainty evidence regarding the effect of antenatal exposure to ECT on infant harms. One pooled analysis of case reports concluded that ECT should be a "last resort" treatment[666] while three narrative reviews of largely case reports concluded that the risk of adverse harms to the fetus were low.[667] (see Section AppD4.3.1.3.1)											
ECT – postnatal exposure											
There was no higher certainty evidence regarding the effect of postnatal exposure to ECT on infant harms. One small prospective comparative study (without adjustment for potential confounding) suggests that breastfeeding following post-partum ECT does not result in adverse effect to the infant.[668] (see Section AppD4.3.1.4.1)											
Evidence Statement: *There is insufficient evidence available to make an Evidence Statement regarding the effect of antenatal or postnatal exposure to ECT on fetal or infant harms.*											
Footnotes: None											

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: CI, confidence interval; ECT, electroconvulsive therapy.

[666] Leikness 2015.

[667] Calaway 2016, Pompili 2014 and Anderson 2009.

[668] Babu 2013.

D3.3.2 Transcranial magnetic stimulation

A summary of the characteristics of the individual included studies can be found in **Table AppD2-31** in **Appendix D2.3.2.2**. A detailed discussion of the evidence for each group or individual intervention type and outcome can be found in **Appendix D4.3.2**.

The EP table reporting the results of the assessment of TMS is presented in **Table D3-44**. No SRs were identified in the SR search and updated search that assessed the impact of antenatal or postnatal exposure to TMS on the fetus, infant or child. The single included study (Eryilmaz 2015) compared the effect of TMS with no TMS in pregnant women with major depressive disorder. This study had a number of methodological deficiencies, the main ones being the use of a non-concurrent control group and a lack of adjustment for potential confounding. As such, there is insufficient evidence available to make an Evidence Statement on the fetal/infant harms associated with use of TMS during pregnancy or the postnatal period. It should be noted that the authors report no significant harms associated with the use of TMS, and showed no significant difference in motor or cognitive development, although there was a non-significant lower prevalence of mothers' perception in language development.

Table D3-44 Evidence Profile table: TMS harms

Certainty assessment							Summary of findings				
Outcome subgroup ***No. participants*** **(No. studies)**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Overall certainty of evidence**	**Population (N)**		**Risk estimate (95% CI)**	**Anticipated absolute effects**	
							Non-exposure	**Exposed**		**Risk with control**	**Risk difference with intervention**
TMS – antenatal exposure											
There was no higher certainty evidence regarding the effect of antenatal exposure to TMS on infant harms. One prospective cohort study with a non-concurrent untreated, depressed control group that did not sufficiently adjust for potential confounding showed no difference in infant adverse events or developmental delay at a mean of 32 months using the ADSI.[669] (see Section D4.3.2.1.2)											
Evidence Statement: *There is insufficient evidence available to make an Evidence Statement regarding the effect of antenatal or postnatal exposure to TMS on infant harms.*											
Footnotes: None											

Notes: Relative effects shown in black bold text denote a statistically significantly greater harm in the intervention group. Relative effects shown in grey bold text denote a statistically significantly greater harm in the control group.

Abbreviations: ADSI, Ankara Developmental Screening Inventory; CI, confidence interval; ECT, electroconvulsive therapy.

[669] Eryilmaz 2015.

D4 REFERENCES

Almeida, N. D., O. Basso, M. Abrahamowicz, R. Gagnon and R. Tamblyn (2016). "Risk of Miscarriage in Women Receiving Antidepressants in Early Pregnancy, Correcting for Induced Abortions." Epidemiology 27(4): 538-546.

Anderson, E. L. and I. M. Reti (2009). "ECT in pregnancy: a review of the literature from 1941 to 2007." Psychosom Med 71(2): 235-242.

Babu, G. N., H. Thippeswamy and P. S. Chandra (2013). "Use of electroconvulsive therapy (ECT) in postpartum psychosis-A naturalistic prospective study." Archives of Women's Mental Health 16(3): 247-251.

Ban, L., J. E. Gibson, J. West, L. Fiaschi, R. Sokal, L. Smeeth, P. Doyle, R. B. Hubbard and L. J. Tata (2014a). "Maternal depression, antidepressant prescriptions, and congenital anomaly risk in offspring: a population-based cohort study." Bjog 121(12): 1471-1481.

Ban, L., J. West, J. E. Gibson, L. Fiaschi, R. Sokal, P. Doyle, R. Hubbard, L. Smeeth and L. J. Tata (2014b). "First trimester exposure to anxiolytic and hypnotic drugs and the risks of major congenital anomalies: a United Kingdom population-based cohort study." PLoS One 9(6): e100996.

Ban, L., L. J. Tata, J. West, L. Fiaschi and J. E. Gibson (2012). "Live and non-live pregnancy outcomes among women with depression and anxiety: a population-based study." PLoS One 7(8): e43462.

Bérard, A., J. P. Zhao and O. Sheehy (2015). "Sertraline use during pregnancy and the risk of major malformations." Am J Obstet Gynecol 212(6): 795.e791-795.e712.

Bérard, A., T. Boukhris and O. Sheehy (2016). "Selective serotonin reuptake inhibitors and autism: additional data on the Quebec Pregnancy/Birth Cohort." Am J Obstet Gynecol 215(6): 803-805. Boukhris, T., O. Sheehy, L. Mottron and A. Berard (2016). "Antidepressant Use During Pregnancy and the Risk of Autism Spectrum Disorder in Children." JAMA Pediatr 170(2): 117-124.

Bromley, R., J. Weston, N. Adab, J. Greenhalgh, A. Sanniti, A. J. McKay, C. Tudur Smith and A. G. Marson (2014). "Treatment for epilepsy in pregnancy: neurodevelopmental outcomes in the child." Cochrane Database Syst Rev(10): Cd010236.

Brown, A. S., D. Gyllenberg, H. Malm, I. W. McKeague, S. Hinkka-Yli-Salomaki, M. Artama, M. Gissler, K. Cheslack-Postava, M. M. Weissman, J. A. Gingrich and A. Sourander (2016). "Association of Selective Serotonin Reuptake Inhibitor Exposure During Pregnancy With Speech, Scholastic, and Motor Disorders in Offspring." JAMA Psychiatry 73(11): 1163-1170.

Calaway, K., S. Coshal, K. Jones, J. Coverdale and R. Livingston (2016). "A Systematic Review of the Safety of Electroconvulsive Therapy Use During the First Trimester of Pregnancy." J ect.

Cohen, L. S., et al. (1994). "A reevaluation of risk of in utero exposure to lithium." JAMA 271(2): 146-150.

Cole, J. A., J. G. Modell, B. R. Haight, I. S. Cosmatos, J. M. Stoler and A. M. Walker (2007a). "Bupropion in pregnancy and the prevalence of congenital malformations." Pharmacoepidemiol Drug Saf 16(5): 474-484.

Cole, J. A., S. A. Ephross, I. S. Cosmatos and A. M. Walker (2007b). "Paroxetine in the first trimester and the prevalence of congenital malformations." Pharmacoepidemiol Drug Saf 16(10): 1075-1085.

Correa-Villasenor, A., et al. (1994). "Ebstein's malformation of the tricuspid valve: genetic and environmental factors. The Baltimore-Washington Infant Study Group." 50: 137-147.

Diav-Citrin, O., B. Okotore, K. Lucarelli and G. Koren (1999). "Pregnancy outcome following first-trimester exposure to zopiclone: a prospective controlled cohort study." Am J Perinatol 16(4): 157-160.

Diav-Citrin, O., S. Shechtman, E. Tahover, V. Finkel-Pekarsky, J. Arnon, D. Kennedy, A. Erebara, A. Einarson and A. Ornoy (2014). "Pregnancy outcome following in utero exposure to lithium: a prospective, comparative, observational study." Am J Psychiatry 171(7): 785-794.

Djulus, J., G. Koren, T. R. Einarson, L. Wilton, S. Shakir, O. Diav-Citrin, D. Kennedy, S. Voyer Lavigne, M. De Santis and A. Einarson (2006). "Exposure to mirtazapine during pregnancy: a prospective, comparative study of birth outcomes." J Clin Psychiatry 67(8): 1280-1284.

Edmonds, L. D. and G. P. Oakley Jr (1990). "Ebstein's anomaly and maternal lithium exposure during pregnancy." Teratology 41: 551-552.

El Marroun, H., T. J. White, N. J. van der Knaap, J. R. Homberg, G. Fernandez, N. K. Schoemaker, V. W. Jaddoe, A. Hofman, F. C. Verhulst, J. J. Hudziak, B. H. Stricker and H. Tiemeier (2014). "Prenatal exposure to selective serotonin reuptake inhibitors and social responsiveness symptoms of autism: population-based study of young children." Br J Psychiatry 205(2): 95-102.

Eros, E., A. E. Czeizel, M. Rockenbauer, H. T. Sorensen and J. Olsen (2002). "A population-based case-control teratologic study of nitrazepam, medazepam, tofisopam, alprazolum and clonazepam treatment during pregnancy." Eur J Obstet Gynecol Reprod Biol 101(2): 147-154.

Eryilmaz, G., G. H. Sayar, E. Ozten, I. G. Gul, O. Yorbik, N. Isiten and E. Bagci (2015). "Follow-up study of children whose mothers were treated with transcranial magnetic stimulation during pregnancy: preliminary results." Neuromodulation 18(4): 255-260.

Figueroa, R. (2010). "Use of antidepressants during pregnancy and risk of attention-deficit/hyperactivity disorder in the offspring." J Dev Behav Pediatr 31(8): 641-648.

Furu, K., H. Kieler, B. Haglund, A. Engeland, R. Selmer, O. Stephansson, U. A. Valdimarsdottir, H. Zoega, M. Artama, M. Gissler, H. Malm and M. Norgaard (2015). "Selective serotonin reuptake inhibitors and venlafaxine in early pregnancy and risk of birth defects: population based cohort study and sibling design." Bmj 350: h1798.

Galbally, M., M. Roberts and A. Buist (2010). "Mood stabilizers in pregnancy: A systematic review." Australian and New Zealand Journal of Psychiatry 44(11): 967-977.

Gould, J., L. Smithers and M. Makrides (2013) "The effect of maternal omega-3 (n?3) LCPUFA supplementation during pregnancy on early childhood cognitive and visual development: a systematic review and meta-analysis of randomized controlled trials (Structured abstract)." American Journal of Clinical Nutrition 97, 531-544.

Grigoriadis, S., E. H. VonderPorten, L. Mamisashvili, A. Eady, G. Tomlinson, C. L. Dennis, G. Koren, M. Steiner, P. Mousmanis, A. Cheung and L. E. Ross (2013b). "The effect of prenatal antidepressant exposure on neonatal adaptation: A systematic review and meta-analysis." Journal of Clinical Psychiatry 74(4): e309-e320.

Grigoriadis, S., E. H. Vonderporten, L. Mamisashvili, G. Tomlinson, C. L. Dennis, G. Koren, M. Steiner, P. Mousmanis, A. Cheung and L. E. Ross (2014). "Prenatal exposure to antidepressants and persistent pulmonary hypertension of the newborn: systematic review and meta-analysis." Bmj 348: f6932.

Grzeskowiak, L. E., A. L. Gilbert and J. L. Morrison (2012). "Neonatal outcomes after late-gestation exposure to selective serotonin reuptake inhibitors." J Clin Psychopharmacol 32(5): 615-621.

Harrington, R. A., L. C. Lee, R. M. Crum, A. W. Zimmerman and I. Hertz-Picciotto (2014). "Prenatal SSRI use and offspring with autism spectrum disorder or developmental delay." Pediatrics 133(5): e1241-1248.

Hayes, R. M., P. Wu, R. C. Shelton, W. O. Cooper, W. D. Dupont, E. Mitchel and T. V. Hartert (2012). "Maternal antidepressant use and adverse outcomes: a cohort study of 228,876 pregnancies." Am J Obstet Gynecol 207(1): 49.e41-49.

Huybrechts, K. F., B. T. Bateman, K. Palmsten, R. J. Desai, E. Patorno, C. Gopalakrishnan, R. Levin, H. Mogun and S. Hernandez-Diaz (2015). "Antidepressant use late in pregnancy and risk of persistent pulmonary hypertension of the newborn." Jama 313(21): 2142-2151.

Huybrechts, K. F., K. Palmsten, J. Avorn, L. S. Cohen, L. B. Holmes, J. M. Franklin, H. Mogun, R. Levin, M. Kowal, S. Setoguchi and S. Hernandez-Diaz (2014a). "Antidepressant use in pregnancy and the risk of cardiac defects." N Engl J Med 370(25): 2397-2407.

Huybrechts, K. F., S. Hernandez-Diaz, E. Patorno, R. J. Desai, H. Mogun, S. Z. Dejene, J. M. Cohen, A. Panchaud, L. Cohen and B. T. Bateman (2016). "Antipsychotic Use in Pregnancy and the Risk for Congenital Malformations." JAMA Psychiatry 73(9): 938-946.

Johnson, K. C., A. K. Smith, Z. N. Stowe, D. J. Newport and P. A. Brennan (2016). "Preschool outcomes following prenatal serotonin reuptake inhibitor exposure: differences in language and behavior, but not cognitive function." J Clin Psychiatry 77(2): e176-182.

Johnson, K. C., J. L. LaPrairie, P. A. Brennan, Z. N. Stowe and D. J. Newport (2012). "Prenatal antipsychotic exposure and neuromotor performance during infancy." Arch Gen Psychiatry 69(8): 787-794.

Juric, S., D. J. Newport, J. C. Ritchie, M. Galanti and Z. N. Stowe (2009). "Zolpidem (Ambien) in pregnancy: placental passage and outcome." Arch Womens Ment Health 12(6): 441-446.

Kallen, B. (1988). "Comments on teratogen update: lithium." Teratology 38(597).

Kallen, B. and A. Tandberg (1983). "Lithium and pregnancy. A cohort study on manic-depressive women." Acta Psychiatr Scand 68(2): 134-139.

Kallen, B., N. Borg and M. Reis (2013). "The use of central nervous system active drugs during pregnancy." Pharmaceuticals (Basel) 6(10): 1221-1286.

Kar, S., M. Wong, E. Rogozinska and S. Thangaratinam (2016). "Effects of omega-3 fatty acids in prevention of early preterm delivery: a systematic review and meta-analysis of randomized studies." Eur J Obstet Gynecol Reprod Biol 198: 40-46.

Kieler, H., M. Artama, A. Engeland, O. Ericsson, K. Furu, M. Gissler, R. B. Nielsen, M. Norgaard, O. Stephansson, U. Valdimarsdottir, H. Zoega and B. Haglund (2012). "Selective serotonin reuptake inhibitors during pregnancy and risk of persistent pulmonary hypertension in the newborn: population based cohort study from the five Nordic countries." Bmj 344: d8012.

Kieviet, N., C. Hoppenbrouwers, K. M. Dolman, J. Berkhof, H. Wennink and A. Honig (2015). "Risk factors for poor neonatal adaptation after exposure to antidepressants in utero." Acta Paediatr 104(4): 384-391.

Kjaer, D., E. Horvath-Puho, J. Christensen, M. Vestergaard, A. E. Czeizel, H. T. Sorensen and J. Olsen (2007). "Use of phenytoin, phenobarbital, or diazepam during pregnancy and risk of congenital abnormalities: a

Laugesen, K., M. S. Olsen, A. B. Telen Andersen, T. Froslev and H. T. Sorensen (2013). "In utero exposure to antidepressant drugs and risk of attention deficit hyperactivity disorder: a nationwide Danish cohort study." BMJ Open 3(9): e003507.

Leiknes, K. A., M. J. Cooke, L. Jarosch-von Schweder, I. Harboe and B. Hoie (2015). "Electroconvulsive therapy during pregnancy: A systematic review of case studies." Archives of Women's Mental Health 18(1): 1-39.

Lin, H. C., I. J. Chen, Y. H. Chen, H. C. Lee and F. J. Wu (2010). "Maternal schizophrenia and pregnancy outcome: does the use of antipsychotics make a difference?" Schizophr Res 116(1): 55-60.

Malm, H., A. S. Brown, M. Gissler, D. Gyllenberg, S. Hinkka-Yli-Salomaki, I. W. McKeague, M. Weissman, P. Wickramaratne, M. Artama, J. A. Gingrich and A. Sourander (2016). "Gestational Exposure to Selective Serotonin Reuptake Inhibitors and Offspring Psychiatric Disorders: A National Register-Based Study." J Am Acad Child Adolesc Psychiatry 55(5): 359-366.

Malm, H., A. Sourander, M. Gissler, D. Gyllenberg, S. Hinkka-Yli-Salomaki, I. W. McKeague, M. Artama and A. S. Brown (2015). "Pregnancy Complications Following Prenatal Exposure to SSRIs or Maternal Psychiatric Disorders: Results From Population-Based National Register Data." Am J Psychiatry 172(12): 1224-1232.

Margulis, A. V., A. Abou-Ali, M. M. Strazzeri, Y. Ding, F. Kuyateh, E. Y. Frimpong, M. S. Levenson and T. A. Hammad (2013). "Use of selective serotonin reuptake inhibitors in pregnancy and cardiac malformations: a propensity-score matched cohort in CPRD." Pharmacoepidemiol Drug Saf 22(9): 942-951.

McDonagh, M., et al. (2014). Antidepressant treatment of depression during pregnancy and the postpartum period, Agency for Healthcare Research and Quality: 1-81.

Moretti, M. E., A. Maxson, F. Hanna and G. Koren (2009). "Evaluating the safety of St. John's wort in human pregnancy." Reprod Toxicol 28(1): 96-99.

Nakhai-Pour, H. R., P. Broy and A. Berard (2010). "Use of antidepressants during pregnancy and the risk of spontaneous abortion." Cmaj 182(10): 1031-1037.

NICE (2015) National Collaborating Centre for Mental Health. Antenatal and postnatal mental health: the NICE guideline on clinical management and service guidance, National Institute for Health and Care Excellence.

Nulman, I., G. Koren, J. Rovet, M. Barrera, D. L. Streiner and B. M. Feldman (2015). "Neurodevelopment of children prenatally exposed to selective reuptake inhibitor antidepressants: Toronto sibling study." J Clin Psychiatry 76(7): e842-847.

Oberlander, T. F., W. Warburton, S. Misri, J. Aghajanian and C. Hertzman (2006). "Neonatal outcomes after prenatal exposure to selective serotonin reuptake inhibitor antidepressants and maternal depression using population-based linked health data." Arch Gen Psychiatry 63(8): 898-906.

Oberlander, T. F., W. Warburton, S. Misri, W. Riggs, J. Aghajanian and C. Hertzman (2008a). "Major congenital malformations following prenatal exposure to serotonin reuptake inhibitors and benzodiazepines using population-based health data." Birth Defects Res B Dev Reprod Toxicol 83(1): 68-76.

Odsbu, I., S. Skurtveit, R. Selmer, C. Roth, S. Hernandez-Diaz and M. Handal (2015). "Prenatal exposure to anxiolytics and hypnotics and language competence at 3 years of age." Eur J Clin Pharmacol 71(3): 283-291.

Petersen, I., C. J. Sammon, R. L. McCrea, D. P. Osborn, S. J. Evans, P. J. Cowen and I. Nazareth (2016a). "Risks associated with antipsychotic treatment in pregnancy: Comparative cohort studies based on electronic health records." Schizophr Res 176(2-3): 349-356.

Petersen, I., S. J. Evans, R. Gilbert, L. Marston and I. Nazareth (2016). "Selective serotonin reuptake inhibitors and congenital heart anomalies: comparative cohort studies of women treated before and during pregnancy and their children." J Clin Psychiatry 77(1): e36-42.

Pompili, M., G. Dominici, G. Giordano, L. Longo, G. Serafini, D. Lester, M. Amore and P. Girardi (2014). "Electroconvulsive treatment during pregnancy: A systematic review." Expert Review of Neurotherapeutics 14(12): 1377-1390.

Rai, D., B. K. Lee, C. Dalman, J. Golding, G. Lewis and C. Magnusson (2013). "Parental depression, maternal antidepressant use during pregnancy, and risk of autism spectrum disorders: population based case-control study." Bmj 346: f2059.

Ramos, E., M. St-Andre, E. Rey, D. Oraichi and A. Berard (2008). "Duration of antidepressant use during pregnancy and risk of major congenital malformations." Br J Psychiatry 192(5): 344-350.

Reis, M. and B. Kallen (2008). "Maternal use of antipsychotics in early pregnancy and delivery outcome." J Clin Psychopharmacol 28(3): 279-288.

Saccone, G., I. Saccone and V. Berghella (2016b). "Omega-3 long-chain polyunsaturated fatty acids and fish oil supplementation during pregnancy: which evidence?" J Matern Fetal Neonatal Med 29(15): 2389-2397.

Simon, G. E., M. L. Cunningham and R. L. Davis (2002). "Outcomes of prenatal antidepressant exposure." Am J Psychiatry 159(12): 2055-2061.

Sorensen, M. J., M. I. Kjaersgaard, H. S. Pedersen, M. Vestergaard, J. Christensen, J. Olsen, E. Parner, L. H. Pedersen and B. H. Bech (2015). "Risk of Fetal Death after Treatment with Antipsychotic Medications during Pregnancy." PLoS One 10(7): e0132280.

Sorensen, M. J., T. K. Gronborg, J. Christensen, E. T. Parner, M. Vestergaard, D. Schendel and L. H. Pedersen (2013). "Antidepressant exposure in pregnancy and risk of autism spectrum disorders." Clin Epidemiol 5: 449-459.

Troyer, W. A., G. R. Pereira, R. A. Lannon, J. Belik and M. C. Yoder (1993). "Association of maternal lithium exposure and premature delivery." J Perinatol 13(2): 123-127.

Vigod, S. N., T. Gomes, A. S. Wilton, V. H. Taylor and J. G. Ray (2015). "Antipsychotic drug use in pregnancy: high dimensional, propensity matched, population based cohort study." Bmj 350: h2298.

Wang, L. H., H. C. Lin, C. C. Lin, Y. H. Chen and H. C. Lin (2010). "Increased risk of adverse pregnancy outcomes in women receiving zolpidem during pregnancy." Clin Pharmacol Ther 88(3): 369-374.

Weston, J., R. Bromley, C. F. Jackson, N. Adab, J. Clayton-Smith, J. Greenhalgh, J. Hounsome, A. J. McKay, C. Tudur Smith and A. G. Marson (2016). "Monotherapy treatment of epilepsy in pregnancy: congenital malformation outcomes in the child." Cochrane Database Syst Rev 11: Cd010224.

Wikner, B. N. and B. Kallen (2011). "Are hypnotic benzodiazepine receptor agonists teratogenic in humans?" J Clin Psychopharmacol 31(3): 356-359.

Wikner, B. N., C. O. Stiller, U. Bergman, C. Asker and B. Kallen (2007). "Use of benzodiazepines and benzodiazepine receptor agonists during pregnancy: neonatal outcome and congenital malformations." Pharmacoepidemiol Drug Saf 16(11): 1203-1210.

Zalzstein, E., Koren, G., Einarson, T. and R. M. Freedom (1990). "A case–control study on the association between first trimester exposure to lithium and Ebstein's anomaly." American Journal of Cardiology 65: 817-818.