

Postnatal care

Supplement 1: Methods

NICE guideline tbc

Development of the guideline and methods

October 2020

Draft for consultation

*Developed by the National Guideline
Alliance part of the Royal College of
Obstetricians and Gynaecologists*

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or their carer or guardian.

Local commissioners and/or providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the [Welsh Government](#), [Scottish Government](#), and [Northern Ireland Executive](#). All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

© NICE 2020. All rights reserved. Subject to [Notice of rights](#).

ISBN:

Contents

Development of the guideline.....	5
Remit.....	5
What this guideline covers.....	5
Groups that are covered.....	5
Key areas that are covered	5
Methods	6
Introduction	6
Developing the review questions and outcomes	6
Searching for evidence.....	9
Scoping search.....	9
Systematic literature search	9
Economic systematic literature search	10
Reviewing evidence.....	10
Systematic review process	10
Type of studies and inclusion/exclusion criteria	11
Methods of combining evidence	12
Appraising the quality of evidence	19
Formal consensus reviews	32
Reviewing economic evidence	34
Inclusion and exclusion of economic studies	35
Consideration of modelling studies that estimated long-term benefits to women and babies and related cost-savings associated with breastfeeding.....	36
Appraising the quality of economic evidence	36
Economic modelling	37
Cost effectiveness criteria	37
Developing recommendations	38
Guideline recommendations.....	38
Research recommendations.....	38
Validation process	38
Updating the guideline.....	38
Funding	39
References.....	40

1 Development of the guideline

2 Remit

- 3 The National Institute for Health and Care Excellence (NICE) commissioned the
4 National Guideline Alliance (NGA) to develop a guideline about postnatal care.

5 What this guideline covers

6 Groups that are covered

- 7 • Women and babies from the birth of the baby until the end of the postnatal period,
8 which, for the purposes of this guideline, is defined as 8 weeks after the birth.
9 Questions on babies' feeding covered all relevant time periods, including the
10 antenatal and postnatal periods.
- 11 • Women having twins and triplets were covered as a subgroup for selected review
12 questions.

13 Key areas that are covered

- 14 • Planning the content and delivery of care.
15 • Identifying and assessing health and wellbeing needs in women.
16 • Identifying and assessing health and wellbeing needs in babies.
17 • Planning and management of babies' feeding.

18

19 For further details of that the guideline does and does not cover see the guideline
20 [scope](#) on the NICE website.

1 Methods

2 Introduction

3 This section summarises methods used to identify and review the evidence, to
4 consider cost effectiveness, and to develop guideline recommendations. This
5 guideline was developed in accordance with methods described in [Developing NICE](#)
6 [guidelines: the manual](#) (NICE 2014). A more up to date version is now available of
7 the NICE manual; however, development of this guideline was initiated before
8 publication of the new manual and to ensure consistent methods the 2014 version
9 was used throughout.

10 Until March 2018, declarations of interest were recorded and managed in accordance
11 with NICE's 2014 conflicts of interest policy. From April 2018, declarations were
12 recorded and managed in accordance with NICE's 2018 [Policy on declaring and](#)
13 [managing interests for NICE advisory committees](#).

14 Developing the review questions and outcomes

15 The review questions considered in this guideline were based on the key areas
16 identified in the guideline [scope](#) .They were drafted by the NGA technical team, and
17 refined and validated by the guideline committee.

18

19 The review questions were based on the following frameworks:

- 20 • intervention reviews – using population, intervention, comparison and outcome
21 (PICO)
- 22 • diagnostic reviews– using population, diagnostic test (index test), reference
23 standard and target condition (PIRT)
- 24 • prognostic reviews – using population, presence or absence of a prognostic, risk
25 or predictive factor and outcome (PPO)
- 26 • qualitative reviews – using population, phenomenon of interest and context.

27 These frameworks guided the development of review protocols, the literature
28 searching process, and critical appraisal and synthesis of evidence. They also
29 facilitated development of recommendations by the committee.

30 Literature searches, critical appraisal and evidence reviews were completed for all
31 review questions.

32 The review questions and evidence reviews corresponding to each question (or
33 group of questions) are summarised in Table 1.

1 **Table 1: Summary of review questions and index to evidence reviews**

Evidence review	Review question	Type of review
[A] Length of stay	<ul style="list-style-type: none"> How does the length of postpartum stay affect women and their babies (single births)? How does length of postpartum stay affect women and their babies (twins or triplets)? 	Intervention
[B] Information transfer	What information needs to be communicated between healthcare professionals at transfer of care from birth care team to community care?	Qualitative
[C] Timing of first contact by midwife	<ul style="list-style-type: none"> When should the first postnatal contact by midwives be made after transfer from place of birth to community care (single births)? When should the first postnatal contact by midwives be made after transfer from place of birth to community care (twins or triplets)? 	Intervention
[D] Timing of first contact by health visitor	When should the first postnatal contact by health visitors be made?	Intervention
[E] Timing of comprehensive assessment	When should a comprehensive, routine assessment of the woman at the end of the postnatal period occur (for example at 6 weeks, 8 weeks or not at all)?	Intervention
[F] Content of postnatal care contacts	What is the essential content of the postnatal care contacts for women and babies?	Formal consensus
[G] Provision of information about the postnatal health of women	When and how should information be given to mothers and their partners about postnatal health of the mother?	Qualitative
[H] Tools for the clinical review of women	What tools for clinical review of women are effective during the first 8 weeks after birth? (including pain scores)	Intervention
[I] Assessment of secondary postpartum haemorrhage	How should early signs and symptoms of postpartum haemorrhage be assessed?	Intervention
[J] Perineal pain	What characteristics of perineal pain suggest the need for further evaluation?	Prognostic
[K] Information on lactation suppression	What information and support should be given to women about lactation suppression? And under what	Qualitative

Evidence review	Review question	Type of review
	circumstances should the information be provided?	
[L1] Signs and symptoms of serious illness in babies	What signs and symptoms (alone or in combination) in babies are associated with serious illness or mortality?	Diagnostic and prognostic
[L2] Scoring systems for illness in babies	Which scoring systems are accurate in identifying or predicting illness severity in babies?	Clinical prediction model and diagnostic
[M] Benefits and harms of bed sharing	What are the benefits and harms of co-sleeping?	Intervention
[N] Co-sleeping risk factors	What are the risk factors in relation to co-sleeping for sudden infant death syndrome (SIDS)?	Prognostic
[O] Emotional attachment	What interventions in the postnatal period are effective at promoting emotional attachment?	Intervention
[P] Breastfeeding interventions ¹	<ul style="list-style-type: none"> • What interventions are effective in starting and maintaining breastfeeding (single births)? • What interventions are effective in starting and maintaining breastfeeding (twins or triplets births)? 	Intervention
[Q] Breastfeeding facilitators and barriers	What are perceived by parents to be the facilitators and barriers for starting and maintaining breastfeeding?	Qualitative
[R] Tools for predicting breastfeeding difficulties	What observations or clinical tools accurately predict breastfeeding difficulties?	Prognostic
[S] Breastfeeding information and support	<ul style="list-style-type: none"> • What information on breastfeeding do parents find helpful (single births)? • What information on breastfeeding do parents find helpful (twins and triplets)? • What support with breastfeeding do parents find helpful (single births)? • What support with breastfeeding do parents find helpful (twins and triplets)? 	Qualitative
[T] Formula feeding information and support	<ul style="list-style-type: none"> • What information on formula feeding do parents find helpful? • What support with formula feeding do parents find helpful? 	Qualitative

1 ¹Original health economic analysis conducted

- 1 Additional information related to development of the guideline is contained in:
- 2 • Supplement 1 (Methods; this document)
- 3 • Supplement 2 (NGA staff list)

4 Searching for evidence

5 Scoping search

6 During the scoping phase, searches were conducted for previous guidelines,
7 economic evaluations, health technology assessments and systematic reviews.
8 Searches of websites of organisations, institutional repositories and internet search
9 engines were also undertaken for relevant policies and related documents, including
10 grey literature.

11 Systematic literature search

12 Systematic literature searches were undertaken to identify published evidence
13 relevant to each review question.

14 Databases were searched using subject headings, free-text terms and, where
15 appropriate, study type filters. Where possible, searches were limited to retrieve
16 studies published in English. All the searches were conducted in the following
17 databases: EMCare, MEDLINE, MEDLINE IN-PROCESS and Embase. For review
18 questions that included systematic reviews and RCTs among the eligibility criteria,
19 the Cochrane Central Register of Controlled Trials (CCTR), Cochrane Database of
20 Systematic Reviews (CDSR), Database of Abstracts of Reviews of Effects (DARE)
21 and the Health Technology Assessment (HTA) database were also searched. For
22 review questions that were highly focused on nursing, the Cumulative Index to
23 Nursing and Allied Health Literature (CINAHL) database was also searched.
24 Additionally, online repositories were searched for questions where guidelines were
25 considered.

26 Where possible, searches were limited to studies published in English.

27

28 Searches were run once for all reviews during development. The guideline committee
29 and the NGA technical team considered the review questions for which the searches
30 might need to be updated, and after prioritising against a number of criteria, made a
31 decision to selectively rerun the searches for evidence reviews A, C, D, G, H, I and
32 O, which were performed between 9-12 weeks in advance of the final guideline
33 committee meetings before consultation on the draft guideline; these reruns were
34 completed during December 2019. Any studies added to the databases after
35 December 2019 (including those published before December 2019 but not yet
36 indexed) were not considered for inclusion.

37 Details of the search strategies, including the study-design filters used and
38 databases searched, are provided in Appendix B of each evidence review.

1

2 **Economic systematic literature search**

3 Systematic literature searches were also undertaken to identify published economic
4 evidence. Databases were searched using subject headings, free-text terms and,
5 where appropriate, an economic evaluations search filter.

6 A single search, using the population search terms used in the evidence reviews,
7 was conducted to identify economic evidence in the NHS Economic Evaluation
8 Database (NHS EED) and HTA. Another single search, using the population search
9 terms used in the evidence reviews, combined with an economic evaluations search
10 filter, was conducted in Emcare, Medline, Medline in Process and Embase.

11 Where wider populations were reviewed, additional searches, using the population
12 and intervention (or equivalent) search terms used in the evidence reviews, were
13 conducted, combined with a search filter for economic evaluations, if and as
14 applicable.

15 Where possible, searches were limited to studies published in English.

16 The systematic economic literature searches were updated during December 2019,
17 11-12 weeks in advance of the final committee meeting before consultation on the
18 draft guideline. Any studies added to the databases after December 2019 (including
19 those published before December 2019 but not yet indexed) were not considered for
20 inclusion.

21 Details of the search strategies, including the study-design filters used and
22 databases searched, are provided in Appendix B of each evidence review, where
23 applicable.

24 **Quality assurance**

25 Search strategies were quality assured by cross-checking reference lists of relevant
26 studies, analysing search strategies from published systematic reviews and asking
27 members of the committee to highlight key studies. The principal search strategies
28 for each search were also quality assured by a second information scientist using an
29 adaptation of the PRESS 2015 Guideline Evidence-Based Checklist
30 (McGowan 2016). In addition, all publications highlighted by stakeholders at the time
31 of the consultation on the draft scope were considered for inclusion.

32 **Reviewing evidence**

33 **Systematic review process**

34 The evidence was reviewed in accordance with the following approach.

- 1 • Potentially relevant articles were identified from the search results for each review
2 question by screening titles and abstracts. Full-text copies of the articles were
3 then obtained.
 - 4 • Full-text articles were reviewed against pre-specified inclusion and exclusion
5 criteria in the review protocol (see Appendix A of each evidence review).
 - 6 • Key information was extracted from each article on study methods and results, in
7 accordance with factors specified in the review protocol. The information was
8 presented in a summary table in the corresponding evidence review and in a more
9 detailed evidence table (see Appendix D of each evidence review).
 - 10 • Included studies were critically appraised using an appropriate checklist as
11 specified in [Developing NICE guidelines: the manual](#) (NICE 2014). Further detail
12 on appraisal of the evidence is provided below.
 - 13 • Summaries of evidence by outcome were presented in the corresponding
14 evidence review and discussed by the committee.
- 15 Review questions selected as priorities for economic analysis and complex review
16 questions were subject to dual screening and study selection through a 10% random
17 sample of articles. Any discrepancies were resolved by discussion between the first
18 and second reviewers or by reference to a third (senior) reviewer. For the remaining
19 review questions, internal (NGA) quality assurance processes included consideration
20 of the outcomes of screening, study selection and data extraction and the committee
21 reviewed the results of study selection and data extraction. The review protocol for
22 each question specifies whether dual screening and study selection was undertaken
23 for that particular question.
- 24 Drafts of all evidence reviews were checked by a senior reviewer.

25 **Type of studies and inclusion/exclusion criteria**

- 26 Inclusion and exclusion of studies was based on criteria specified in the
27 corresponding review protocol.
- 28 Systematic reviews with meta-analyses were considered to be the highest quality
29 evidence that could be selected for inclusion.
- 30 For intervention reviews, randomised controlled trials (RCTs) were prioritised for
31 inclusion because they are considered to be the most robust type of study design
32 that could produce an unbiased estimate of intervention effects. Where there was
33 limited evidence from RCTs, non-randomised studies were considered for inclusion.
- 34 For prognostic reviews, prospective and retrospective cohort and case-control
35 studies and case series were considered for inclusion. Studies that included
36 multivariable analysis were prioritised. Review L1 was designed to include diagnostic
37 and prognostic data and L2 was a clinical prediction model review designed to
38 include both model performance and diagnostic accuracy data. For both reviews
39 cross-sectional studies were considered for inclusion as well as cohort studies.

1 For qualitative reviews, studies using focus groups, structured interviews or semi-
2 structured interviews were considered for inclusion. Where qualitative evidence was
3 sought, data from surveys or other types of questionnaire were considered for
4 inclusion only if they provided data from open-ended questions, but not if they
5 reported only quantitative data.

6 For the formal consensus review (F), published guidelines were included (see formal
7 consensus review section).

8 The committee were consulted about any uncertainty regarding inclusion or exclusion
9 of studies. A list of excluded studies for each review question, including reasons for
10 exclusion is presented in Appendix K of the corresponding evidence review. The
11 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) for
12 the search of clinical evidence is presented in Appendix C of each evidence review.

13 Narrative reviews, posters, letters, editorials, comment articles, unpublished studies
14 and studies published in languages other than English were excluded. Conference
15 abstracts were not considered for inclusion because conference abstracts typically
16 do not have sufficient information to allow for full critical appraisal.

17 **Methods of combining evidence**

18 When planning reviews (through preparation of protocols), the following approaches
19 for data synthesis were discussed and agreed with the committee.

20 **Data synthesis for intervention reviews**

21 ***Pairwise meta-analysis***

22 Meta-analysis to pool results from RCTs was conducted where possible using
23 Cochrane Review Manager (RevMan5) software. Where non-randomised evidence
24 was used, this was not meta-analysed.

25 For dichotomous outcomes, such as mortality, the Mantel–Haenszel method with a
26 random effects model was used to calculate risk ratios (RRs), due to the variability of
27 included interventions. For all outcomes with zero events in both arms the risk
28 difference was presented. For outcomes with low event rates (<1%), Peto odds
29 ratios (ORs) were calculated as this method performs well when events are rare
30 (Bradburn 2007). In meta-analysis, if some (but not all) studies have zero events in
31 both arms, RD was presented to account for study weights, as in evidence review A
32 on length of stay.

33 If a study reported only the summary statistic and 95% confidence interval (CI) the
34 generic-inverse variance method was used to enter data into RevMan5. If the control
35 event rate was reported this was used to generate the absolute risk difference in
36 GRADEpro. If multivariable analysis was used to derive the summary statistic but no
37 adjusted control event rate was reported, no absolute risk difference was calculated.

38 No continuous data was pooled because there was not enough evidence.

1 Subgroups for stratified analyses were agreed for some review questions as part of
2 protocol development.

3 When meta-analysis was undertaken, the results were presented visually using forest
4 plots generated using RevMan5 (see Appendix F of relevant evidence reviews).

5 **Meta-regression**

6 Meta-regression analysis was considered appropriate for evidence review P, the
7 effectiveness of breastfeeding interventions, specifically for the interventions that
8 started in the antenatal period and finished in the postnatal period. Meta-regression
9 is used in meta-analysis to simultaneously investigate the impact of moderator
10 variables on study effect size. In this case meta-regression was considered
11 appropriate because there was a large volume of included studies (n=63) each with
12 different intervention characteristics (or 'moderator variables'), for example where the
13 intervention was delivered, how long it lasted for, how the intervention was delivered
14 and how often.

15 For the purpose of this meta-regression analysis, each study was categorised using
16 the following variables.

- 17 • Number of contact visits: 0, 1, 2-3, 4-8 and 9+.
- 18 • How delivered: face to face on an individual basis, face to face in a group,
19 remote, self-help.
- 20 • Duration of contact: contact with the intervention lasted less than 8 weeks,
21 contact with the intervention lasted more than 8 weeks.
- 22 • Where the intervention was delivered: at home, in a healthcare setting,
23 combination of both home and healthcare setting.

24 The following analyses were conducted for each outcome.

- 25 • Initiation of breastfeeding
 - 26 ○ How delivered
 - 27 ▪ Face to face as an individual versus standard care
 - 28 ▪ Remote versus standard care
 - 29 ▪ Self-help versus standard care
 - 30 ○ Number of contacts
 - 31 ▪ 0-1 versus standard care
 - 32 ▪ 2-3 versus standard care
 - 33 ▪ 4-8 versus standard care
 - 34 ▪ 9+ versus standard care
 - 35 ○ Duration of contact
 - 36 ▪ Less than 8 weeks versus standard care
 - 37 ▪ More than 8 weeks versus standard care
 - 38 ○ Where delivered
 - 39 ▪ Healthcare setting versus standard care
 - 40 ▪ Home setting versus standard care
 - 41 ▪ Both healthcare and home setting versus standard care

- 1 ○ A final model including number of contacts, how delivered and where
- 2 delivered
- 3 ▪ 2-3 contacts versus 0-1 contacts
- 4 ▪ 4-8 contacts versus 0-1 contacts
- 5 ▪ 9+ contacts versus 0-1 contacts
- 6 ▪ Face to face as an individual versus standard care
- 7 ▪ Self-help versus standard care
- 8 ▪ Healthcare setting versus both healthcare and home setting
- 9 ▪ Home setting versus both healthcare and home setting

10 NB. Duration of contact was not included in the final model because there was not
11 enough data, furthermore, duration of contacts was considered to be captured by the
12 number of contacts.

- 13 • Any breastfeeding at 3-14 days
- 14 ○ How delivered
- 15 ▪ Face to face as an individual versus standard care
- 16 ▪ Face to face as part of a group versus standard care
- 17 ▪ Remote versus standard care
- 18 ▪ Self-help versus standard care
- 19 ○ Number of contacts
- 20 ▪ 0 versus standard care
- 21 ▪ 1 versus standard care
- 22 ▪ 2-3 versus standard care
- 23 ▪ 4-8 versus standard care
- 24 ▪ 9+ versus standard care
- 25 ○ Duration of contact
- 26 ▪ Less than 8 weeks versus standard care
- 27 ▪ More than 8 weeks versus standard care
- 28 ○ Where delivered
- 29 ▪ Healthcare setting versus standard care
- 30 ▪ Home setting versus standard care
- 31 ▪ Both healthcare and home setting versus standard care
- 32 ○ A final model including number of contacts, how delivered and where
- 33 delivered
- 34 ▪ 1 contact versus 0 contacts
- 35 ▪ 2-3 contacts versus 0 contacts
- 36 ▪ 4-8 contacts versus 0 contacts
- 37 ▪ 9+ contacts versus 0 contacts
- 38 ▪ Face to face as an individual versus standard care
- 39 ▪ Face to face as part of a group versus standard care
- 40 ▪ Remote versus standard care
- 41 ▪ Self-help versus standard care
- 42 ▪ Healthcare setting versus both healthcare and home setting
- 43 ▪ Home setting versus both healthcare and home setting
- 44
- 45 • Exclusive breastfeeding at 3-14 days

- 1 ○ How delivered
- 2 ▪ Face to face as an individual versus standard care
- 3 ▪ Face to face as part of a group versus standard care
- 4 ▪ Remote versus standard care
- 5 ▪ Self-help versus standard care
- 6 ○ Number of contacts
- 7 ▪ 0 versus standard care
- 8 ▪ 1 versus standard care
- 9 ▪ 2-3 versus standard care
- 10 ▪ 4-8 versus standard care
- 11 ▪ 9+ versus standard care
- 12 ○ Duration of contact
- 13 ▪ Less than 8 weeks versus standard care
- 14 ▪ More than 8 weeks versus standard care
- 15 ○ Where delivered
- 16 ▪ Healthcare setting versus standard care
- 17 ▪ Home setting versus standard care
- 18 ▪ Both healthcare and home setting versus standard care
- 19 ○ A final model including number of contacts, how delivered and where
- 20 delivered
- 21 ▪ 1 contact versus 0 contacts
- 22 ▪ 2-3 contacts versus 0 contacts
- 23 ▪ 4-8 contacts versus 0 contacts
- 24 ▪ 9+ contacts versus 0 contacts
- 25 ▪ Face to face as an individual versus standard care
- 26 ▪ Face to face as part of a group versus standard care
- 27 ▪ Remote versus standard care
- 28 ▪ Self-help versus standard care
- 29 ▪ Healthcare setting versus both healthcare and home setting
- 30 ▪ Home setting versus both healthcare and home setting
- 31
- 32 • Any breastfeeding at 6-12 weeks
- 33 ○ How delivered
- 34 ▪ Face to face as an individual versus standard care
- 35 ▪ Face to face as part of a group versus standard care
- 36 ▪ Remote versus standard care
- 37 ▪ Self-help versus standard care
- 38 ○ Number of contacts
- 39 ▪ 0 versus standard care
- 40 ▪ 1 versus standard care
- 41 ▪ 2-3 versus standard care
- 42 ▪ 4-8 versus standard care
- 43 ▪ 9+ versus standard care
- 44 ○ Duration of contact
- 45 ▪ Less than 8 weeks versus standard care
- 46 ▪ More than 8 weeks versus standard care
- 47 ○ Where delivered

- 1 ▪ Healthcare setting versus standard care
- 2 ▪ Home setting versus standard care
- 3 ▪ Both healthcare and home setting versus standard care
- 4 ○ Model results
- 5 ▪ 1 contact versus 0 contacts
- 6 ▪ 2-3 contacts versus 0 contacts
- 7 ▪ 4-8 contacts versus 0 contacts
- 8 ▪ 9+ contacts versus 0 contacts
- 9 ▪ Face to face as an individual versus standard care
- 10 ▪ Face to face as part of a group versus standard care
- 11 ▪ Remote versus standard care
- 12 ▪ Self-help versus standard care
- 13 ▪ Healthcare setting versus both healthcare and home setting
- 14 ▪ Home setting versus both healthcare and home setting
- 15
- 16 • Exclusive breastfeeding at 6-12 weeks
- 17 ○ How delivered
- 18 ▪ Face to face as an individual versus standard care
- 19 ▪ Face to face as part of a group versus standard care
- 20 ▪ Remote versus standard care
- 21 ▪ Self-help versus standard care
- 22 ○ Number of contacts
- 23 ▪ 0 versus standard care
- 24 ▪ 1 versus standard care
- 25 ▪ 2-3 versus standard care
- 26 ▪ 4-8 versus standard care
- 27 ▪ 9+ versus standard care
- 28 ○ Duration of contact
- 29 ▪ Less than 8 weeks versus standard care
- 30 ▪ More than 8 weeks versus standard care
- 31 ○ Where delivered
- 32 ▪ Healthcare setting versus standard care
- 33 ▪ Home setting versus standard care
- 34 ▪ Both healthcare and home setting versus standard care
- 35 ○ A final model including number of contacts, how delivered and where
- 36 delivered
- 37 ▪ 1 contact versus 0 contacts
- 38 ▪ 2-3 contacts versus 0 contacts
- 39 ▪ 4-8 contacts versus 0 contacts
- 40 ▪ 9+ contacts versus 0 contacts
- 41 ▪ Face to face as an individual versus standard care
- 42 ▪ Face to face as part of a group versus standard care
- 43 ▪ Remote versus standard care
- 44 ▪ Self-help versus standard care
- 45 ▪ Healthcare setting versus both healthcare and home setting
- 46 ▪ Home setting versus both healthcare and home setting
- 47

- 1 • Any breastfeeding at 16 to 24 weeks:
 - 2 ○ How delivered
 - 3 ▪ Face to face as an individual versus standard care
 - 4 ▪ Face to face as part of a group versus standard care
 - 5 ▪ Remote versus standard care
 - 6 ▪ Self-help versus standard care
 - 7 ○ Number of contacts
 - 8 ▪ 0 versus standard care
 - 9 ▪ 1 versus standard care
 - 10 ▪ 2-3 versus standard care
 - 11 ▪ 4-8 versus standard care
 - 12 ▪ 9+ versus standard care
 - 13 ○ Duration of contact
 - 14 ▪ Less than 8 weeks versus standard care
 - 15 ▪ More than 8 weeks versus standard care
 - 16 ○ Where delivered
 - 17 ▪ Healthcare setting versus standard care
 - 18 ▪ Home setting versus standard care
 - 19 ▪ Both healthcare and home setting versus standard care
 - 20 ○ A final model including number of contacts, how delivered and where
 - 21 delivered
 - 22 ▪ 1 contact versus 0 contacts
 - 23 ▪ 2-3 contacts versus 0 contacts
 - 24 ▪ 4-8 contacts versus 0 contacts
 - 25 ▪ 9+ contacts versus 0 contacts
 - 26 ▪ Face to face as an individual versus standard care
 - 27 ▪ Face to face as part of a group versus standard care
 - 28 ▪ Remote versus standard care
 - 29 ▪ Self-help versus standard care
 - 30 ▪ Healthcare setting versus both healthcare and home setting
 - 31 ▪ Home setting versus both healthcare and home setting

32 Sample Win BUGS code for the analysis of any breastfeeding at 16 to 26 weeks,
33 including the variables how the intervention was delivered, the number of contacts for
34 the intervention and where the intervention was delivered is given in evidence report
35 P, appendix M. Other analyses used the same substantive code as the one provided,
36 but was modified to include the relevant predictor variables for the model under
37 consideration.

38 See evidence report P for further details of the meta-regression and results.

39 **Data synthesis for reviews of diagnostic test accuracy and prediction tools**

40 When diagnostic test accuracy was measured dichotomously, sensitivity, specificity,
41 and positive and negative likelihood ratios were used as outcomes. These diagnostic
42 test accuracy parameters were calculated by the NGA technical team using data
43 reported in the articles.

1 Meta-analysis of diagnostic test accuracy parameters was conducted, using the
2 statistical software STATA (v13), if there were data from four or more studies that
3 could be pooled.

4 **Data synthesis for prognostic reviews**

5 ORs or RRs with 95% CIs reported in published studies were extracted or calculated
6 by the NGA technical team to examine relationships between risk factors and
7 outcomes of interest. Ideally analyses would have adjusted for key confounders
8 (such as age) to be considered for inclusion. Recognising variation across studies in
9 terms of populations, risk factors, outcomes and statistical analysis methods
10 (including adjustments for confounding factors), prognostic data were not pooled, but
11 results from individual studies were presented in the evidence reviews.

12 Additional calculations were required to be undertaken by the NGA technical team for
13 evidence review N, co-sleeping risk factors. The aim of this review was to identify risk
14 factors in relation to co-sleeping for sudden unexpected death in infancy. The
15 exposure of interest was co-sleeping with a risk factor and the reference standard
16 was co-sleeping. Where the papers reported the OR for the exposures of interest
17 (co-sleeping with a risk factor) against a reference standard not of interest (typically
18 'not co-sleeping') and also within the same paper, the OR for co-sleeping with no
19 additional risk factor against the same reference standard, the NGA technical team
20 calculated the added risk of co-sleeping with a risk factor compared to co-sleeping by
21 using an equation by Franchini (2012) which calculates the difference between the
22 two ORs, using similar methods to that of indirect treatment comparisons. For
23 example, if the paper reports A versus B and also C versus B and we are interested
24 in A versus C, the Franchini method can be used to calculate the added risk.

25 **Data synthesis for qualitative reviews**

26 Where possible, a meta-synthesis was conducted to combine evidence from
27 qualitative studies. Whenever a qualitative theme relevant to the protocol was
28 identified, these were extracted along with supporting quotes where available. When
29 all themes had been extracted from studies, common concepts were categorised and
30 tabulated and usually broken down into themes and subthemes, depending on the
31 complexity of the data. Information about how many studies contributed to each
32 theme or subtheme was also captured.

33 In qualitative synthesis, a theme being reported more than other themes across
34 included studies does not necessarily mean that the theme is more important than
35 other themes. The aim of qualitative research is to identify new perspectives on a
36 particular topic. Study types and populations in qualitative research can differ widely,
37 meaning that themes identified by just one or a few studies can provide important
38 new information on a given topic. Therefore, for the purpose of the qualitative reviews
39 in this guideline, it was planned that further studies would not be added when they
40 reported the same themes as had already been identified from other studies because
41 the emphasis was to be on conceptual robustness rather than quantitative
42 completeness of the evidence. This would have implications for the types and

1 numbers of studies included in the qualitative reviews, with study inclusion continuing
2 until no new relevant data could be found regarding a topic that would add to or
3 refute it. This concept is referred to in the literature as ‘theoretical saturation’ (Dixon-
4 Woods 2005). In this guideline theoretical saturation was reached for 4 qualitative
5 reviews (described in evidence reports G, Q, S and T).

6 Themes from individual studies were integrated into a wider context and, when
7 possible, overarching categories of themes with sub-themes were identified. Themes
8 were derived from data presented in individual studies. When themes were extracted
9 from 1 primary study only, theme names used in the guideline mirrored those in the
10 source study. However, when themes were based on evidence from multiple studies,
11 the theme names were assigned by the NGA technical team. The names of
12 overarching categories of themes were also assigned by the NGA technical team.

13 Themes and subthemes were listed in the evidence reports for the committee to
14 understand the relationship between them.

15 **Appraising the quality of evidence**

16 **Intervention studies**

17 ***Pairwise meta-analysis***

18 **GRADE methodology for intervention reviews**

19 For intervention reviews, the evidence for outcomes from included RCTs and
20 comparative non-randomised studies was evaluated and presented using the
21 Grading of Recommendations Assessment, Development and Evaluation (GRADE)
22 methodology developed by the international [GRADE working group](#). More
23 information about this tool can be found on the [developer’s website](#).

24 When GRADE was applied, software developed by the GRADE working group
25 (GRADEpro) was used to assess the quality of each outcome, taking account of
26 individual study quality factors and any meta-analysis results. Results were
27 presented in GRADE profiles (GRADE tables). The clinical evidence profile tables
28 include details of the quality assessment and pooled outcome data, where
29 appropriate, a relative and an absolute measure of intervention effect and the
30 summary of quality of evidence for that outcome. In this table, the columns for
31 intervention and control indicate the sum across studies of the number of participants
32 in each arm for continuous outcomes and frequency of events (n/N; the sum across
33 studies of the number of participants with events divided by sum of the number of
34 participants) for dichotomous outcomes.

35 The selection of outcomes for each review question was agreed during development
36 of the associated review protocol in discussion with the committee. The evidence for
37 each outcome was examined separately for the quality elements summarised in
38 Table 2. Criteria considered in the rating of these elements are discussed below.
39 Each element was graded using the quality ratings summarised in Table 3. Footnotes

1 to GRADE tables were used to record reasons for grading a particular quality
 2 element as having a 'serious' or 'very serious' quality issue. The ratings for each
 3 component were combined to obtain an overall assessment of quality for each
 4 outcome as described in Table 4.

5 The initial quality rating was based on the study design: RCTs started as 'high'
 6 quality evidence, non-randomised studies started as 'low' quality evidence. The
 7 rating was then modified according to the assessment of each quality element (Table
 8 2). Each quality element considered to have a 'serious' or 'very serious' quality issue
 9 was downgraded by 1 or 2 levels respectively (for example, evidence starting as
 10 'high' quality was downgraded to 'moderate' or 'low' quality). In addition, there was a
 11 possibility to upgrade evidence from non-randomised studies (provided the evidence
 12 for that outcome had not previously been downgraded) if there was a large
 13 magnitude of effect, a dose–response gradient, or if all plausible confounding would
 14 reduce a demonstrated effect or suggest a spurious effect when results showed no
 15 effect.

16 **Table 2: Summary of quality elements in GRADE for intervention reviews**

Quality element	Description
Risk of bias ('Study limitations')	This refers to limitations in study design or implementation that reduce the internal validity of the evidence
Inconsistency	This refers to unexplained heterogeneity in the results
Indirectness	This refers to differences in study populations, interventions, comparators or outcomes between the available evidence and inclusion criteria specified in the review protocol
Imprecision	This occurs when a study has few participants or few events of interest, resulting in wide confidence intervals that cross minimally important thresholds
Publication bias	This refers to systematic under- or over-estimation of the underlying benefit or harm resulting from selective publication of study results

17 **Table 3: GRADE quality ratings (by quality element)**

Quality issues	Description
None or not serious	No serious issues with the evidence for the quality element under consideration
Serious	Issues with the evidence sufficient to downgrade by 1 level for the quality element under consideration
Very serious	Issues with the evidence sufficient to downgrade by 2 levels for the quality element under consideration

1 **Table 4: Overall quality of the evidence in GRADE (by outcome)**

Overall quality grading	Description
High	Further research is very unlikely to change the level of confidence in the estimate of effect
Moderate	Further research is likely to have an important impact on the level of confidence in the estimate of effect and may change the estimate
Low	Further research is very likely to have an important impact on the level of confidence in the estimate of effect and is likely to change the estimate
Very low	The estimate of effect is very uncertain

2 *Assessing risk of bias in intervention reviews*

3 Bias is a systematic error, or consistent deviation from the truth in results obtained.
 4 When a risk of bias is present the true effect can be either under- or over-estimated.

5 Risk of bias in RCTs was assessed using the Cochrane risk of bias tool (see
 6 Appendix H in [Developing NICE guidelines: the manual](#); NICE 2014). The original
 7 version of the tool was updated in August 2019 and from this point, version 2 was
 8 used to assess included RCTs, with the result that it was used in review P, breast
 9 feeding interventions.

10 The Cochrane risk of bias tool assesses the following possible sources of bias:

- 11 • selection bias
- 12 • performance bias
- 13 • attrition bias
- 14 • detection bias
- 15 • reporting bias.

16 A study with a poor methodological design does not automatically imply high risk of
 17 bias; the bias is considered individually for each outcome and it is assessed whether
 18 the chosen design and methodology will impact on the estimation of the intervention
 19 effect. It is sometimes argued that if the nature or design of a study makes it difficult
 20 or unethical to reduce the risk of bias, for example through blinding participants, then
 21 the study should not be penalised (downgraded). However, the NGA technical team
 22 took the view that a lack of blinding should lead to downgrading on performance bias,
 23 because regardless of the reason, a lacking of blinding could still potentially lead to
 24 deviations from the intended interventions and bias in measurement of the outcomes.
 25 The same standard was applied across all reviews.

26 More details about the Cochrane risk of bias tool can be found in Section 8 of the
 27 [Cochrane Handbook for Systematic Reviews of Interventions](#) (Higgins and Thomas
 28 2019).

1 For systematic reviews of RCTs the AMSTAR checklist was used and for systematic
2 reviews of other study types the ROBIS checklist was used (see Appendix H in
3 [Developing NICE guidelines: the manual](#); NICE 2014).

4 For non-randomised studies the ROBINS-I checklist was used (see Appendix H in
5 [Developing NICE guidelines: the manual](#); NICE 2014).

6 *Assessing inconsistency in intervention reviews*

7 Inconsistency refers to unexplained heterogeneity in results of meta-analysis. When
8 estimates of treatment effect vary widely across studies (that is, there is
9 heterogeneity or variability in results), this suggests true differences in underlying
10 effects. Inconsistency is, thus, only truly applicable when statistical meta-analysis is
11 conducted (that is, results from different studies are pooled). When outcomes were
12 derived from a single study the rating 'no serious inconsistency' was used when
13 assessing this domain, as per GRADE methodology (Santesso 2016).

14 Inconsistency was assessed visually by inspecting forest plots and observing
15 whether there was considerable heterogeneity in the results of the meta-analysis (for
16 example if the point estimates of the individual studies consistently showed benefits
17 or harms). This was supported by calculating the I-squared statistic for the meta-
18 analysis with an I-squared value of more than 50% indicating considerable
19 heterogeneity, and more than 75% indicating very serious heterogeneity. When
20 considerable or very serious heterogeneity was observed, possible reasons were
21 explored and subgroup analyses were performed as pre-specified in the review
22 protocol where possible. In the case of unexplained heterogeneity, sensitivity
23 analyses were planned based on the quality of studies, eliminating studies at high
24 risk of bias (in relation to randomisation, allocation concealment and blinding, and/or
25 missing outcome data).

26 When no plausible explanation for the heterogeneity could be found, the quality of
27 the evidence was downgraded in GRADE for inconsistency.

28 *Assessing indirectness in intervention reviews*

29 Directness refers to the extent to which populations, interventions, comparisons and
30 outcomes reported in the evidence are similar to those defined in the inclusion
31 criteria for the review and was assessed by comparing the PICO elements in the
32 studies to the PICO defined in the review protocol. Indirectness is important when
33 such differences are expected to contribute to a difference in effect size, or may
34 affect the balance of benefits and harms considered for an intervention.

35 *Assessing imprecision and importance in intervention reviews*

36 Imprecision in GRADE methodology refers to uncertainty around the effect estimate
37 and whether or not there is an important difference between interventions (that is,
38 whether the evidence clearly supports a particular recommendation or appears to be
39 consistent with several candidate recommendations). Therefore, imprecision differs
40 from other aspects of evidence quality because it is not concerned with whether the

1 point estimate is accurate or correct (has internal or external validity). Instead, it is
2 concerned with uncertainty about what the point estimate actually represents. This
3 uncertainty is reflected in the width of the CI.

4 The 95% CI is defined as the range of values within which the population value will
5 fall on 95% of repeated samples, were the procedure to be repeated. The larger the
6 study, the smaller the 95% CI will be and the more certain the effect estimate.

7 Imprecision was assessed in the guideline evidence reviews by considering whether
8 the width of the 95% CI of the effect estimate was relevant to decision making,
9 considering each outcome independently. This is illustrated in Figure 1, which
10 considers a positive outcome for the comparison of two treatments. Three decision-
11 making zones can be differentiated, bounded by the thresholds for minimal
12 importance (minimally important differences; MIDs) for benefit and harm.

13 When the CI of the effect estimate is wholly contained in 1 of the 3 zones there is no
14 uncertainty about the size and direction of effect, therefore, the effect estimate is
15 considered precise; that is, there is no imprecision.

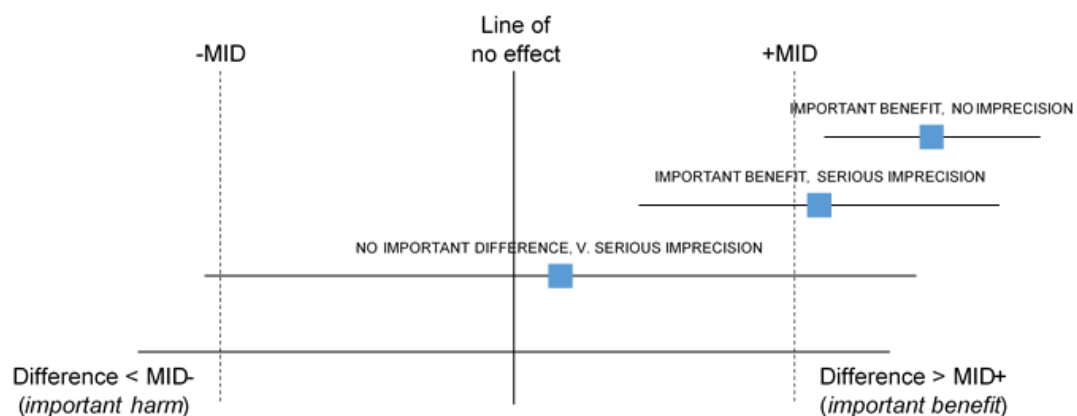
16 When the CI crosses 2 zones, it is uncertain in which zone the true value of the effect
17 estimate lies and therefore there is uncertainty over which decision to make. The CI
18 is consistent with 2 possible decisions, therefore, the effect estimate is considered to
19 be imprecise in the GRADE analysis and the evidence is downgraded by 1 level
20 ('serious imprecision').

21 When the CI crosses all 3 zones, the effect estimate is considered to be very
22 imprecise because the CI is consistent with 3 possible decisions and there is
23 therefore a considerable lack of confidence in the results. The evidence is therefore
24 downgraded by 2 levels in the GRADE analysis ('very serious imprecision').

25 Implicitly, assessing whether a CI is in, or partially in, an important zone, requires the
26 guideline committee to estimate an MID or to say whether they would make different
27 decisions for the 2 confidence limits.

28 Exceptions to the above approach are described in the section below.

1 **Figure 1: Assessment of imprecision and importance in intervention reviews**
 2 **using GRADE**



3
 4 *MID, minimally important difference*

5 *Defining minimally important differences for intervention reviews*

6 The committee was asked whether there were any recognised or acceptable MID in
 7 the published literature and community relevant to the review questions under
 8 consideration. The committee was not aware of any MID that could be used for the
 9 guideline.

10 In the absence of published or accepted MID, the committee agreed to use the
 11 GRADE default MID to assess imprecision for most outcomes. For dichotomous
 12 outcomes minimally important thresholds for a RR of 0.8 and 1.25 respectively were
 13 used as default MID in most of the evidence reviews. The committee also chose to
 14 use 0.8 and 1.25 as the MID for OR in the absence of published or accepted MID.
 15 OR were predominantly used in the guideline when Peto OR were indicated due to
 16 low event rates, as low event rates OR are mathematically similar to RR making the
 17 extrapolation appropriate. There were some exceptions when the default MID were
 18 not used and these are outlined below.

19 When there were zero events in both arms, for which the risk difference was
 20 presented, as in review A. In this context imprecision cannot be assessed against
 21 relative CI (such as 0.8 to 1.25) so instead, a sample size rule (akin to optimal
 22 information size criteria) of 300 was used, with MID as the null effect. Outcomes were
 23 downgraded once if 1 of these applied and twice if they both applied to the risk
 24 difference. If risk difference was used for meta-analysis, for example if the majority of
 25 studies had zero events in either arm, imprecision was assessed based on sample
 26 size using 300 and 500 as cut-offs for very serious and serious imprecision
 27 respectively. The committee used these numbers based on commonly used optimal
 28 information size thresholds.

29 Where the committee felt that any improvement represented an important difference
 30 (as in breastfeeding outcomes in review P and infant mortality in review N) the line of

1 no statistically significant effect was applied. That is, any statistically significant
 2 change was considered to be important in practice and in that case, there was no
 3 imprecision. If there was no statistically significant effect (that is that the line of null
 4 effect was crossed), the effect estimate was considered to have serious imprecision.
 5 If the point estimate was greater than the MID and the 95% CI crossed the line of no
 6 effect but the 90% CI did not, it was judged that while there may be an important
 7 effect, there is uncertainty around the estimate.

8 For continuous outcomes default MIDs are equal to half the median standard
 9 deviation (SD) of the control groups at baseline (or at follow-up if the SD is not
 10 available at baseline).

11 *Assessing publication bias in intervention reviews*

12 Where 10 or more studies were included as part of a single meta-analysis, a funnel
 13 plot was produced to graphically assess the potential for publication bias. Where
 14 fewer than 10 studies were included for an outcome, the committee subjectively
 15 assessed the likelihood of publication bias based on factors such as the proportion of
 16 trials funded by industry and the propensity for publication bias in the topic area.

17 **Diagnostic reviews**

18 ***Adapted GRADE methodology for diagnostic reviews***

19 For diagnostic reviews an adapted GRADE approach was used. GRADE
 20 methodology is designed for intervention reviews but the quality assessment
 21 elements and outcome presentation were adapted by the guideline developers for
 22 diagnostic test accuracy reviews. For example, GRADE tables were modified to
 23 include diagnostic test accuracy measures (sensitivity, specificity and likelihood
 24 ratios).

25 The evidence for each outcome in the diagnostic reviews was examined separately
 26 for the quality elements listed and defined in Table 5. The criteria considered in the
 27 rating of these elements are discussed below. Each element was graded using the
 28 quality levels summarised in Table 3. Footnotes to GRADE tables were used to
 29 record reasons for grading a particular quality element as having a 'serious' or 'very
 30 serious' quality issue. The ratings for each component were combined to obtain an
 31 overall assessment of quality for each outcome as described in Table 4.

32 The initial quality rating was based on the study design: cross-sectional or cohort
 33 studies start as 'high' quality.

34 **Table 5: Adaptation of GRADE quality elements for diagnostic reviews**

Quality element	Description
Risk of bias ('Study limitations')	Limitations in study design and implementation may bias estimates of diagnostic accuracy. High risk of bias for the majority of the evidence reduces confidence in the estimated effect. Diagnostic accuracy studies are not usually randomised and therefore would

Quality element	Description
	not be downgraded for study design from the outset (they start as high quality)
Inconsistency	This refers to unexplained heterogeneity in test accuracy measures (such as sensitivity and specificity) between studies
Indirectness	This refers to differences in study populations, index tests, reference standards or outcomes between the available evidence and inclusion criteria specified in the review protocol
Imprecision	This occurs when a study has relatively few participants and the probability of a correct diagnosis is low. Accuracy measures would therefore have wide confidence intervals around the estimated effect

1 *Assessing risk of bias in diagnostic reviews*

2 Risk of bias in diagnostic reviews was assessed using the Quality Assessment of
3 Diagnostic Accuracy Studies version 2 (QUADAS-2) checklist (see Appendix H in
4 [Developing NICE guidelines: the manual](#); NICE 2014).

5 Risk of bias in primary diagnostic accuracy reviews in QUADAS-2 consists of 4
6 domains:

- 7 • participant selection
- 8 • index test
- 9 • reference standard
- 10 • flow and timing.

11 More details about the QUADAS-2 tool can be found on the [developer's website](#).

12 *Assessing inconsistency in diagnostic reviews*

13 Inconsistency refers to the unexplained heterogeneity of the results in meta-analysis.
14 When estimates of diagnostic accuracy vary widely across studies (that is, there is
15 heterogeneity or variability in results), this suggests true differences in underlying
16 effects. Inconsistency is, thus, only truly applicable when statistical meta-analysis is
17 conducted (that is, results from different studies are pooled).

18 Inconsistency for diagnostic reviews was assessed visually by inspecting forest plots
19 and observing whether there was considerable heterogeneity in the results of the
20 meta-analysis. This was supported by calculating the I-squared statistic for the meta-
21 analysis with an I-squared value of more than 50% indicating considerable
22 heterogeneity, and more than 80% indicating very serious heterogeneity. If there was
23 considerable or very serious heterogeneity the evidence was downgraded for
24 inconsistency.

25

1 *Assessing indirectness in diagnostic reviews*

2 Indirectness in diagnostic reviews was assessed using the QUADAS-2 checklist by
3 assessing the applicability of the studies in relation to the review question in the
4 following domains:

- 5 • participant selection
- 6 • index test
- 7 • reference standard.

8 More details about the QUADAS-2 tool can be found on the [developer's website](#).

9 *Assessing imprecision and importance in diagnostic reviews*

10 The judgement of precision for diagnostic evidence was based on the CIs of the
11 sensitivity and specificity. The committee defined 2 decision thresholds for each
12 measure, a value above which the test could be recommended and a value below
13 which the test would be considered of no use. These thresholds were based on the
14 committee's experience and consensus.

15 The thresholds were:

- 16 • sensitivity: low threshold 75%, high threshold 90%
- 17 • specificity: low threshold 75%, high threshold 90%.

18 Outcomes were downgraded for imprecision when their 95% CI crossed at least 1
19 threshold. If the CI crossed 1 threshold, the outcome was downgraded once for
20 imprecision. If the CI crossed 2 thresholds, the outcome was downgraded twice for
21 imprecision. These assessments were made on the meta-analysed outcomes where
22 applicable or if outcomes were not meta-analysed, on the individual study results
23 themselves.

24 In review L2, the following cut-offs were used when summarising the performance of
25 the scoring systems:

- 26 • very useful test: sensitivity $\geq 90\%$
- 27 • moderately useful test: sensitivity 75% to 89%
- 28 • not a useful test: sensitivity $\leq 75\%$

29 **Prognostic studies**

30 ***Adapted GRADE methodology for prognostic reviews***

31 For prognostic reviews with evidence from comparative studies an adapted GRADE
32 approach was used. As noted above, GRADE methodology is designed for
33 intervention reviews but the quality assessment elements were adapted for
34 prognostic reviews. Adapted GRADE was not used for evidence from case series;
35 instead quality of case series evidence was assessed using the Checklist for Case
36 Series developed by the Joanna Briggs Institute. More information about this tool can
37 be found on the [developer's website](#).

1 The evidence for each outcome in the prognostic reviews was examined separately
 2 for the quality elements listed and defined in Table 6. The criteria considered in the
 3 rating of these elements are discussed below. Each element was graded using the
 4 quality levels summarised in Table 3. Footnotes to GRADE tables were used to
 5 record reasons for grading a particular quality element as having 'serious' or 'very
 6 serious' quality issues. The ratings for each component were combined to obtain an
 7 overall assessment of quality for each outcome as described in Table 4.

8 **Table 6: Adaptation of GRADE quality elements for prognostic reviews**

Quality element	Description
Risk of bias ('Study limitations')	Limitations in study design and implementation may bias estimates and interpretation of the effect of the prognostic/risk factor. High risk of bias for the majority of the evidence reduces confidence in the estimated effect. Prognostic studies are not usually randomised and therefore would not be downgraded for study design from the outset (they start as high quality)
Inconsistency	This refers to unexplained heterogeneity between studies looking at the same prognostic/risk factor, resulting in wide variability in estimates of association (such as RRs or ORs), with little or no overlap in confidence intervals
Indirectness	This refers to any departure from inclusion criteria listed in the review protocol (such as differences in study populations or prognostic/risk factors), that may affect the generalisability of results
Imprecision	This occurs when a study has relatively few participants and also when the number of participants is too small for a multivariable analysis (as a rule of thumb, 10 participants are needed per variable). This was assessed by considering the confidence interval in relation to the point estimate for each outcome reported in the included studies

9 *RR, relative risk; OR, odds ratio*

10 *Assessing risk of bias in prognostic reviews*

11 The Quality in Prognosis Studies (QUIPS) tool developed by Hayden 2013 was used
 12 to assess risk of bias in studies included in prognostic reviews (see Appendix H in
 13 the [Developing NICE guidelines: the manual](#); NICE 2014). The risk of bias in each
 14 study was determined by assessing the following domains:

- 15 • selection bias
- 16 • attrition bias
- 17 • prognostic factor bias
- 18 • outcome measurement bias
- 19 • control for confounders
- 20 • appropriate statistical analysis.

1 *Assessing inconsistency in prognostic reviews*

2 Where multiple results were deemed appropriate to meta-analyse (that is, there was
3 sufficient similarity between risk factor and outcome under investigation)
4 inconsistency was assessed by visually inspecting forest plots and observing
5 whether there was considerable heterogeneity in the results of the meta-analysis.
6 This was assessed by calculating the I-squared statistic for the meta-analysis with an
7 I-squared value of more than 50% indicating considerable heterogeneity, and more
8 than 80% indicating very serious heterogeneity. When considerable or very serious
9 heterogeneity was observed, possible reasons were explored and subgroup analyses
10 were performed as pre-specified in the review protocol where possible.

11 When no plausible explanation for the heterogeneity could be found, the quality of
12 the evidence was downgraded in GRADE for inconsistency.

13 *Assessing indirectness in prognostic reviews*

14 Indirectness in prognostic reviews was assessed by comparing the populations,
15 prognostic factors and outcomes in the evidence to those defined in the review
16 protocol.

17 *Assessing imprecision and importance in prognostic reviews*

18 Prognostic studies may have a variety of purposes, for example, establishing typical
19 prognosis in a broad population, establishing the effect of patient characteristics on
20 prognosis, and developing a prognostic model. While by convention MIDs relate to
21 intervention effects, the committee agreed to use GRADE default MIDs for
22 dichotomous outcomes (RR) as a starting point from which to assess whether the
23 size of an outcome effect in a prognostic study would be large enough to be
24 meaningful in practice.

25 **Qualitative reviews**

26 ***GRADE-CERQual methodology for qualitative reviews***

27 For qualitative reviews an adapted GRADE Confidence in the Evidence from
28 Reviews of Qualitative research (GRADE-CERQual) approach (Lewin 2015) was
29 used. In this approach the quality of evidence is considered according to themes in
30 the evidence. The themes may have been identified in the primary studies or they
31 may have been identified by considering the reports of a number of studies. Quality
32 elements assessed using GRADE-CERQual are listed and defined in Table 7. Each
33 element was graded using the levels of concern summarised in Table 8. The ratings
34 for each component were combined (as with other types of evidence) to obtain an
35 overall assessment of quality for each theme as described in Table 9.

1 **Table 7: Adaptation of GRADE quality elements for qualitative reviews**

Quality element	Description
Risk of bias ('Methodological limitations')	Limitations in study design and implementation may bias interpretation of qualitative themes identified. High risk of bias for the majority of the evidence reduces confidence in review findings. Qualitative studies are not usually randomised and therefore would not be downgraded for study design from the outset (they start as high quality)
Relevance (or applicability) of evidence	This refers to the extent to which the evidence supporting the review findings is applicable to the context specified in the review question
Coherence of findings	This refers to the extent to which review findings are well grounded in data from the contributing primary studies and provide a credible explanation for patterns identified in the evidence
Adequacy of data (theme saturation or sufficiency)	This corresponds to a similar concept in primary qualitative research, that is, whether a theoretical point of theme saturation was achieved, at which point no further citations or observations would provide more insight or suggest a different interpretation of the particular theme. Individual studies that may have contributed to a theme or sub-theme may have been conducted in a manner that by design would have not reached theoretical saturation at an individual study level

2 **Table 8: CERQual levels of concern (by quality element)**

Level of concern	Definition
None or very minor concerns	Unlikely to reduce confidence in the review finding
Minor concerns	May reduce confidence in the review finding
Moderate concerns	Will probably reduce confidence in the review finding
Serious concerns	Very likely to reduce confidence in the review finding

3 **Table 9: Overall confidence in the evidence in CERQual (by review finding)**

Overall confidence level	Definition
High	It is highly likely that the review finding is a reasonable representation of the phenomenon of interest
Moderate	It is likely that the review finding is a reasonable representation of the phenomenon of interest
Low	It is possible that the review finding is a reasonable representation of the phenomenon of interest
Very low	It is unclear whether the review finding is a reasonable representation of the phenomenon of interest

1 *Assessing methodological limitations in qualitative reviews*

2 Methodological limitations in qualitative studies were assessed using the Critical
 3 Appraisal Skills Programme (CASP) checklist for qualitative studies (see appendix H
 4 in [Developing NICE guidelines: the manual](#); NICE 2014). Overall methodological
 5 limitations were derived by assessing the methodological limitations across the 6
 6 domains summarised in Table 10.

7 **Table 10: Methodological limitations in qualitative studies**

Aim and appropriateness of qualitative evidence	This domain assesses whether the aims and relevance of the study were described clearly and whether qualitative research methods were appropriate for investigating the research question
Rigour in study design or validity of theoretical approach	This domain assesses whether the study approach was documented clearly and whether it was based on a theoretical framework (such as ethnography or grounded theory). This does not necessarily mean that the framework has to be stated explicitly, but a detailed description ensuring transparency and reproducibility should be provided
Sample selection	This domain assesses the background, the procedure and reasons for the method of selecting participants. The assessment should include consideration of any relationship between the researcher and the participants, and how this might have influenced the findings
Data collection	This domain assesses the documentation of the method of data collection (in-depth interviews, semi-structured interviews, focus groups or observations). It also assesses who conducted any interviews, how long they lasted and where they took place
Data analysis	This domain assesses whether sufficient detail was documented for the analytical process and whether it was in accordance with the theoretical approach. For example, if a thematic analysis was used, the assessment would focus on the description of the approach used to generate themes. Consideration of data saturation would also form part of this assessment (it could be reported directly or it might be inferred from the citations documented that more themes could be found)
Results	This domain assesses any reasoning accompanying reporting of results (for example, whether a theoretical proposal or framework is provided)

8 *Assessing relevance of evidence in qualitative reviews*

9 Relevance (applicability) of findings in qualitative research is the equivalent of
 10 indirectness for quantitative outcomes, and refers to how closely the aims and

1 context of studies contributing to a theme reflect the objectives outlined in the
2 guideline review protocol.

3 *Assessing coherence of findings in qualitative reviews*

4 For qualitative research, a similar concept to inconsistency is coherence, which
5 refers to the way findings within themes are described and whether they make sense.
6 This concept was used in the quality assessment across studies for individual
7 themes. This does not mean that contradictory evidence was automatically
8 downgraded, but that it was highlighted and presented, and that reasoning was
9 provided. Provided the themes, or components of themes, from individual studies fit
10 into a theoretical framework, they do not necessarily have to reflect the same
11 perspective. It should, however, be possible to explain these by differences in context
12 (for example, the views of healthcare professionals might not be the same as those
13 of family members, but they could contribute to the same overarching themes).

14 *Assessing adequacy of data in qualitative reviews*

15 Adequacy of data (theme saturation or sufficiency) corresponds to a similar concept
16 in primary qualitative research in which consideration is made of whether a
17 theoretical point of theme saturation was achieved, meaning that no further citations
18 or observations would provide more insight or suggest a different interpretation of the
19 theme concerned. As noted above, it is not equivalent to the number of studies
20 contributing to a theme, but rather to the depth of evidence and whether sufficient
21 quotations or observations were provided to underpin the findings.

22 *Assessing importance in qualitative reviews*

23 For themes stemming from qualitative findings, importance was agreed by the
24 committee taking account of the generalisability of the context from which the theme
25 was derived and whether it was sufficiently convincing to support or warrant a
26 change in current practice, as well as the quality of the evidence.

27 **Formal consensus reviews**

28 Formal consensus was carried out using the nominal group technique (Murphy 1998)
29 for evidence review F. This is a structured method focusing on the opinions of
30 individuals within a group. Due to this focus on individuals it is referred to as a
31 'nominal group' technique. It usually involves anonymous voting with an opportunity
32 to provide comments. It is usually conducted by an iterative process in which options
33 with low agreement are eliminated and options with high agreement are retained.
34 Using the comments that individuals provided, options with medium agreement are
35 revised and then considered in a second round.

36 *Details of the nominal group technique as used in this guideline*

37 A search was conducted for relevant published guidelines and systematic reviews.
38 Systematic reviews were included in order to plug any gaps highlighted by the
39 guidelines in terms of coverage of the areas that the committee agreed would be

1 essential content of postnatal care contacts. These were listed a priori in the review
2 protocol.

3 Only international or national guidelines that had been developed on the basis of an
4 evidence review were included. It was agreed that these would be more
5 generalisable than locally developed guidelines and would have a more robust basis
6 for drafting recommendations. In order to identify the most relevant literature, only
7 those guidelines published since the previous NICE recommendations on the content
8 of postnatal care were published in NG37 (2006) were considered. For consistency,
9 the same date cut-off was used in a search for published systematic reviews which
10 were considered in parallel with the published guidelines. All potentially relevant
11 guidelines were assessed for quality using the Appraisal of Guidelines for Research
12 and Evaluation ([AGREE II](#)) instrument (see assessing quality below). The 2 published
13 systematic reviews that were identified were assessed for quality using the Risk of
14 Bias in Systematic Reviews (ROBIS) checklist.

15 Once the guidelines had been assessed, the NGA technical team extracted relevant
16 recommendations from these guidelines and derived a set of statements for all
17 included topics. All statements were checked for practical content by the NGA clinical
18 advisor and the committee chair. If no recommendations existed within the included
19 guidelines for a particular element of postnatal care content, then findings from the
20 published systematic reviews were used to produce statements.

21 The formal consensus exercise was conducted over 2 committee meetings. At the
22 initial meeting the statements were presented to the committee in a questionnaire
23 format. All committee members were invited to take part in the formal consensus
24 exercise (this did not include the chair or guideline clinical advisor as they had been
25 involved in deriving the statements, nor co-opted members). Committee members
26 were asked to rate each statement based on their personal opinion of what they
27 believed 'best practice' would be. The statements were rated using a 9-point Likert
28 scale, where 1 represents 'strongly disagree', 5 represents 'neither agree nor
29 disagree', and 9 represents 'strongly agree'. The participants were also able to
30 record for any statement that they believed they had insufficient knowledge to
31 provide a rating. There was also space for written comments about each statement.
32 Once this first round of voting had been conducted, the NGA technical team
33 calculated overall percentage agreement for each individual statement. Statements
34 with 80% or greater agreement were kept, and were to be used to inform
35 recommendations. Statements with less than 60% agreement were discarded unless
36 there were obvious and addressable issues identified from any comments. Those
37 statements with 60% to 80% agreement were redrafted by the NGA technical team
38 (using the written comments if provided).

39 The redrafted statements were placed into the same questionnaire format as round 1
40 of the formal consensus process. Committee members were sent the revised
41 statements electronically and asked to rate them in the same way as in the first
42 round. Responses were emailed back to the NGA technical team, who calculated
43 agreement as above.

1 At the following committee meeting, statements with 80% or greater agreement (from
2 rounds 1 and 2) were presented as the evidence to inform the development of
3 recommendations. The statements were discussed and the committee used them in
4 combination with their knowledge and experience to develop the recommendations.

5 *Assessing quality of guidelines in formal consensus reviews*

6 Potentially relevant guidelines were assessed for quality using [AGREE II](#) instrument
7 (Table 11). The tool assesses 6 domains: scope and purpose, stakeholder
8 involvement, rigour of development, clarity of presentation, applicability and editorial
9 independence.

10 Within each domain there is a set of questions, each of which is scored using a 7-
11 point scale (1 – ‘strongly disagree’ to 7 – ‘strongly agree’). Each section is rated and
12 then an overall score for that domain is calculated. Two reviewers independently
13 rated all identified guidelines using this method (see the [AGREE II](#) for detailed
14 instructions). The committee took account of the quality ratings during discussions
15 and when they agreed recommendations based on the formal consensus review
16 findings.

17 **Table 11: Assessing quality of guidelines**

Domain	Description
Scope and purpose	Assesses the aim of the guideline, the specific health questions, and the target population
Stakeholder involvement	Assesses the extent to which the guideline involved the appropriate stakeholders, and whether it represents the views of intended users
Rigour of development	Assesses the methods used to gather and synthesise the evidence and to construct the recommendations
Clarity of presentation	Assesses the language, format and structure of the guideline
Applicability	Assesses likely barriers and facilitators of implementation, uptake and resource implications of the guideline
Editorial independence	Assesses the likelihood of the recommendations being biased and potential conflict of interests

18 **Reviewing economic evidence**

19 Systematic reviews of economic literature were conducted for all review questions
20 covered in the guideline, unless economic evidence was not relevant to a review
21 question. In addition, modelling studies that estimated long-term benefits to women
22 and babies and related cost-savings associated with breastfeeding (any or exclusive)
23 were reviewed in order to identify modelling components such as model structures,
24 clinical outcomes associated with breastfeeding, clinical and cost data and further
25 assumptions that could be adopted or used after adaptation when developing the
26 guideline economic model on the same topic.

1 Inclusion and exclusion of economic studies

2 Titles and abstracts of articles identified through the economic literature searches
3 were independently assessed for inclusion using the predefined eligibility criteria
4 listed in Table 12.

5 **Table 12: Inclusion and exclusion criteria for systematic reviews of economic** 6 **evaluations**

Inclusion criteria
Only studies from the Organisation for Economic Co-operation and Development (OECD) member countries were included, as the aim of the review was to identify economic information transferable to the UK context.
Only studies published from 2004 onwards were included in the review. This date restriction was imposed so that retrieved economic evidence was relevant to current healthcare settings and costs.
Selection criteria regarding the populations and interventions assessed were identical to the clinical literature review.
Full economic evaluations that compared 2 or more relevant options and considered both costs and consequences as well as costing analyses that compared only costs between 2 or more interventions.
Clinical effectiveness data utilised in the analysis should have been derived from a literature review, a clinical trial, a prospective or retrospective cohort study, or a study with a before-and-after design.
The outcome measure of the economic analysis should be the quality-adjusted life-year (QALY) or one of the measures considered in the clinical review.
Studies should be reporting separately costs for each option assessed, from a healthcare perspective.
Exclusion criteria
Poster presentations and abstracts in conference proceedings.
Non-English language papers.
Non-comparative studies.
Studies that adopted a non-healthcare perspective and did not consider healthcare costs.

7 Once the screening of titles and abstracts was completed, full-text copies of
8 potentially relevant articles were requested for detailed assessment. Inclusion and
9 exclusion criteria were applied to articles obtained as full-text copies.

10 Evidence tables of included economic studies and lists of economic studies excluded
11 after obtaining full text with reasons for exclusion, are provided in Appendix H and
12 Appendix K, respectively, of the relevant evidence reviews. The PRISMA for the
13 search of economic evaluations is presented in Appendix G of each evidence review.

1 **Consideration of modelling studies that estimated long-term benefits to**
 2 **women and babies and related cost-savings associated with**
 3 **breastfeeding**

4 A systematic review of modelling studies that assessed the long-term benefits and
 5 cost-savings associated with breastfeeding was undertaken, to identify parameters
 6 that could inform the economic model that was developed to inform the cost-
 7 effectiveness of interventions for starting and maintaining breastfeeding, assessed in
 8 evidence review P.

9 The titles and abstracts of papers identified through the searches were assessed for
 10 inclusion in this review using broad eligibility criteria defined in Table 13.

11 **Table 13: Inclusion and exclusion criteria for the systematic review of**
 12 **modelling studies that estimated long-term benefits to women and**
 13 **babies and related cost-savings associated with breastfeeding**

Inclusion criteria
Modelling studies from any country were considered.
The study population should be women breastfeeding term babies that were healthy at birth (single or multiple births).
The studies should estimate projected benefits to women and/or babies and/or related cost-savings associated with breastfeeding.
Exclusion criteria
Poster presentations and abstracts in conference proceedings.
Non-English language papers.

14 Once the screening of titles and abstracts was complete, full-text versions of the
 15 selected papers were acquired.

16 Modelling studies that met inclusion criteria and those that were excluded after full
 17 text was obtained are reported in Appendix N of evidence review P.

18 **Appraising the quality of economic evidence**

19 The applicability and quality of economic evaluations in this guideline were appraised
 20 using the methodology checklist reported in [Developing NICE guidelines: the manual](#)
 21 (NICE 2014), Appendix H, for all studies that met the inclusion criteria.

22 The methodological assessment of economic studies considered in this guideline has
 23 been summarised in economic evidence profiles that were developed for each review
 24 question for which economic evidence was available. All studies that fully or partially
 25 met the applicability and quality criteria described in the methodology checklist were
 26 considered during the guideline development process.

27 Economic profiles of all economic studies that were considered during guideline
 28 development, including de novo economic analyses undertaken for this guideline, are
 29 provided in Appendix I of the respective evidence reviews.

1 Economic modelling

2 The aims of the economic input to the guideline were to inform the guideline
3 committee of potential economic issues related to the postnatal care of women and
4 their babies in order to ensure that recommendations represented a cost-effective
5 use of healthcare resources. Health economic evaluations aim to integrate data on
6 care benefits (ideally in terms of quality-adjusted life-years, QALYs) with the costs of
7 different care options. In addition, the economic input aimed to identify areas of high
8 resource impact, as these need to be supported by robust evidence on cost
9 effectiveness.

10 Areas for economic modelling were prioritised by the committee. The rationale for
11 prioritising review questions for economic modelling was set out in an economic plan
12 agreed between NICE, the committee, and members of the NGA technical team.
13 Economic modelling was undertaken in areas with likely major resource implications,
14 where the current extent of uncertainty over cost effectiveness was significant and
15 economic analysis was expected to reduce this uncertainty. The following economic
16 questions were selected as key issues that were addressed by economic modelling.

- 17 • Cost-effectiveness of interventions for the initiation and maintenance of
18 breastfeeding. The methods and results of the de novo economic analysis are
19 fully reported in Appendix J of evidence review P.
- 20 • Cost-effectiveness of shorter versus longer postpartum stay. This question was
21 originally prioritised for economic modelling, however, clinical evidence
22 suggested that there were no significant differences in outcomes between early
23 and late discharge; therefore, this question was de-prioritised at a later stage
24 as there was no need for economic modelling.
- 25 • Cost-effectiveness of clinical tools for the clinical review of women postnatally.
26 This question was not possible to model, due to lack of relevant clinical
27 evidence.
- 28 • Cost-effectiveness of different approaches for the assessment of early signs
29 and symptoms of postpartum haemorrhage. This question was not possible to
30 model, due to lack of relevant clinical evidence.

31 When new economic analysis was not prioritised or was not possible to conduct, the
32 committee made a qualitative judgement regarding cost effectiveness by considering
33 expected differences in resource use and costs between options, alongside clinical
34 effectiveness evidence identified from the clinical evidence review.

35 Cost effectiveness criteria

36 NICE's report [The NICE Principles](#) sets out the principles that committees should
37 consider when judging whether an intervention offers good value for money. In
38 general, an intervention was considered to be cost effective if any of the following
39 criteria applied (provided that the estimate was considered plausible):

- 1 • the intervention dominated other relevant strategies (that is, it was both less costly
2 in terms of resource use and more effective compared with all the other relevant
3 alternative strategies)
- 4 • the intervention cost less than £20,000 per QALY gained compared with the next
5 best strategy.

6 The committee's considerations of cost effectiveness are discussed explicitly under
7 the heading 'The committee's discussion of the evidence' under subheading 'Cost
8 effectiveness and resource use' in the relevant evidence reviews.

9 **Developing recommendations**

10 **Guideline recommendations**

11 Recommendations were drafted on the basis of the committee's interpretation of the
12 available evidence, taking account of the balance of benefits, harms and costs
13 between different courses of action. When effectiveness and economic evidence was
14 of poor quality, conflicting or absent, the committee drafted recommendations based
15 on their expert opinion. The considerations for making consensus-based
16 recommendations include the balance between potential benefits and harms, the
17 economic costs or implications compared with the economic benefits, current
18 practices, recommendations made in other relevant guidelines, person's preferences
19 and equality issues.

20 The main considerations specific to each recommendation are outlined under the
21 heading 'The committee's discussion of the evidence' within each evidence review.

22 For further details refer to [Developing NICE guidelines: the manual](#) (NICE 2014).

23 **Research recommendations**

24 When areas were identified for which evidence was lacking, the committee
25 considered making recommendations for future research. For further details refer to
26 [Developing NICE guidelines: the manual](#) (NICE 2014).

27 **Validation process**

28 This guideline was subject to a 6-week public consultation and feedback process. All
29 comments received from registered stakeholders were responded to in writing and
30 posted on the NICE website at publication. For further details refer to [Developing
31 NICE guidelines: the manual](#) (NICE 2014).

32 **Updating the guideline**

33 Following publication, NICE will undertake a surveillance review to determine
34 whether the evidence base has progressed sufficiently to consider altering the

- 1 guideline recommendations and warrant an update. For further details refer to
- 2 [Developing NICE guidelines: the manual](#) (NICE 2014).

3 **Funding**

- 4 The NGA was commissioned by NICE to develop this guideline.

1 **References**

2 **Bradburn 2007**

3
4 Bradburn, M. J., Deeks, J. J., Berlin, J. A., & Localio, A. R. Much ado about nothing:
5 A comparison of the performance of meta-analytical methods with rare events.
6 *Statistics in Medicine*, 26, 53–77, 2007.

7 **Dixon-Woods 2005**

8 Dixon-Woods M, Agarwal S, Jones D et al. Synthesising qualitative and quantitative
9 evidence: a review of possible methods. *Journal of Health Services Research &*
10 *Policy* 10(1), 45–53, 2005

12 **Franchini 2012**

13 Franchini, A. J., S. Dias, A. E. Ades, J. P. Jansen, N. J. Welton. Accounting for
14 correlation in network meta-analysis with multi-arm trials. *Research Synthesis Methods*
15 3, 142-160, 2012

16 **Hayden 2013**

17 Jill A. Hayden, Danielle A. van der Windt, Jennifer L. Cartwright, Pierre Côté, Claire
18 Bombardier. Assessing Bias in Studies of Prognostic Factors. *Ann Intern Med.* 158,
19 280–286, 2013

20 **Higgins 2011**

21 Higgins JPT, Thomas, J (editors). *Cochrane Handbook for Systematic Reviews of*
22 *Interventions Version 6* The Cochrane Collaboration, 2019. Available from
23 <https://training.cochrane.org/handbook/current> (accessed 18 March 2020)

24 **Lewin 2015**

25 Lewin S, Glenton C, Munthe-Kaas H et al. Using qualitative evidence in decision
26 making for health and social interventions: an approach to assess confidence in
27 findings from qualitative evidence syntheses (GRADE-CERQual). *PLoS Med* 12, 10,
28 e1001895, 2015

29 **McGowan 2016**

30 McGowan J, Sampson M, Salzwedel DM et al. [PRESS Peer Review of Electronic](#)
31 [Search Strategies: 2015 guideline statement](#). *Journal of Clinical Epidemiology* 75:
32 40–6, 2016

33 **Murphy 2016**

34 Murphy MK, Black NA, Lamping DL, McKee CM, Standerson CFB, Askam, J.
35 Consensus development methods, and their use in clinical guideline development.
36 *Heath Technology Assessment*, 2, 1998

1

2 **NICE 2014**

3 National Institute for Health and Care Excellence (NICE). Developing NICE
4 guidelines: the manual, 2014 (updated 2018). Available from
5 <https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview> (accessed
6 18 March 2020)

7 **NICE 2018**

8 National Institute for Health and Care Excellence (NICE). NICE Policy on conflicts of
9 interest, 2014 (updated 2017). Available from
10 [https://www.nice.org.uk/Media/Default/About/Who-we-are/Policies-and-](https://www.nice.org.uk/Media/Default/About/Who-we-are/Policies-and-procedures/declaration-of-interests-policy.pdf)
11 [procedures/declaration-of-interests-policy.pdf](https://www.nice.org.uk/Media/Default/About/Who-we-are/Policies-and-procedures/declaration-of-interests-policy.pdf) (accessed 18 March 2020)

12 **Santesso 2016**

13 Santesso N, Carrasco-Labra A, Langendam M et al. Improving GRADE evidence
14 tables part 3: detailed guidance for explanatory footnotes supports creating and
15 understanding GRADE certainty in the evidence judgments. Journal of clinical
16 epidemiology 74, 28-39, 2016