National Institute for Health and Care Excellence

Final

Intrapartum care for women with existing medical conditions or obstetric complications and their babies

Supplement 1: Methods

NICE guideline NG121

Development of the guideline and methods

March 2019

Final

Developed by the National Guideline Alliance hosted by 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 All NICE guidance is subject to regular review and may be updated or withdrawn.

ISBN: 978-1-4731-3296-2

Contents

Development of the guideline

Remit

The National Institute for Health and Care Excellence (NICE) commissioned the National Guideline Alliance (NGA) to develop a guideline for intrapartum care of 'high risk' women.

This guideline will sit alongside NICE's existing guideline on intrapartum care for healthy women and babies (CG190). It covers intrapartum care when either the woman or her baby is at high risk of adverse outcomes because of an existing medical condition affecting the woman or an obstetric complication.

What this guideline covers

Groups that are covered

Women in labour (spontaneous or induced) who are at high risk of adverse outcomes for themselves and/or their baby.

Two groups of women in labour are the main focus of this guideline:

- women in spontaneous or induced labour (or who have a planned caesarean section) who are identified as being at high risk of adverse outcomes because of existing maternal medical conditions
- women in spontaneous or induced labour who are identified as being at high risk of adverse outcomes because:
  - of obstetric complications, in the current and/or previous pregnancy, labour and/or birth
  - the baby is identified during labour to be at risk of adverse outcomes
  - they have had no antenatal care.

Clinical areas that are covered

The guideline covers the following clinical areas.

Women at high risk of adverse outcomes for themselves and/or their baby because of existing maternal medical conditions

- Information provision
- Antenatal care planning involving a multidisciplinary team
- Intrapartum care for women with cardiac (heart) disease:
  - stratification of risk
  - management of anticoagulation for valvular disease
  - mode of birth

  - fluid management
  - diagnosis and management of cardiomyopathy
  - anaesthesia and analgesia
  - management of the third stage of labour
- Intrapartum care for women with asthma:
  - analgesia
  - use of prostaglandins and other uterotonics
- Intrapartum care for women on long-term systemic steroid medication:
  - steroid replacement regimens
- Intrapartum care for women with haemostatic (bleeding) disorders:
  - use of regional anaesthesia and analgesia
  - management of the third stage of labour
  - thresholds for platelet count and/or function requiring plans for the birth to be modified
- Intrapartum care for women with a history of subarachnoid haemorrhage or arterio-venous malformation of the brain:
  - mode of birth
  - management of the second stage of labour
- Intrapartum care for women who develop an acute kidney injury or have chronic kidney disease:
  - fluid management
  - mode of birth
- Intrapartum care for women with obesity:
  - fetal presentation
  - anaesthesia and analgesia
  - fetal monitoring
  - delivery position
  - equipment needs

Women at high risk of adverse outcomes for themselves and/or their baby because of obstetric complications or other reasons

- Information provision
- Risk assessment
- Intrapartum care for women with pyrexia:
  - fetal blood sampling
  - use of anti-pyretics
- Intrapartum care for women with sepsis:
  - mode of birth
  - anaesthesia and analgesia

  - fetal monitoring
  - antimicrobial therapy
  - management for the woman immediately after the birth
- Intrapartum care for women with intrapartum haemorrhage:
  - management of intrapartum haemorrhage
- Intrapartum care for women with breech presenting in labour:
  - mode of birth
- Intrapartum care for women with a small-for-gestational age baby:
  - fetal monitoring
- Intrapartum care for women with a large-for-gestational age baby:
  - mode of birth
- Intrapartum care for women who present in labour having had no antenatal care:
  - risk assessment and management of labour
- Intrapartum care for women with previous caesarean section:
  - management of the first and second stages of labour
- Intrapartum care for women in labour after 42 weeks of pregnancy:
  - maternal and fetal monitoring

For further details see the guideline scope on the NICE website.

What this guideline does not cover

Groups that are not covered

- Women in labour whose baby is identified antenatally to be at high risk of adverse outcomes exclusively because the baby has a congenital disorder.
- Women in labour who are identified before or during labour to be at high risk of adverse outcomes solely because of personal or social circumstances.
- Women in labour without known medical conditions who are having a caesarean section that was planned during their antenatal care.

Clinical areas that are not covered

- Women with mental health problems requiring medication
- Women with thrombotic disorders
- Women with musculoskeletal disorders, including back problems
- Women with hepatitis B or C, or with HIV
- Women with previous myomectomy or hysterotomy
- Women with pelvic girdle pain
- Women with neurological disorders such as epilepsy
- Women with neuromuscular disorders such as multiple sclerosis

- Women with sickle cell disease
- Women with thyroid disease
- Women with liver disease
- Women with multiple pregnancy
- Women with hypertension in pregnancy
- Women with a third- or fourth-degree tear
- Women with diabetes in pregnancy
- Women with obstetric cholestasis
- Women in suspected preterm labour without medical or obstetric complications
- Women with cord prolapse
- Women who collapse in labour
- Women with suspected amniotic fluid embolism
- Women colonised by group B streptococcus in pregnancy
- Women with planned caesarean section for reasons other than existing maternal medical conditions
- Women with placenta accreta
- Women who have undergone female genital mutilation
- Women whose baby is stillborn
- Women whose baby is identified as having shoulder dystocia
- Women with malpresentation other than breech

The rationale for excluding these areas is outlined in the guideline scope on the NICE website.

Methods

Preamble

This section summarises methods used to identify and review the evidence, to consider cost effectiveness, and to develop guideline recommendations. This guideline was developed in accordance with methods described in Developing NICE guidelines: the manual (NICE 2014).

Until March 2018, declarations of interest were recorded and managed in accordance with NICE's 2014 conflicts of interest policy. From April 2018, declarations were recorded and managed in accordance with NICE's 2018 Policy on declaring and managing interests for NICE advisory committees.

For this guideline there were 2 committees, each developing recommendations for part of the guideline (1 for women at high risk of adverse outcomes for themselves and/or their baby because of existing maternal medical conditions, and the other for women at high risk of adverse outcomes for themselves and/or their baby because of obstetric complications or other reasons; see separate documents listing members of each committee and their declarations of interest). The chairs of the 2 committees attended the final pre-consultation meeting of each other's committee to promote transparent and consistent decision making across the guideline as a whole. Both committees had sight of all the draft recommendations before consultation and were able to provide feedback on the other committee's recommendations via the chair of their own committee. Additionally, the chairs of the 2 committees led a joint discussion, involving both committees, during the post-consultation guideline committee meetings in relation to stakeholder comments on the topic of women-centred language. The committees jointly agreed principles for revisions to be made to recommendations in the light of the comments.

Developing the review questions and outcomes

The 43 review questions considered in this guideline were based on the key areas identified in the guideline scope .They were drafted by the NGA technical team, and refined and validated by the relevant guideline committee (see Table 1: Summary of review questions and index to evidence reports).

The review questions were based on the following frameworks:

- intervention reviews – using population, intervention, comparison and outcome (PICO)
- diagnostic reviews and reviews of clinical prediction model accuracy – using population, diagnostic test (index test), reference standard and target condition (PIRT)
- prognostic reviews – using population, presence or absence of a prognostic, risk or predictive factor and outcome (PPO)

- qualitative reviews – using population, phenomenon of interest and context (PICo).

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

Full literature searches, critical appraisal and evidence reviews were completed for all review questions.

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

Table 1: Summary of review questions and index to evidence reports

Evidence report	Subtopic in scope	Review question	Type of review
Women at high risk of adverse outcomes for themselves and/or their baby because of existing maternal medical conditions			
[A] Information provision	–	What are the main areas of information about labour and birth that are needed by pregnant women with existing medical conditions?	Qualitative and intervention
[B] Antenatal care planning involving a multidisciplinary team	–	Does antenatal care planning for birth involving an expanded multidisciplinary team compared with routine antenatal care planning improve intrapartum outcomes for women with existing medical conditions?	Intervention
[C] Intrapartum care for women with cardiac disease	Stratification of risk	What history, clinical examination and investigation is most useful to stratify the intrapartum risk for women with cardiac disease?	Intervention, diagnostic and prognostic
[C] Intrapartum care for women with cardiac disease	Management of anticoagulation for valvular disease	What is the appropriate management of anticoagulation for women with valvular disease in planning for childbirth?	Intervention
[C] Intrapartum care for women with cardiac disease	Mode of birth	Which women with cardiac disease should be offered elective caesarean section or assisted second stage for reasons specific to cardiac disease?	Intervention

Evidence report	Subtopic in scope	Review question	Type of review
[C] Intrapartum care for women with cardiac disease	Fluid management	Which women with cardiac conditions need additional haemodynamic monitoring or management during childbirth: • input-output chart of fluid balance with a urinary catheter or urometer • invasive monitoring using an arterial line and central venous pressure • cardiac monitoring • fluid restriction?	Intervention
[C] Intrapartum care for women with cardiac disease	Diagnosis and management of cardiomyopathy	What is the most appropriate method of diagnosis for women with suspected cardiomyopathy in labour?	Diagnostic and prognostic
[C] Intrapartum care for women with cardiac disease	Diagnosis and management of cardiomyopathy	What is the optimal management for women with peripartum cardiomyopathy in labour?	Intervention
[C] Intrapartum care for women with cardiac disease	Anaesthesia and analgesia	Is regional or general anaesthesia safer for women with cardiac disease for peripartum surgical procedures including caesarean section?	Intervention
[C] Intrapartum care for women with cardiac disease	Anaesthesia and analgesia	What are the risks and benefits of central neuraxial analgesia compared with systemic analgesia, inhaled analgesia or no analgesia for women with cardiac disease who are in labour?	Intervention
[C] Intrapartum care for women with cardiac disease	Management of the third stage of labour	How should the third stage of labour be managed for women with cardiac disease?	Intervention
[D] Intrapartum care for women with asthma	Analgesia	What are the risks and benefits of central neuraxial analgesia compared with systemic analgesia, inhaled analgesia or no analgesia for women with asthma in labour?	Intervention
[D] Intrapartum care for women with asthma	Use of prostaglandins and other uterotonics	What is the safety of drugs commonly used in labour in women with difficult asthma, including prostaglandins for	Intervention

Evidence report	Subtopic in scope	Review question	Type of review
		inducing labour and prostaglandins and other uterotonics for treating postpartum haemorrhage?	
[E] Intrapartum care for women on long-term systemic steroid medication	Steroid replacement regimens	What steroid replacement regimen should be used during the peripartum period for women on long-term systemic steroid medication?	Intervention
[F] Intrapartum care for women with haemostatic disorders	Use of regional anaesthesia and analgesia	In which women with haemostatic disorders should regional anaesthesia and analgesia be avoided?	Prognostic
[F] Intrapartum care for women with haemostatic disorders	Thresholds for platelet count and/or function requiring plans for the birth to be modified	What is the threshold level of platelet count and/or function below which plans for the birth need to be modified in women with haemostatic disorders?	Prognostic
[F] Intrapartum care for women with haemostatic disorders	Management of the third stage of labour	How should the third stage of labour be managed for women who are at increased risk of bleeding because of haemostatic disorders?	Intervention
[G] Intrapartum care for women with a history of subarachnoid haemorrhage or arterio-venous malformation of the brain	Mode of birth	Which women with a history of intracranial haemorrhage or a cerebrovascular malformation should avoid labour?	Intervention
[G] Intrapartum care for women with a history of subarachnoid haemorrhage or arterio-venous malformation of the brain	Management of the second stage of labour	How should the second stage of labour be managed for women with a history of intracranial haemorrhage or with a cerebrovascular malformation?	Intervention
[H] Intrapartum care for women who develop an	Fluid management	What is the most effective fluid management regimen for women who develop an acute	Intervention

Evidence report	Subtopic in scope	Review question	Type of review
acute kidney injury or have chronic kidney disease		kidney injury or have chronic kidney disease and who are in the peripartum period?	
[H] Intrapartum care for women who develop an acute kidney injury or have chronic kidney disease	Mode of birth	Which women who develop an acute kidney injury or have chronic kidney disease should be offered early birth (via induction of labour or elective caesarean section) for reasons specific to kidney disease?	Intervention
[I] Intrapartum care for women with obesity	Fetal presentation	What is the value of assessing fetal presentation early in labour for women with obesity?	Intervention
[I] Intrapartum care for women with obesity	Anaesthesia and analgesia	Does an ultrasound scan of the woman's back improve needle siting for central neuraxial blockade anaesthesia and analgesia for women with obesity in the peripartum period?	Intervention
[I] Intrapartum care for women with obesity	Fetal monitoring	What is the effectiveness of intermittent auscultation compared with electronic fetal monitoring during labour in women with obesity?	Intervention
[I] Intrapartum care for women with obesity	Delivery position	What is the optimal position in the second stage of labour for women with obesity?	Intervention
[I] Intrapartum care for women with obesity	Equipment needs	What additional equipment is needed to ensure optimal care of women with obesity in the peripartum period?	Intervention
Women at high risk of adverse outcomes for themselves and/or their baby because of obstetric complications or other reasons			
[J] Information provision	–	What are the information needs of women at high risk of adverse outcomes in labour due to obstetric complications that arise before or during the intrapartum period?	Qualitative and intervention
[K] Risk assessment	–	What maternal observations should be performed for women at high risk of adverse outcomes in labour for the woman or the baby, and what	Intervention

Evidence report	Subtopic in scope	Review question	Type of review
		is the optimal frequency of making these observations?	
[L] Intrapartum care for women with pyrexia	Fetal blood sampling	Does the use of fetal blood sampling (in conjunction with electronic fetal monitoring) for women with pyrexia in labour improve outcomes for the baby?	Intervention and diagnostic
[L] Intrapartum care for women with pyrexia	Use of anti-pyretics	Does the use of anti-pyretics in women with pyrexia in labour improve outcomes for the woman or the baby?	Intervention
[M] Intrapartum care for women with sepsis	Mode of birth	What is the optimal mode of birth for women with sepsis?	Intervention
[M] Intrapartum care for women with sepsis	Anaesthesia and analgesia	What are the most effective and safe methods of anaesthesia for women with sepsis in labour?	Intervention
[M] Intrapartum care for women with sepsis	Anaesthesia and analgesia	What are the most effective and safe methods of analgesia for women with sepsis in labour?	Intervention
[M] Intrapartum care for women with sepsis	Fetal monitoring	How should fetal monitoring be managed for women with sepsis who present in labour?	Intervention and diagnostic
[M] Intrapartum care for women with sepsis	Antimicrobial therapy	What is the most clinical and cost effective antimicrobial therapy for women with sepsis in labour?	Intervention
[M] Intrapartum care for women with sepsis	Management for the woman immediately after the birth	What is the most appropriate management for women with sepsis in the first 24 hours after the birth?	Intervention
[N] Intrapartum care for women with intrapartum haemorrhage	Management of intrapartum haemorrhage	What is the optimal management for intrapartum haemorrhage?	Intervention
[O] Intrapartum care for women with breech presenting in labour	Mode of birth	What is the optimal mode of birth (emergency caesarean section or continuation of labour) for women with breech presenting in the first or second stage of labour?	Intervention

Evidence report	Subtopic in scope	Review question	Type of review
[P] Intrapartum care for women with a small-for-gestational age baby	Fetal monitoring	How should fetal monitoring be managed during labour for women with a small-for-gestational age baby?	Intervention
[Q] Intrapartum care for women with a large-for-gestational age baby	Mode of birth	What is the optimal mode of birth (emergency caesarean section or continuation of labour) for women with a large for gestational age baby?	Intervention
[R] Intrapartum care for women who present in labour having had no antenatal care	Risk assessment and management of labour	What are the most appropriate systems for risk assessment and management of labour for women who present in labour having had no antenatal care?	Intervention
[S] Intrapartum care for women with previous caesarean section	Management of the first and second stages of labour	How should the first and second stages of labour be managed for women with previous caesarean section?	Intervention
[T] Intrapartum care for women in labour after 42 weeks of pregnancy	Maternal and fetal monitoring	What maternal and fetal monitoring should be carried out for women in labour after 42 weeks of pregnancy?	Intervention

Additional information related to development of the guideline is contained in:

- Supplement 1 (Development of the guideline and methods; this document)
- Supplement 2 (Health economics)
- Supplement 3 (NGA staff list).

Searching for evidence

Clinical literature search

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

Databases were searched using medical subject headings, free-text terms and study type filters where appropriate. Where possible, searches were restricted to retrieve articles published in English. All searches were conducted in the following databases: Medline, Medline-in-Process, Cochrane Central Register of Controlled Trials (CCTR), Cochrane Database of Systematic Reviews (CDSR), Database of Abstracts of Reviews of Effects (DARE), Health Technology Assessments (HTA) and Embase.

For review questions related to information provision, PsycInfo and Maternity and Infant Care Database (MIDIRS) were also searched. Except for the review question about management of intrapartum haemorrhage (for which the search had been completed during March 2018) all searches were updated at least 6–8 weeks in advance of the final guideline committee meetings before consultation on the draft guideline; these updates were completed during April 2018. Any studies added to the databases after April 2018 (including those published before April 2018 but not yet indexed) were not considered for inclusion.

Search strategies were quality assured by cross-checking reference lists of relevant articles, analysing search strategies from other systematic reviews and asking members of the relevant committee to highlight key studies. All search strategies were also quality assured by an information scientist who was not involved in developing the primary search strategy. Details of the search strategies, including study-design filters applied and databases searched, are presented in Appendix B of each evidence report.

All publications highlighted by stakeholders at the time of the consultation on the draft scope were considered for inclusion. During the scoping phase, searches were conducted for guidelines, health technology assessments, systematic reviews, economic evaluations and reports on biomedical databases and websites of organisations relevant to the topic. Formal searching for grey literature and unpublished literature was not undertaken routinely.

Health economic literature search

Systematic literature searches were also undertaken to identify published health economic evidence. A broad search was conducted to identify health economic evidence related to intrapartum care in the following databases: NHS Economic Evaluation Database (NHS EED) and HTA. A broad search was also conducted to identify health economic evidence related to intrapartum care in the following databases with an economic search filter applied: Medline, CCTR and Embase. A specific health economic search was undertaken for the review question about antimicrobial therapy for women with sepsis as the full title of the question contained the phrase ‘cost effectiveness’. For this question the NHS EED and HTA databases were searched as well as the Medline, CCTR and Embase databases, where an economic search filter was applied. Where possible, the searches were restricted to retrieve articles published in English; studies published in languages other than English were not eligible for inclusion.

The search strategies for the health economic literature search are included in Supplement 2 (Health economics). As for the clinical literature searches, economic literature searches were updated at least 6–8 weeks in advance of the final committee meetings before consultation on the draft guideline; these updates were completed during April 2018.

Call for evidence

No call for evidence was made.

Reviewing evidence

Systematic review process

The evidence was reviewed in accordance with the following approach.

- Potentially relevant articles were identified from the search results for each review question by screening titles and abstracts. Full-text copies of the articles were then obtained.
- Full-text articles were reviewed against pre-specified inclusion and exclusion criteria in the review protocol (see Appendix A of each evidence report).
- Key information was extracted from each article on study methods and results, in accordance with factors specified in the review protocol. The information was presented in a summary table in the corresponding evidence report and in a more detailed evidence table (see Appendix E of each evidence report).
- Included studies were critically appraised using an appropriate checklist as specified in Developing NICE guidelines: the manual (NICE 2014). Further detail on appraisal of the evidence is provided below.
- Summaries of evidence by outcome were presented in the corresponding evidence report and discussed by the relevant committee.

Review questions selected as high priorities for health economic analysis (and those selected as medium priorities and where health economic analysis could influence recommendations) were subject to dual screening and study selection through a 10% random sample of articles. Any discrepancies were resolved by discussion between the first and second reviewers or by reference to a third (senior) reviewer. For the remaining review questions, internal (NGA) quality assurance processes included consideration of the outcomes of screening, study selection and data extraction and the relevant committee reviewed the results of study selection and data extraction. The review protocol for each question specifies whether dual screening and study selection was undertaken for that particular question.

Drafts of all evidence reviews were checked by a senior reviewer.

Type of studies and inclusion/exclusion criteria

Inclusion and exclusion of studies was based on criteria specified in the corresponding review protocol.

Systematic reviews with meta-analyses were considered to be the highest quality evidence that could be selected for inclusion.

For intervention reviews, randomised controlled trials (RCTs) were prioritised for inclusion because they are considered to be the most robust type of study design that could produce an unbiased estimate of intervention effects. Where there was limited evidence from RCTs, non-randomised controlled trials and/or observational studies were considered for inclusion, including cohort studies, case–control studies, cross-sectional studies and case series. Where data from observational studies were included, results for each outcome were presented separately for each study and meta-analysis was not conducted.

For diagnostic or clinical prediction rule reviews, test-and-treat RCTs were prioritised for inclusion. In the absence of such studies, cross-sectional studies and prospective or retrospective cohort studies were considered for inclusion. When limited evidence was available, case–control studies and case series were also considered for inclusion.

For prognostic reviews, prospective and retrospective cohort and case–control studies and case series were considered for inclusion.

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

The relevant committee was consulted about any uncertainty regarding inclusion or exclusion of studies. A list of excluded studies for each review question, including reasons for exclusion is presented in Appendix D of the corresponding evidence report.

Narrative reviews, posters, letters, editorials, comment articles, unpublished studies and studies published in languages other than English were excluded. Conference abstracts were generally not considered for inclusion except in a few review questions where the relevant committee anticipated that no other published evidence was likely to be identified (see the review protocols for details).

Methods of combining evidence

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

Data synthesis for intervention reviews

Meta-analysis to pool results from RCTs was conducted where possible using Cochrane Review Manager (RevMan5) software. As noted above, results from observational studies were not pooled using meta-analysis.

For dichotomous outcomes, such as mortality, the Mantel–Haenszel method with a fixed effect model was used to calculate risk ratios (relative risks; RRs).

For continuous outcomes, measures of central tendency (mean) and variation (standard deviation; SD) are required for meta-analysis. Data for continuous outcomes, such as duration of hospital stay, were meta-analysed using an inverse-variance method for pooling weighted mean differences (WMDs). Where SDs were not reported for each intervention group, the standard error (SE) of the mean difference was calculated from other reported statistics (p values or 95% confidence intervals; CIs) and then meta-analysis was conducted as described above.

When evidence was based on studies that reported descriptive data or medians with interquartile ranges or p values, this information was included in the corresponding GRADE tables (see below) without calculating relative or absolute effects. Consequently, certain aspects of quality assessment such as imprecision of the effect estimate could not be assessed for this type of evidence. The limited reporting was interpreted as representing a risk of bias when assessing study limitations.

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

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

Statistical heterogeneity was assessed by visually examining forest plots and calculating the I-squared inconsistency statistic (with an I-squared value of more than 50% indicating considerable heterogeneity, and more than 80% indicating very serious heterogeneity). When considerable heterogeneity was present, the meta-analysis was re-run using the Der-Simonian and Laird method with a random effects model and the results from the two models were compared. In addition, predefined subgroup analyses were performed where possible. In the case of unexplained heterogeneity, sensitivity analyses were planned based on the quality of studies, eliminating studies at high risk of bias (in relation to randomisation, allocation concealment and blinding, and/or missing outcome data). In cases where there was no plausible explanation for the heterogeneity, the quality of the evidence was downgraded in GRADE for inconsistency and highlighted in the discussion of results.

When case series were included, descriptive data from the studies were included and no further analysis was performed.

Data synthesis for reviews of diagnostic test accuracy and clinical prediction tools

When diagnostic test accuracy was measured dichotomously, sensitivity, specificity, and positive and negative likelihood ratios were used as outcomes. When diagnostic test accuracy was measured continuously, the area under the receiver-operating characteristic (ROC) curve (AUC) was used. These diagnostic test accuracy parameters were obtained directly from results reported in the source articles or calculated by the NGA technical team using data reported in the articles. Where possible, 95% CIs for diagnostic test accuracy parameters were reported;

alternatively, median values and corresponding ranges were used if CIs were not reported and could not be calculated by the NGA technical team.

Sensitivity and specificity measure the ability of a test to correctly classify participants as having or not having the target condition. When sensitivity is high, a negative test result rules out the condition. When specificity is high, a positive test result rules in the condition. An ideal test would be both highly sensitive and highly specific, but this is frequently unachievable and typically there is a trade-off between the 2 measures.

The following cut-offs were used when summarising sensitivity and specificity:

- high: more than 90%
- moderate: 75% to 90%
- low: less than 75%.

Positive and negative likelihood ratios measure the association between a test result and the target condition. A positive likelihood ratio (LR+) greater than 1 indicates a positive test result and is associated with having the condition, while a negative likelihood ratio (LR-) less than 1 indicates a negative test result and is associated with not having the condition. A high value of LR+ would indicate that the test is useful in ruling in the condition whereas a low value of LR- would indicate that the test is useful in ruling out the condition.

The following cut-offs were used when summarising likelihood ratios:

- very useful test: LR+ more than 10, LR- less than 0.1
- moderately useful test: LR+ 5 to 10, LR- 0.1 to 0.2
- not a useful test: LR+ less than 5, LR- more than 0.2.

The AUC shows the true-positive rate (sensitivity) as a function of false-positive rate (1 – specificity). The following cut-offs were used when summarising AUC:

- excellent or perfect test: 0.91–1.00
- good: 0.81–0.92
- moderate: 0.71–0.80
- poor: 0.61–0.70
- very poor: 0.50–0.60
- the index test is worse than chance: lower than 0.50.

Meta-analysis of diagnostic test accuracy parameters was planned if there was data from three or more studies that could be pooled. However, this was not the case in any of the reviews, therefore, meta-analysis for diagnostic test accuracy parameters was not performed.

Data synthesis for prognostic reviews

Determining risk factors for complications during labour and birth could aid early identification and subsequent management. Odds ratios (ORs) or RRs with 95% CIs

reported in published studies were extracted or calculated by the NGA technical team to examine relationships between risk factors and outcomes of interest. Ideally analyses would have adjusted for key confounders (such as age or parity) to be considered for inclusion. Recognising variation across studies in terms of populations, risk factors, outcomes and statistical analysis methods (including adjustments for confounding factors), prognostic data were not pooled, but results from individual studies were presented in the evidence reports.

When case series were included, descriptive data from the studies were included and no further analysis was performed.

Data synthesis for qualitative reviews

Where possible, a meta-synthesis was conducted to combine evidence from qualitative studies. The main aim of qualitative data synthesis in this guideline was to describe topics that might influence a woman's experience of labour and birth, including experience of her birth companion(s), rather than building new theories or reconceptualising topics under review. Whenever studies identified a qualitative theme, this was extracted and the main characteristics were summarised. When all themes were extracted from studies, common concepts were categorised and tabulated. This included information on how many studies had contributed to each theme identified by the NGA technical team.

In qualitative synthesis, a theme being reported more than other themes across included studies does not necessarily mean that the theme is more important than other themes. The aim of qualitative research is to identify new perspectives on a particular topic. Study types and populations in qualitative research can differ widely, meaning that themes identified by just 1 or a few studies can provide important new information on a given topic. Therefore, for the purpose of the qualitative reviews in this guideline, it was planned that further studies would not be added when they reported the same themes as had already been identified from other studies because the emphasis was to be on conceptual robustness rather than quantitative completeness of the evidence. This would have implications for the types and numbers of studies included in the qualitative reviews, with study inclusion continuing until no new relevant data could be found regarding a topic that would add to or refute it. This concept is referred to in the literature as 'theoretical saturation' (Dixon-Woods 2005). However, there was limited evidence available for the qualitative reviews considered in this guideline, and so the methods for managing data saturation were not needed.

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

Emerging themes were placed into a thematic map representing the relationship between themes and overarching categories. The purpose of such a map is to show relationships between overarching categories and associated themes.

Appraising the quality of evidence

Intervention studies

GRADE methodology for intervention reviews

For intervention reviews, the evidence for outcomes from included RCTs and comparative observational studies was evaluated and presented using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology developed by the international GRADE working group. GRADE was not used for evidence from case series; instead quality of case series evidence was assessed using the Checklist for Case Series developed by the Joanna Briggs Institute. More information about this tool can be found on the developer's website.

When GRADE was applied, software developed by the GRADE working group (GRADEpro) was used to assess the quality of each outcome, taking account of individual study quality factors and any meta-analysis results. Results were presented in GRADE profiles (GRADE tables).

The selection of outcomes for each review question was agreed during development of the associated review protocol in discussion with the relevant committee. The evidence for each outcome was examined separately for the quality elements summarised in Table 2. Criteria considered in the rating of these elements are discussed below. Each element was graded using the quality ratings summarised in Table 3. Footnotes to GRADE tables were used to record reasons for grading a particular quality element as having a 'serious' or 'very serious' quality issue. The ratings for each component were combined to obtain an overall assessment of quality for each outcome as described in Table 4.

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

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

Quality element	Description
Risk of bias ('Study limitations')	Limitations in study design and implementation may bias estimates of treatment effect. High risk of bias for the majority of the evidence reduces confidence in the estimated effect
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 relatively few participants or few events of interest, resulting in wide confidence intervals around estimates of effect that include clinically 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

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

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

Assessing risk of bias in intervention reviews

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

Risk of bias in RCTs was assessed using the Cochrane risk of bias tool (see Appendix H in Developing NICE guidelines: the manual; NICE 2014).

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

- selection bias
- performance bias
- attrition bias
- detection bias
- reporting bias.

A study with a poor methodological design does not automatically imply high risk of bias; the bias is considered individually for each outcome and it is assessed whether the chosen design and methodology will impact on the estimation of the intervention effect.

More details about the Cochrane risk of bias tool can be found in Section 8 of the Cochrane Handbook for Systematic Reviews of Interventions (Higgins 2011).

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

For observational studies the Newcastle-Ottawa checklist was used (see Appendix H in Developing NICE guidelines: the manual; NICE 2014).

Assessing inconsistency in intervention reviews

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

Inconsistency was assessed by visually inspecting forest plots and observing whether there was considerable heterogeneity in the results of the meta-analysis. This was assessed by calculating the I-squared statistic for the meta-analysis with an I-squared value of more than 50% indicating considerable heterogeneity, and more than 80% indicating very serious heterogeneity. When considerable or very serious heterogeneity was observed, possible reasons were explored and subgroup analyses were performed as pre-specified in the review protocol where possible. In the case of unexplained heterogeneity, sensitivity analyses were planned based on the quality of studies, eliminating studies at high risk of bias (in relation to randomisation, allocation concealment and blinding, and/or missing outcome data).

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

Assessing indirectness in intervention reviews

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

Assessing imprecision and clinical importance in intervention reviews

Imprecision in GRADE methodology refers to uncertainty around the effect estimate and whether or not there is a clinically important difference between interventions (that is, whether the evidence clearly supports a particular recommendation or appears to be consistent with several candidate recommendations). Therefore, imprecision differs from other aspects of evidence quality because it is not concerned with whether the point estimate is accurate or correct (has internal or external validity). Instead, it is concerned with uncertainty about what the point estimate actually represents. This uncertainty is reflected in the width of the CI.

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

Imprecision was assessed in the guideline evidence reviews by considering whether the width of the 95% CI of the effect estimate was relevant to decision making, considering each outcome independently. This is illustrated in Figure 1, which considers a positive outcome for the comparison of treatment ‘A’ versus treatment ‘B’. Three decision-making zones can be differentiated, bounded by the thresholds for clinical importance (minimally important differences; MIDs) for benefit and harm. The MID for harm for a positive outcome means the threshold at which treatment A is less effective than treatment B by an amount that is clinically important to people with the condition of interest (favours B).

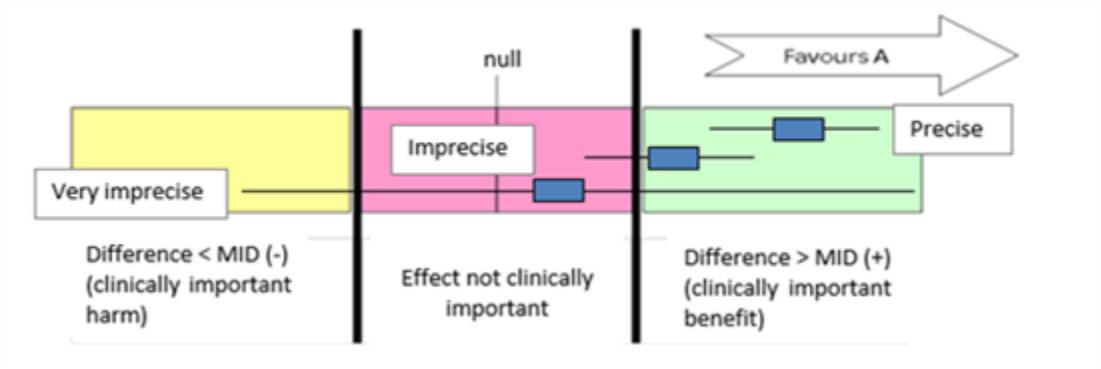
When the CI of the effect estimate is wholly contained in 1 of the 3 zones there is no uncertainty about the size and direction of effect, therefore, the effect estimate is considered precise; that is, there is no imprecision.

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

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

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

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



MID, minimally important difference

Defining minimally important differences for intervention reviews

Each committee was asked whether there were any recognised or acceptable MIDs in the clinical literature and community relevant to the review questions under consideration. Neither committee was aware of any MIDs that could be used for their parts of the guideline.

In the absence of published or accepted MIDs, the committee agreed to use the GRADE default MIDs to assess imprecision. For dichotomous outcomes clinically important thresholds for a RR of 0.8 and 1.25 respectively were used as default MIDs in the guideline. The same thresholds were used as default MIDs in the guideline for all dichotomous outcomes considered in intervention evidence reviews. For continuous outcomes default MIDs are equal to half the median SD of the control groups at baseline (or at follow-up if the SD is not available a baseline).

Diagnostic reviews and clinical prediction models

Adapted GRADE methodology for diagnostic reviews and prediction models

For diagnostic reviews and prediction models, an adapted GRADE approach was used. GRADE methodology is designed for intervention reviews but the quality assessment elements and outcome presentation were adapted for diagnostic test accuracy reviews and prediction models. For example, GRADE tables were modified to include diagnostic test accuracy measures (sensitivity, specificity and likelihood ratios).

The evidence for each outcome in the diagnostic reviews and prediction models was examined separately for the quality elements listed and defined in Table 5. The

criteria considered in the rating of these elements are discussed below. Each element was graded using the quality levels summarised in Table 3. Footnotes to GRADE tables were used to record reasons for grading a particular quality element as having a 'serious' or 'very serious' quality issue. The ratings for each component were combined to obtain an overall assessment of quality for each outcome as described in Table 4.

The initial quality rating was based on the study design: cross-sectional or cohort studies start as 'high' quality and case–control studies start as 'low' quality.

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 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

Assessing risk of bias in diagnostic reviews and prediction models

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

Risk of bias in primary diagnostic accuracy reviews or prediction models in QUADAS-2 consists of 4 domains:

- participant selection
- index test
- reference standard
- flow and timing.

More details about the QUADAS-2 tool can be found on the developer's website.

Assessing inconsistency in diagnostic reviews and prediction models

Inconsistency refers to the unexplained heterogeneity of the results in meta-analysis. When estimates of diagnostic accuracy and prediction model parameters vary widely across studies (that is, there is heterogeneity or variability in results), this suggests

true differences in underlying effects. Inconsistency is, thus, only truly applicable when statistical meta-analysis is conducted (that is, results from different studies are pooled). No meta-analysis was performed for diagnostic reviews and prediction models in this guideline. However, 'no serious inconsistency' is nevertheless used to describe this quality assessment in the GRADE tables for outcomes from single studies.

Assessing indirectness in diagnostic reviews and prediction models

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

- participant selection
- index test
- reference standard.

More details about the QUADAS-2 tool can be found on the developer's website.

Assessing imprecision and clinical significance in diagnostic reviews and prediction models

The judgement of precision for diagnostic and prediction model evidence was based on the CI for test sensitivity as this was considered to be the primary measure of interest in this guideline. A difference in 95% confidence limits for sensitivity of 0-20 percentage points was considered to represent 'no imprecision', whereas differences of 20-40 percentage points and more than 40 percentage points were taken to represent 'serious imprecision' and 'very serious imprecision', respectively.

Prognostic studies

Adapted GRADE methodology for prognostic reviews

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

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

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

RR, relative risk; OR, odds ratio

Assessing risk of bias in prognostic reviews

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

- selection bias
- attrition bias
- prognostic factor bias
- outcome measurement bias
- control for confounders
- appropriate statistical analysis.

Assessing inconsistency in prognostic reviews

No meta-analysis was performed for prognostic reviews in this guideline. 'No serious inconsistency' was nevertheless used to describe this quality assessment in the GRADE tables for outcomes from single studies.

Assessing indirectness in prognostic reviews

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

Assessing imprecision and clinical importance in prognostic reviews

Prognostic studies may have a variety of purposes, for example, establishing typical prognosis in a broad population, establishing the effect of patient characteristics on prognosis, and developing a prognostic model. While by definition MIDs relate to treatment effects, the committee agreed to use GRADE default MIDs for intervention studies as a starting point from which to assess whether the size of an outcome effect in a prognostic study would be large enough to be meaningful in clinical practice.

Qualitative reviews

Adapted GRADE-CERQual methodology for qualitative reviews

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

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

Quality element	Description
	may have been conducted in a manner that by design would have not reached theoretical saturation at an individual study level

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

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

Assessing risk of bias in qualitative reviews

The risk of bias in qualitative studies was assessed using the Critical Appraisal Skills Programme (CASP) checklist for qualitative studies (see Appendix H in Developing NICE guidelines: the manual; NICE 2014). The overall risk of bias was derived by assessing the risk of bias across the 6 domains summarised in Table 10.

Table 10: Risk of bias 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)

Assessing relevance of evidence in qualitative reviews

Relevance (applicability) of findings in qualitative research is the equivalent of indirectness for quantitative outcomes, and refers to how closely the aims and context of studies contributing to a theme reflect the objectives outlined in the guideline review protocol.

Assessing coherence of findings in qualitative reviews

For qualitative research, a similar concept to inconsistency is coherence, which refers to the way findings within themes are described and whether they make sense. This concept was used in the quality assessment across studies for individual

themes. This does not mean that contradictory evidence was automatically downgraded, but that it was highlighted and presented, and that reasoning was provided. Provided the themes, or components of themes, from individual studies fit into a theoretical framework, they do not necessarily have to reflect the same perspective. It should, however, be possible to explain these by differences in context (for example, the views of healthcare professionals might not be the same as those of family members, but they could contribute to the same overarching themes).

Assessing adequacy of data in qualitative reviews

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

Assessing clinical significance in qualitative reviews

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

Evidence statements

Evidence statements are presented after the GRADE tables in each evidence report. They summarise key features in the available clinical evidence. The wording reflects the certainty or uncertainty in the estimate of effect (quantitative evidence) or review finding (qualitative evidence). Evidence statements are presented by outcome or theme, and encompass the following features in the evidence:

- the quality of the evidence
- the numbers of studies and participants for the outcome concerned or prognostic/risk factor or prediction model (quantitative evidence) or that contributed to themes (qualitative evidence)
- a brief description of the participants
- where relevant, an indication of the direction of effect (for example, if a treatment is beneficial or harmful compared with another, or whether there is no difference between the tested treatments or a summary of the effect size of the prognostic/risk factor or accuracy of the prediction model)
- where relevant, whether or not the estimate of effect is clinically important.

Reviewing economic evidence

Inclusion and exclusion of economic studies

A global health economic literature search was undertaken for women at high risk of adverse outcomes for themselves and/or their baby because of existing maternal medical conditions. This covered all 26 review questions considered in this part of the guideline.

Two further health economic literature searches were undertaken for women at high risk of adverse outcomes for themselves and/or their baby because of obstetric complications or other reasons:

- a global search that covered all 17 review questions considered in this part of the guideline
- a search tailored specifically to the review question about clinical and cost effectiveness of antimicrobial therapy for women in labour with sepsis.

Titles and abstracts of articles identified through the economic literature searches were independently assessed for inclusion using the predefined eligibility criteria listed in Table 11.

Table 11: Inclusion and exclusion criteria for systematic reviews of economic evaluations

Inclusion criteria
Intervention or comparators in accordance with the guideline scope
Study population in accordance with the guideline scope
Full economic evaluations (cost-utility, cost-effectiveness, cost-benefit or cost-consequence analyses) assessing both costs and outcomes associated with interventions of interest
Exclusion criteria
Abstracts containing insufficient methodological details
Cost-of-illness type studies

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

Details of economic evidence study selection, lists of excluded studies, economic evidence tables, the results of quality assessment of economic evidence (see below) and health economic evidence profiles are presented in Supplement 2 (Health economics).

Appraising the quality of economic evidence

The quality of economic evidence was assessed using the economic evaluations checklist specified in Developing NICE guidelines: the manual (NICE 2014). See Supplement 2 (Health economics) for further details.

Health economic modelling

The aims of the health economic input to the guideline were to inform the guideline committees of potential economic issues to ensure that recommendations represented a cost effective use of healthcare resources. Health economic evaluations aim to integrate data on healthcare benefits (ideally in terms of quality-adjusted life-years; QALYs) with the costs of different care options. In addition, the health economic input aimed to identify areas of high resource impact; these are recommendations which (while cost effective) might have a large impact on Clinical Commissioning Group or Trust finances and so need special attention.

For women at high risk of adverse outcomes for themselves and/or their baby because of existing maternal medical conditions, the guideline committee prioritised the following review questions where it was thought that economic considerations would be particularly important in formulating recommendations.

- Does antenatal care planning for birth involving an expanded multidisciplinary team compared with routine antenatal care planning improve intrapartum outcomes for women with existing medical conditions?
- Does an ultrasound scan of the woman's back improve needle siting for central neuraxial blockade anaesthesia and analgesia for women with obesity in the peripartum period?
- What additional equipment is needed to ensure optimal care of women with obesity in the peripartum period?

Clinical effectiveness evidence was identified for the review question about antenatal care planning involving a multidisciplinary team for women with existing medical conditions and for the question about ultrasound needle siting of central neuraxial blockade for women with obesity. Original health economic modelling was undertaken for both of these questions. In the absence of clinical effectiveness evidence for the question about equipment needs for women with obesity, cost analyses were undertaken.

For women at high risk of adverse outcomes for themselves and/or their baby because of obstetric complications or other reasons the guideline committee prioritised the following review questions for economic considerations.

- What is the optimal mode of birth (emergency caesarean section or continuation of labour) for women with breech presenting in the first or second stage of labour?
- What is the optimal mode of birth (emergency caesarean section or continuation of labour) for women with a large for gestational age baby?

Clinical effectiveness evidence was identified for the review question about mode of birth for women with a large for gestational age baby and original health economic modelling was undertaken for this question. Clinical effectiveness evidence was identified for the review question about mode of birth for women with breech presenting in labour, but original health economic modelling was not undertaken because of the high risk of selection bias in the included studies.

Cost effectiveness criteria

NICE's report Social value judgements: principles for the development of NICE guidance sets out the principles that committees should consider when judging whether an intervention offers good value for money. In general, an intervention was considered to be cost effective if any of the following criteria applied (provided that the estimate was considered plausible):

- the intervention dominated other relevant strategies (that is, it was both less costly in terms of resource use and more clinically effective compared with all the other relevant alternative strategies)
- the intervention cost less than £20,000 per QALY gained compared with the next best strategy
- the intervention provided clinically important benefits at an acceptable additional cost when compared with the next best strategy.

The committees' considerations of cost effectiveness are discussed explicitly under the heading 'Consideration of economic benefits and harms' in the relevant evidence reports.

Details of the cost effectiveness analyses undertaken for the guideline are presented in Supplement 2 (Health economics).

Developing recommendations

Guideline recommendations

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

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

For further details refer to Developing NICE guidelines: the manual (NICE 2014).

Research recommendations

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

Validation process

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

Updating the guideline

Following publication, NICE will undertake a surveillance review to determine whether the evidence base has progressed sufficiently to consider altering the guideline recommendations and warrant an update. For further details refer to Developing NICE guidelines: the manual (NICE 2014).

Funding

The NGA was commissioned by NICE to develop this guideline.

References

Dixon-Woods 2005

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

Hayden 2013

Jill A. Hayden, Danielle A. van der Windt, Jennifer L. Cartwright, Pierre Côté, Claire Bombardier. Assessing Bias in Studies of Prognostic Factors. Ann Intern Med. 2013;158:280–286. doi: 10.7326/0003-4819-158-4-201302190-00009

Higgins 2011

Higgins JPT, Green S (editors) (2011) Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011] The Cochrane Collaboration. Available from www.handbook.cochrane.org (accessed 15 June 2018)

Lewin 2015

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

NICE 2014

National Institute for Health and Care Excellence (NICE) (2014) Developing NICE guidelines: the manual (updated 2017). Available from https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview (accessed 15 June 2018)

NICE 2018

National Institute for Health and Care Excellence (NICE) (2014) NICE Policy on conflict of interest (updated 2017). Available from https://www.nice.org.uk/Media/Default/About/Who-we-are/Policies-and-procedures/declaration-of-interests-policy.pdf (accessed 15 June 2018)

Santesso 2016

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