National Institute for Health and Care Excellence

Final

Inducing labour

Development of the guideline and methods

NICE guideline NG207

Supplement 5

November 2021

Final

Developed by the National Guideline Alliance which is a part of the Royal College of Obstetricians and Gynaecologists

NICE accredited
www.nice.org.uk/accreditation

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 All NICE guidance is subject to regular review and may be updated or withdrawn.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 2021. All rights reserved. Subject to Notice of rights

ISBN: 978-1-4731-4327-2

Contents

Development of the guideline

Remit

The National Institute for Health and Care Excellence (NICE) commissioned the National Guideline Alliance (NGA) to update the existing NICE clinical guideline on Inducing labour (CG70, July 2008).

What this guideline update covers

Groups that are covered

- Women undergoing induction of labour for the following reasons:
  - prolonged pregnancy
  - preterm rupture of membrane
  - prelabour rupture of membranes
  - presence of fetal growth restriction
  - previous caesarean section
  - history of precipitate labour
  - maternal request
  - breech presentation
  - intrauterine fetal death
  - suspected macrosomia

Clinical areas that are covered

The 2021 update to the guideline covers the following clinical issues:

- The gestational age at which induction of labour should be offered if spontaneous labour does not ensue
- The benefits and harms associated with induction of labour in women with suspected fetal macrosomia
- Methods to induce labour in women with intrauterine fetal death who have had a previous caesarean birth
- The benefits and harms associated with pharmacological and mechanical methods of induction of labour in women

For further details please refer to the surveillance report on the NICE website that defined which sections of this guideline should be updated.

What this guideline update does not cover

Clinical areas that are not covered

This guideline update does not cover the following clinical issues:

- information and decision-making
- induction of labour in clinical circumstances, other than fetal macrosomia and intrauterine fetal death after previous caesarean birth
- setting of induction of labour

- monitoring and pain relief
- prevention and management of complications

Methods

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).

Declarations of interest were recorded according to NICE's 2014 conflicts of interest policy until 31 March 2018. From 1 April 2018, declarations of interest were recorded according to NICE's 2018 conflicts of interest policy. Those interests declared until April 2018 were reclassified according to NICE's 2018 conflicts of interest policy (see Register of Interests).

Developing the review questions and outcomes

The 4 review questions included in the update to the guideline were based on the key areas identified by the NICE surveillance program as requiring an update. Two questions were identified by a routine surveillance report, 1 question was flagged by relevant stakeholders during development as requiring an update, and 1 was flagged by the guideline committee during development as requiring an update, due to the publication of new evidence. The review questions were drafted by the NGA technical team and were refined and validated by the committee. Originally, there were two separate questions on the pharmacological methods and the mechanical methods to induce labour, but the committee highlighted that mechanical and pharmacological methods of inducing labour were often combined or considered as direct alternatives, and therefore the review of methods for the induction of labour should be combined, and this was agreed with NICE.

The review questions were based on the following framework for intervention reviews:

- population, intervention, comparator and outcome (PICO).

This framework guided the development of the review protocols, the literature searching process, the critical appraisal and synthesis of evidence. It also facilitated the development of recommendations by the committee.

Literature searches, critical appraisals and evidence reviews were completed for each review question.

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

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

Evidence review	Review question	Type of review
A	1. What are the benefits and harms of induction of labour in women with suspected fetal macrosomia?	Intervention
B	2. What are the benefits and harms of pharmacological and mechanical methods in induction of labour?	Intervention[1]

Evidence review	Review question	Type of review
C	3. At what gestational age should induction of labour be offered if spontaneous labour does not ensue?	Intervention
D	4. How should labour be induced in women with intrauterine fetal death who have had a previous caesarean birth?	Intervention

Original health economic analysis conducted for evidence review B

The COMET database was searched for core outcome sets relevant to this guideline. No core outcome sets were identified at the time of this search and therefore the outcomes for evidence review A, C and D were based on committee discussions. The outcomes for evidence review B were based on guidance from Cochrane, which had in turn been used to inform the outcomes of the previous Health Technology Assessment and upon which our review was based. An additional outcome (epidural) was added based on committee discussions.

Additional information on the network meta-analysis methods used in the development of the guideline is contained in appendix N of evidence review B on the pharmacological and mechanical methods of induction of labour.

Searching for evidence

Systematic literature search

Systematic literature searches were undertaken to identify all published clinical evidence relevant to each review question. This is a partial update of an existing guideline. New review protocols were drafted for the updated guideline, but the review protocols for the 2008 version of the guideline were taken into consideration at this stage. Evidence presented in the existing guideline was considered according to the new review protocol, and included in the updated guideline if it met the inclusion criteria for an individual review.

Databases were searched using subject headings, free-text terms and, where appropriate, study type filters were used. Where possible, searches were restricted to retrieve articles published in English. All searches were conducted in the following databases: Medline, Medline-in-process, Embase, Cochrane Central Register of Controlled Trials (CCTR), and Cochrane Database of Systematic Reviews (CDSR). Some searches were conducted in the following databases: Health Technology Assessments (HTA), and Database of Abstracts of Reviews of Effects (DARE). No date restrictions were placed on the searches for review A or review C. The searches for review B were restricted to 2014 onwards since a combined systematic review, network meta-analysis (NMA) and cost-effectiveness study (Alfirevic 2016) had run searches up until March 2014. The searches for this study were assessed and deemed to be robust. The searches for review D were restricted to 2007 onwards, to cover the period from when they were last run for the 2008 guideline.

Searches were run once for all reviews during development. Searches for evidence reviews A and B were updated in May 2020. Searches for evidence reviews C and D were not re-run because it was not anticipated that additional evidence would be

available that would lead to changes in the recommendations in the short timeframe over which this update was carried out.

Any studies added to the databases after the date of the search (even those published prior to this date) were not included unless specifically stated in the text.

Details of the search strategies, including study type filters that were applied and databases that were searched, can be found in appendix B of each evidence report.

Searching for grey literature or unpublished literature was not undertaken.

Economic systematic literature search

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

Searches using the search strategies derived from the review questions, combined with a search filter for economic evaluations, were conducted in Medline, Medline-in-Process, CCTR and Embase. A single search, using the population search terms used in the evidence reviews, was also conducted in the NHS Economic Evaluation Database (NHS EED) and HTA. Where possible, searches were limited to studies published in English.

Searches were run once for all reviews during development. Searches for evidence reviews A and B were updated in May 2020. Searches for evidence reviews C and D were not re-run because it was not anticipated that additional evidence would be available that would lead to changes in the recommendations in the short timeframe over which this update was carried out.

Details of the search strategies, including study type filters that were applied and databases that were searched, can be found in appendix B of each evidence report.

Quality assurance

Search strategies were quality assured by cross-checking reference lists of relevant studies, analysing search strategies from published systematic reviews and asking members of the committee to highlight key studies. The principal search strategies for each search were also quality assured by a second information scientist using an adaptation of the PRESS 2015 Guideline Evidence-Based Checklist (McGowan 2016).

Reviewing evidence

Systematic review process

The evidence was reviewed following these steps.

- Potentially relevant studies 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 papers were reviewed against pre-specified inclusion and exclusion criteria in the review protocols (see appendix A of each evidence review).

- Key information was extracted on the 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 review and in a more detailed evidence table (see appendix D of each evidence review).
- Included studies were critically appraised using an appropriate checklist as specified in Developing NICE guidelines: the manual (NICE). Further detail on appraisal of the evidence is provided below.
- Summaries of evidence by outcome were presented in the corresponding evidence review and discussed by the committee.

The review question informing the NMA was selected as a high priority for economic analysis and was subject to dual screening and study selection through a 10% random sample of articles. In addition, data extraction and critical appraisal for this review question was carried out in duplicate by 2 independent reviewers. Any discrepancies in screening, study selection, data extraction or critical appraisal were resolved by discussion between the first and second reviewers or by reference to a third (senior) reviewer. Additional specific methods for this review question are described in evidence review B. For the remaining review questions, internal (NGA) quality assurance processes included consideration of the outcomes of screening, study selection and data extraction and the 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 (SRs) with meta-analyses were considered the highest quality evidence to 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 were considered for inclusion.

The 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 K of the corresponding evidence review.

Narrative reviews, posters, letters, editorials, comment articles, unpublished studies and studies published in languages other than English were excluded. Conference abstracts were only considered for inclusion in evidence review B for consistency with the approach taken by the authors of the NMA and cost-effectiveness study (Alfirevic 2016). Conference abstracts for evidence reviews A, C and D were not considered for inclusion because these do not typically have sufficient information to allow full critical appraisal.

Methods of combining evidence

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

Data synthesis for intervention reviews

Pairwise meta-analysis

Meta-analysis to pool results from RCTs was conducted where possible using Cochrane Review Manager (RevMan5) software.

For dichotomous outcomes, such as mortality, the Mantel–Haenszel method with a fixed effect model was used to calculate risk ratios (RRs). For all outcomes with zero events in both arms the risk difference was presented. For outcomes in which the majority of arms had low event rates (<1%), Peto odds ratios (ORs) were calculated as this method performs well when events are rare (Bradburn 2007).

For continuous outcomes, a measure 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.

While continuous outcomes were considered and searched for, the majority of evidence included in both reviews in this update was dichotomous in nature.

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 as for data presented as means with SDs. Subjective assessments for the various GRADE domains (see below) were made based on all pertinent available information (for example the sample size and range of data).

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 E of relevant evidence reviews).

Network meta-analysis

In the review looking at the effectiveness of the different pharmacological and mechanical methods of inducing labour, critical outcomes were synthesised using NMA techniques with the NMA review methods described in the relevant evidence review (B), appendix N.

We performed a Hierarchical Bayesian Network Meta-Analysis (NMA) using WinBUGS version 1.4.3, based on the fixed and random effects models available from NICE Decision Support Unit (DSU) technical support document number 2: http://nicedsu.org.uk/wp-content/uploads/2016/03/A-general-linear-modelling-

framework-for-pair-wise-and-network-meta-analysis-of-randomised-controlled-trials..pdf

For the analyses, a series of burn-in simulations was run to allow the posterior distributions to converge and then further simulations were run to produce the posterior outputs. Convergence was assessed by examining the history, autocorrelation and Brooks-Gelman-Rubin plots.

Goodness-of-fit of the model was also estimated by using the posterior mean of the sum of the deviance contributions for each item by calculating the residual deviance and deviance information criteria (DIC). If the residual deviance was close to the number of unconstrained data points (the number of trial arms in the analysis) then the model was explaining the data at a satisfactory level. The choice of a fixed effect or random effects model can be made by comparing their goodness-of-fit to the data. Treatment-specific posterior effects were generated for every possible pair of comparisons by combining direct and indirect evidence in each network. The probability that each treatment is best, based on the proportion of Markov chain iterations in which the treatment effect for an intervention is ranked best, second best and so forth. This was calculated by taking the treatment effect of each intervention compared to the reference treatment and counting the proportion of simulations of the Markov chain in which each intervention had the highest treatment effect.

One of the main advantages of the Bayesian approach is that the method leads to a framework that supports decision making. The Bayesian approach also allows the probability that each intervention is best for achieving a particular outcome, as well as its ranking, to be calculated.

We adapted standard fixed and random effects models available from NICE Decision Support Unit (DSU) technical support document number 2.

To determine if there is evidence of inconsistency, the selected consistency model (fixed or random effects) was compared to an “inconsistency”, or unrelated mean effects, model (see below). We performed further checks for evidence of inconsistency through node-splitting.

For further description of the model used, specific methods, outcomes and the results of the NMA please see the evidence review for question 2, evidence review B.

The running (of all NMAs except for use of epidural which was run by the NGA), inconsistency checking and quality assurance of all the NMA work was undertaken by the NICE Guidelines Technical Support Unit, University of Bristol (TSU).

Appraising the quality of evidence

Intervention reviews

Pairwise meta-analysis

GRADE methodology for intervention reviews

For intervention reviews, the evidence for outcomes from included RCTs and comparative non-randomised studies was evaluated and presented using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology developed by the international GRADE working group.

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 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 non-randomised 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 non-randomised 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')	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

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 a consistent deviation from the truth in the results. 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 as described in appendix H in Developing NICE guidelines: the manual (NICE).

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 the Cochrane Handbook for Systematic Reviews of Interventions (Higgins 2019).

For systematic reviews of RCTs the ROBIS checklist was used (see appendix H in Developing NICE guidelines: the manual (NICE).

Assessing inconsistency in intervention reviews

Inconsistency in GRADE terms refers to unexplained heterogeneity in results of meta-analysis (note this is distinct from the use of inconsistency specifically in the context of NMA, see below for more information). 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 visually by inspecting forest plots and observing whether there was considerable heterogeneity in the results of the meta-analysis (for

example if the point estimates of the individual studies consistently showed benefits or harms). This was supported 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 considerable heterogeneity was present, the meta-analysis was re-run using the Der-Simonian and Laird method with a random effects model and this was used for the final analysis.

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 importance in intervention reviews

Imprecision in GRADE methodology refers to uncertainty around the effect estimate and whether or not there is an 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 95% 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: Assessment of imprecision and importance in intervention reviews using GRADE, 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 minimal 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 important to people with the condition of interest (favours B).

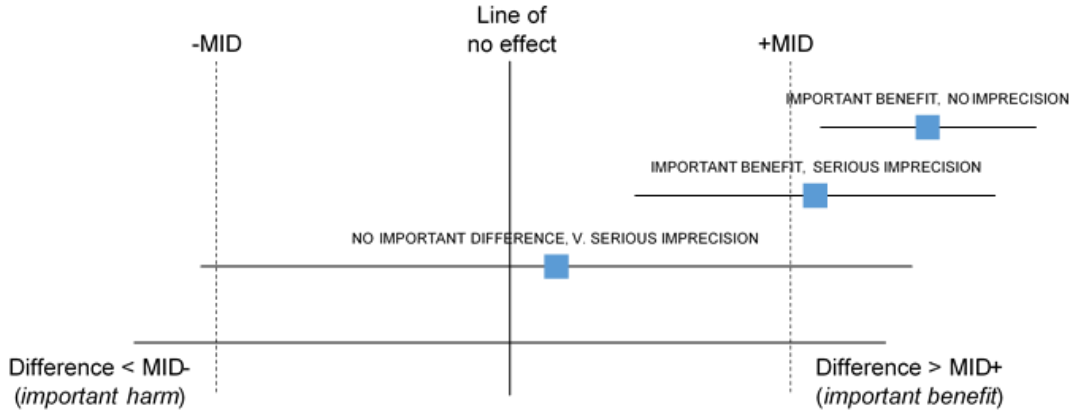
When the 95% CI of the effect estimate is wholly contained in 1 of the 3 zones there is no uncertainty about the size and direction of effect, therefore, the effect estimate is considered precise; that is, there is no imprecision.

When the 95% 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 95% 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 95% CI crosses all 3 zones, the effect estimate is considered to be very imprecise because the 95% CI is consistent with 3 possible 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 95% CI is in, or partially in, an 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 importance in intervention reviews using GRADE



MID, minimally important difference

Minimally important differences

The committee was asked whether there were any recognised or acceptable MIDs in the published literature and community relevant to the review questions under consideration. The committee was not aware of any MIDs that could be used for 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 minimally 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 dichotomous outcomes assessed by Peto OR considered in intervention evidence reviews, as OR and RR are mathematically very similar at low event rates (the principle indication for the use of Peto OR). For specific serious adverse events (such as neonatal mortality) the committee agreed to use statistical significance as the MID, such that any statistically significant increase/decrease in the outcome would be considered clinically important. These are described in the individual protocols for the

review questions. 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 was not available a baseline).

Where zero events occurred in either arm of the majority of studies contributing to an outcome, risk difference was used for meta-analysis. In this case the committee chose to use sample size to assess imprecision in the absence of ratio measure confidence intervals. The committee chose to use 300 and 500 as cut-offs to determine precision in this case based on the numbers used by convention for optimal information size assessments, in other words an outcome with a sample size >500 would be rated as no imprecision, >300 but less than or equal to 500 would be rated as serious imprecision and less than or equal to 300 would be rated as very serious imprecision.

Network meta-analysis

The GRADE approach is not yet well established for use with NMA. Therefore, for the NMAs, quality was assessed by looking at risk of bias across the included evidence using the Cochrane risk of bias tool. This is presented as a summary figure for each NMA result.

The consistency between direct and indirect evidence can be assessed in closed treatment loops within the network. These closed treatment loops are regions within a network where direct evidence is available on at least 3 different treatments that form a closed ‘circuit’ of treatment comparisons (for example, A versus B, B versus C, C versus A). If closed treatment loops existed, then discrepancies between direct and indirect evidence was assessed.

To determine if there is evidence of inconsistency, the selected consistency model (fixed or random effects) was compared to an “inconsistency”, or unrelated mean effects, model. The latter is equivalent to having separate, unrelated, meta-analyses for every pairwise contrast, with a common variance parameter assumed in the case of random effects models. Further checks for evidence of inconsistency either through Bucher’s method or node-splitting were undertaken. Bucher’s method compares the direct and indirect estimates for a contrast in a loop (for example, A-B-C) where the direct estimate of contrast B versus C is compared to its corresponding indirect estimate, which is informed from the direct estimates of the other contrasts in the loop (A versus B and A versus C). This method was used to assess consistency in networks, where there was a single loop and the network contained sparse evidence with zero events, limiting the stability of the results of more sophisticated methods such as the node-splitting method. The node-splitting method allowed the direct and indirect evidence contributing to an estimate of a relative effect to be split and compared. The consistency checks were undertaken by the TSU.

The TSU conducted threshold analyses for the NMA in this guideline. These analyses are a method for assessing the impact of potential bias and quantify how much the evidence in an analysis could change before the recommendation would be expected to change and what the revised recommendation would be. More details on the methods and results of these analyses are provided in the relevant evidence review.

Reviewing economic evidence

Reviewing economic evidence

Reviewing economic evidence

Titles and abstracts of articles identified through the economic literature searches were independently assessed for inclusion using the predefined eligibility criteria summarised in Table 5.

Table 5: Inclusion and exclusion criteria for the systematic reviews of economic evaluations

Inclusion criteria
Studies from Organisation for Economic Co-operation and Development (OECD) countries were included, as the aim of the review was to identify economic information transferable to the UK context
Study population matches scope
Clinical condition and interventions assessed identical to those considered in the clinical evidence review
Studies include sufficient details regarding methods and results to enable methodological quality to be assessed and results to be extracted
Full economic evaluations (cost utility, cost effectiveness, cost benefit or cost consequence analyses) that assess both the costs and outcomes associated with the interventions of interest
Exclusion criteria
Conference abstracts, poster presentations or dissertation abstracts with 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 the economic evidence study selection for each question, list of excluded studies, economic evidence tables, the results of quality assessment of economic evidence (see below) and health economic evidence profiles are presented in appendices G, K, H and I of the evidence report. Existing economic evidence considered in the guideline is provided in the respective evidence reviews.

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).

Economic modelling

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

The committee prioritised the following review question where it was thought that economic considerations would be particularly important in formulating recommendations:

- Evidence review B. What are the benefits and harms of pharmacological and mechanical methods in induction of labour?

For this guideline it was possible to update a previously published cost-effectiveness analysis (Alfirevic 2016). The methods and results of the updated economic analyses are reported in appendix J of the relevant evidence reports. When new economic analysis was not prioritised, the committee made a qualitative judgement regarding cost effectiveness by considering expected differences in resource and cost use between options, alongside clinical effectiveness evidence identified from the clinical evidence review.

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 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 important benefits at an acceptable additional cost when compared with the next best strategy.

The committee's considerations of cost effectiveness are discussed explicitly in the committee's discussion of the evidence section on 'Cost effectiveness and resource use'.

Details of the cost effectiveness analyses undertaken for the guideline are presented in appendix J of each evidence review.

Developing recommendations

Updating existing recommendations

Although a number of sections of the 2008 guideline had not been prioritised for updating by the NICE surveillance report, the committee identified some recommendations in these sections where practice had changed, new technology had become available, or health policy had changed. In addition the committee identified a number of recommendations which were not written in the current NICE style or terminology. As part of the update process the committee therefore reviewed the sections of the guideline which were not being formally updated and made minor edits to some of the recommendations to improve clarity, ensure they reflected current best practice, or correct recommendations that no longer were applicable. These changes are clearly marked in yellow in the guideline version for consultation, and the changes and reasons for them summarised in Table 2 of the update information at the end of the guideline.

Guideline recommendations

Recommendations were drafted on the basis of the committee's interpretation of the available evidence, taking account of the balance of benefits, harms and costs between different courses of action. When effectiveness 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, person'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 review.

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

Research recommendations

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

Validation process

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

Updating the guideline

Following publication, and in accordance with the NICE guidelines manual, NICE will undertake a review of whether the evidence base has progressed significantly to alter the guideline recommendations and warrant an update. For further details please refer to Developing NICE guidelines: the manual (NICE).

Funding

The NGA was commissioned by NICE to develop this guideline.

References

Alfirevic 2016

Alfirevic, Z., Keeney, E., Dowswell, T., Welton, N. J., Medley, N., Dias, S., Jones, L. V., Gyte, G., Caldwell, D. M., Which method is best for the induction of labour? A systematic review, network meta-analysis and cost-effectiveness analysis, Health Technology Assessment (Winchester, England)Health Technol Assess, 20, 1-584, 2016

DerSimonian 1986

DerSimonian, R, Laird, N. (1986). Meta-analysis in clinical trials. Controlled clinical trials, 7(3), 177-188

Bradburn 2007

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

Higgins 2011

Higgins JPT, Green S (editors) Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated 2019] The Cochrane Collaboration. Available from: www.handbook.cochrane.org (accessed 25 May 2021)

McGowan 2016

McGowan J, Sampson M, Salzwedel DM et al. PRESS Peer Review of Electronic Search Strategies: 2015 guideline statement. Journal of Clinical Epidemiology 75: 40–6, 2016

NICE 2010

National Institute for Health and Care Excellence (NICE), Inducing labour. Clinical guideline 70; 2008.

NICE

National Institute for Health and Care Excellence (NICE) (2014) Developing NICE guidelines: the manual (updated 2018 and 2020). Available from: Developing NICE guidelines: the manual (accessed 25 May 2020)

NICE 2018

National Institute for Health and Care Excellence (NICE) (2014) NICE Policy on conflicts of interest (updated 2017). Available from: Policy on declaring and managing interests for NICE advisory committees (accessed 25 May 2020)

Santesso 2016

Santesso, N., Carrasco-Labra, A., Langendam, M., Brignardello-Petersen, R., Mustafa, R.A., Heus, P., Lasserson, T., Opiyo, N., Kunnamo, I., Sinclair, D. and Garner, P., 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, 2016.