National Institute for Health and Care Excellence

Final

Abortion care

Supplement 1: Methods

NICE guideline NG140

Development of the guideline and methods

September 2019

Final

Evidence reviews were 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 All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

ISBN: 978-1-4731-3539-0

Contents

Development of the guideline

Remit

The National Institute for Health and Care Excellence (NICE) commissioned the National Guideline Alliance (NGA) to develop a new guideline on termination of pregnancy.

The guideline was developed using the methods and processes outlined in Developing NICE guidelines: the manual.

At the time of development, the title of this guideline was 'Termination of pregnancy' and this term was used throughout the guideline. In response to comments from stakeholders, the title was changed to 'Abortion care' and abortion has been used throughout. Therefore, both terms appear in this document.

What this guideline covers

Groups that are covered

- Women (of any age) requesting termination of pregnancy under the terms of the Abortion Act 1967 (as amended by the Human Fertilisation and Embryology Act 1990).
- Women (of any age) undergoing a termination of pregnancy as a life-saving procedure.

Specific consideration is given to women with complex pre-existing medical conditions.

Settings that are covered

- Settings licensed to provide termination of pregnancy services.
- All settings that provide publicly funded commissioned assessment for termination of pregnancy or care after the procedure.

For further details please refer to the scope on the NICE website.

Clinical areas covered

The guideline covers the following clinical areas.

- Assessment for termination of pregnancy
- Termination of pregnancy care
- Care after termination of pregnancy
- Service configuration

What this guideline does not cover

Settings that are not covered

- Settings that are not licensed to provide termination of pregnancy services.
- Settings that are not commissioned by publicly funded bodies to provide assessment for termination of pregnancy or care after the procedure.

Clinical areas that are not covered

- Care between conception and the request for termination of pregnancy
- Ongoing care for women who decide not to terminate their pregnancy

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.

Developing the review questions and outcomes

The 21 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 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: population, intervention, comparator and outcome (PICO)
- prognostic reviews: population, prognostic or predictive factor and outcome (PPO)
- qualitative reviews: population or problem, interest (that is, defined event, activity, experience or process) and context (PICo)

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

Full literature searches, critical appraisals 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	Key question(s) in scope	Review question	Type of review
[A] Accessibility and sustainability of abortion services	Service configuration	What strategies ensure the sustainability of a safe and accessible termination of pregnancy service? What strategies enhance access to termination of pregnancy services?	What factors help or hinder the accessibility and sustainability of a safe abortion service?	Qualitative
[A] Accessibility and sustainability of abortion services	Service configuration	What strategies ensure the sustainability of a safe and accessible termination of pregnancy service? What strategies enhance access to	What strategies that improve the factors that help or hinder the accessibility and sustainability of a safe abortion service?	Intervention

Evidence report	Subtopic in scope	Key question(s) in scope	Review question	Type of review
		termination of pregnancy services?		
[B] Information needs of women undergoing am abortion	Assessment for termination of pregnancy	What information should women who have requested a termination of pregnancy be given before they have the procedure?	What information would women who have requested an abortion like?	Qualitative
[C] Anti-D prophylaxis for women up to 13^{+6} weeks' gestation	Termination of pregnancy	Should women who are Rhesus negative and having termination of a first trimester pregnancy receive Rhesus prophylaxis?	Should women who are RhD (or D) negative and having an abortion up to 13^{+6} weeks' gestation receive anti-D prophylaxis?	Intervention
[D] Antibiotic prophylaxis for medical and surgical abortion	Termination of pregnancy	What is the optimal antibiotic prophylaxis regimen (including no antibiotic prophylaxis as an option) for women who are having medical termination of pregnancy?	What is the optimal antibiotic prophylaxis regimen (including no antibiotic prophylaxis as an option) for women who are having medical abortion?	Intervention
[D] Antibiotic prophylaxis for medical and surgical abortion	Termination of pregnancy	What is the optimal antibiotic prophylaxis regimen for women who are having surgical termination of pregnancy?	What is the optimal antibiotic prophylaxis regimen for women who are having surgical abortion?	Intervention
[E] VTE prophylaxis for women having abortion	Termination of pregnancy	Question not in scope	In women who are undergoing an abortion up to 24 weeks' gestation, and who are identified as requiring pharmacological thromboprophylaxis, what is the optimal timing and duration of VTE prophylaxis?	Intervention
[F] Abortion before ultrasound evidence	Termination of pregnancy	Is it safe and effective to start termination before there is ultrasound evidence of an intrauterine pregnancy?	Is it safe and effective to start abortion before there is ultrasound evidence of an intrauterine pregnancy?	Intervention
[G] Expulsion at home for early medical abortion	Termination of pregnancy	For women who are having medical termination of pregnancy, what gestational limit for expulsion at home offers the best balance of benefits and harms?	For women who are having medical abortion, what gestational limit for expulsion at home (i.e., setting outside of clinical facility) offers the best balance of benefits and harms?	Prognostic
[H] Simultaneous versus delayed mifepristone + misoprostol administration for medical abortion up to 10^{+0} weeks' gestation	Termination of pregnancy	For women who are having an early (up to 10 weeks) medical termination of pregnancy, what is the effectiveness, safety and acceptability of mifepristone and misoprostol given simultaneously compared with other time intervals?	For women who are having an early (up to 10^{+0} weeks' gestation) medical abortion, what is the effectiveness, safety and acceptability of mifepristone and misoprostol given simultaneously compared with other time intervals?	Intervention

Evidence report	Subtopic in scope	Key question(s) in scope	Review question	Type of review
[I] Follow-up after medical abortion up to 10^{+0} weeks	Care after termination of pregnancy	What is the best method of excluding an ongoing pregnancy after early (up to 10 weeks) medical termination of pregnancy, when the expulsion has not been witnessed by healthcare professionals (for example, expulsion at home)?	What is the best method of excluding an ongoing pregnancy after early (up to 10^{+0} weeks) medical abortion, when the expulsion has not been witnessed by healthcare professionals (for example, expulsion at home)?	Intervention
[J] Misoprostol after mifepristone for inducing medical abortion between 10^{+1} and 24^{+0} weeks' gestation	Termination of pregnancy	What is the optimal dose and route of administration of misoprostol after mifepristone, for inducing medical termination in the second trimester?	What is the optimal regimen and route of administration of misoprostol after mifepristone, for inducing medical abortion from 10^{+1} to 24^{+0} weeks?	Intervention
[K] Medical versus surgical abortion between 13^{+0} and 24^{+0} weeks' gestation	Termination of pregnancy	What is the effectiveness, safety and acceptability of surgical compared to medical termination in the second trimester?	What is the effectiveness, safety and acceptability of surgical compared to medical abortion between 13^{+0} and 24^{+0} weeks' gestation?	Intervention
[L] Medical abortion after 24 weeks' gestation	Termination of pregnancy	What is the optimal regimen for termination of pregnancy after 24 weeks, for example, for fetal anomaly?	What is the optimal regimen for medical abortion after 24 weeks' gestation?	Intervention
[M] Cervical priming before surgical abortion	Termination of pregnancy	What is the optimal regimen for cervical priming (including no cervical priming as an option) before surgical termination of pregnancy in the first trimester?	What is the optimal regimen for cervical priming (including no cervical priming as an option) before surgical abortion up to and including 13^{+6} weeks' gestation?	Intervention
[M] Cervical priming before surgical abortion	Termination of pregnancy	What is the optimal regimen for cervical priming before surgical termination of pregnancy in the second trimester?	What is the optimal regimen for cervical priming before surgical abortion between 14^{+0} and 24^{+0} weeks' gestation?	Intervention
[N] Anaesthesia or sedation for surgical abortion	Termination of pregnancy	What is the optimal method of anaesthesia or sedation for surgical termination of pregnancy?	What is the optimal method of anaesthesia or sedation for surgical abortion?	Intervention
[O] Support after abortion	Care after termination of pregnancy	What support should women be offered after a termination of pregnancy?	What support would women like after an abortion?	Qualitative
[P] Contraception after medical abortion	Service configuration	What strategies are effective at facilitating uptake of effective contraception after termination of pregnancy?	What strategies are effective at facilitating access to contraception after abortion?	Intervention
[P] Contraception after abortion	Termination of pregnancy	For women who are having medical termination of pregnancy and plan to use a progestogen-only contraceptive	For women who are having medical abortion and plan to use a progestogen-only contraceptive implant or depot	Intervention

Evidence report	Subtopic in scope	Key question(s) in scope	Review question	Type of review
		implant or depot injection, does administration of the contraception at the same time as mifepristone influence the efficacy of the termination?	injection, does administration of the contraception at the same time as mifepristone influence the efficacy of the abortion?	
[P] Contraception after medical abortion	Care after termination of pregnancy	For women who have had medical termination of pregnancy, how soon afterwards is it safe to insert an intrauterine contraceptive device?	For women who have had a medical abortion, how soon afterwards is it safe to insert an intrauterine contraceptive device?	Intervention

RhD: Rhesus D; VTE: Venous thromboembolism

Searching for evidence

Clinical search literature

Systematic literature searches were undertaken to identify all published clinical evidence relevant to the review questions.

Databases were searched using relevant medical subject headings, free-text terms and study type filters where appropriate. Studies published in languages other than English were not reviewed, and a standard exclusions filter was applied (letters, animals, etc). All searches were conducted in MEDLINE, Embase and The Cochrane Library, with some additional database searching in CINAHL, PsycINFO and Web of Science Core Collection for:

- the contraceptive topics (Evidence Review P)
- information and support needs (Evidence Review B and O)
- expulsion at home (Evidence Review G)
- access and sustainability of services topics (Evidence Review A)

Re-run searches were carried out in November 2018, but were not conducted:

- When the initial search was completed in October 2018
- For the information and support topics where we would be unlikely to find additional evidence to change our recommendations given the nature of the evidence used in these reviews
- For surgical versus medical abortion as the results of the economic model and corresponding sensitivity analysis suggest very strongly that the conclusions are unlikely to change as a result of any update search. Moreover, the Guideline Committee were not aware of any new relevant studies
- When the evidence base wasn't fast-moving such as antibiotic prophylaxis, anaesthesia/sedation for abortion, medical abortion and access to contraception

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

Search strategies were quality assured by cross-checking reference lists of relevant papers, analysing search strategies in other systematic reviews and asking committee members to highlight any additional studies. The questions, the study types applied, the databases searched and the years covered can be found in appendix B in each evidence review chapter.

Searching for grey literature or unpublished literature was not undertaken. During the scoping stage, a search was conducted for guidelines and reports on websites of organisations relevant to the topic. Any references suggested by stakeholders at the scoping consultation were considered. Clinical search strategies can be found in appendix B of each evidence review.

Health economics search literature

A global search of economic evidence was undertaken and re-run in November 2018. The following databases were searched:

- MEDLINE (Ovid)
- EMBASE (Ovid)
- Health Technology Assessment database (HTA)
- NHS Economic Evaluation Database (NHS EED)

Further to the database searches, the committee was contacted with a request for details of relevant published and unpublished studies of which they may have knowledge; reference lists of key identified studies were also reviewed for any potentially relevant studies. Finally, the NICE website was searched for any recently published guidance relating to abortion that had not been already identified via the database searches.

The search strategy for existing economic evaluations combined terms capturing abortion and, for searches undertaken in MEDLINE and EMBASE, terms to capture economic evaluations. No restrictions on language or setting were applied to the economic evidence search, but a standard exclusions filter was applied (letters, animals, etc). Full details of the search strategy are presented in Supplementary material -2: Health economics..

Reviewing clinical 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 D of each evidence report).
- Included studies were critically appraised using an appropriate checklist as specified in Developing NICE guidelines: the manual 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 guideline committee.

Review questions which were not intervention reviews, the first intervention review undertaken and any complex intervention reviews were subject to dual screening through a 10% random sample of articles. These tended to include the review

questions selected as high priorities for health economic analysis. 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 guideline 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 or SRs of qualitative studies with thematic syntheses 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 and/or observational studies were considered for inclusion, including cohort studies, before-and-after studies, and cross-sectional studies.

For prognostic reviews, prospective and retrospective cohort studies 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 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 report.

Narrative reviews, posters, letters, editorials, comment articles, unpublished studies and studies published in languages other than English were excluded.

Methods of combining evidence

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

Data synthesis for intervention reviews

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 or random effects model was used to calculate relative risks (RRs).

For continuous outcomes, such as duration of hospital stay, measures of central tendency (mean) and variation (standard deviation; SD) are required for meta-analysis. Data for continuous outcomes 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.

Subgroups for stratified analyses were agreed for the 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 reports).

Data synthesis for prognostic or predictive factor reviews

Univariate RRs with their 95% CIs were calculated and results were plotted with their 95% CIs in forest plots in Review Manager. Wherever possible, the results were pooled. Results from unadjusted and adjusted analyses for key confounders were considered.

Data synthesis for qualitative reviews

Where possible, a meta-synthesis was conducted to combine evidence from qualitative studies. 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.

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. The names of themes and overarching categories of themes were assigned 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. However, additional studies may provide further information that improves the adequacy of the data. 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 if their inclusion would not improve the adequacy of the data, because the emphasis was to be on conceptual robustness rather than quantitative completeness of the evidence. This has 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). There was limited evidence available for 2 of the qualitative reviews considered in this guideline (information needs and support after abortion), and so the methods for

managing data saturation were only applied to 1 review question (factors that help or hinder the accessibility and sustainability of a safe abortion service).

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 reviews

GRADE methodology (the Grading of Recommendations Assessment, Development and Evaluation)

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.

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 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

Quality element	Description
	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 of outcome evidence in GRADE	Description
High	Further research is very unlikely to change our confidence in the estimate of effect.
Moderate	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low	Any 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 (see appendix H in Developing NICE guidelines: the manual 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 would have been used and for systematic reviews of other study types the ROBIS checklist would have been used, but none of either type were actually included (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 serious 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 serious heterogeneity, and more than 80% indicating very serious heterogeneity. When serious 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. If the I-squared statistics was above 50% or 80%, respectively, the evidence was downgraded by 1 or 2 levels for inconsistency. Moreover, if the I-squared statistic was above 80% the data were not pooled, but instead the individual study estimates were presented.

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).

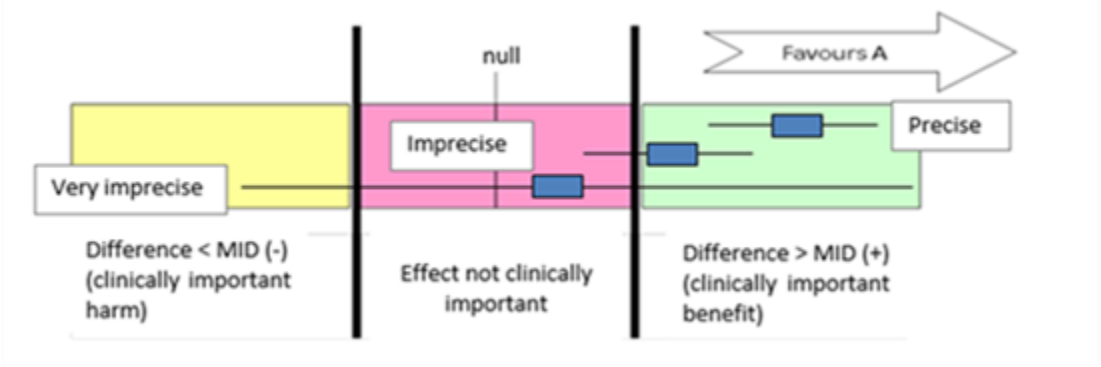
When the CI of the effect estimate is wholly contained in 1 of the 3 zones there is no uncertainty about the size and direction of effect, therefore, the effect estimate is considered precise; that is, there is no imprecision.

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

The guideline committee was asked whether there were any recognised or acceptable MIDs in the clinical literature and community relevant to the review questions under consideration. For some outcomes such as 'need for emergency care/hospital admission' and 'haemorrhage requiring transfusion or > 500 ml of blood

loss', the committee decided to use statistical significance as the MID. For the outcome, 'complete abortion without the need for surgical intervention', the committee agreed to use 3 percentage points as the MID.

When statistical significance was used as the MID, the imprecision ratings were undertaken on that basis by using the optimal information size so that for dichotomous outcomes if the total event rate was ≥300, then the quality was not downgraded, if the event rate was150-299, then the quality was downgraded by 1 level and if the event rate was <150, then the quality was downgraded by 2 levels. However, the GRADE handbook states that, if sample sizes are sufficiently large it is likely that prognostic balance will be achieved even if event rates are low. Therefore, if n was ≥4000, dichotomous outcomes were not downgraded, regardless of the number of events. For continuous outcomes if the sample size n, was ≥400, then the quality was not downgraded, if n was 200-399, then the quality was downgraded by 1 level and if n was <200, then the quality was downgraded by 2 levels.

When the MID was 3 percentage points, the imprecision ratings were undertaken by using the absolute effect estimates so that if the CI crossed 30 fewer (3% of 1000) or 30 more, then the quality was downgraded by 1 level. If the CI crossed both thresholds (fewer and more), then the quality was downgraded by 2 levels.

For outcomes without 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).

For outcome which were only reported as medians and ranges, for which there are no established or default GRADE MIDs, the imprecision ratings were undertaken by using the optimal information size so that if n was ≥400, then the quality was not downgraded, if n was 200-399, then the quality was downgraded by 1 level and if n was <200, then the quality was downgraded by 2 levels.

Prognostic reviews

GRADE methodology for prognostic reviews

For prognostic reviews with evidence from comparative observational studies an adapted GRADE approach was used.

The evidence for each outcome in the prognostic reviews 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 '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 5: 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

Quality element	Description
	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 relative risks or odds ratios), 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

Assessing risk of bias in prognostic reviews

Risk of bias in individual prognostic studies was assessed using the Newcastle-Ottawa scale, which evaluates the risk of selection bias, the comparability between the comparison groups and the adequacy of outcome assessment, including length of follow up. If bias was identified, the evidence was downgraded by 1 or 2 levels, depending on the extent of bias.

Assessing inconsistency in prognostic reviews

Inconsistency denotes unexplained heterogeneity in a meta-analysis and was assessed by calculating the I-squared statistics of the pooled effect estimate. If the I-squared statistics was above 50% or 80%, respectively, the evidence was downgraded by 1 or 2 levels for inconsistency. Moreover, if the I-squared statistic was above 80% the data were not pooled, but instead the individual study estimates were presented.

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 for prognostic reviews

For the outcomes ‘need for emergency care/hospital admission’ and ‘haemorrhage requiring transfusion or > 500 ml of blood loss’ a statistically significant difference between the groups was considered a clinically significant effect. For these dichotomous outcomes imprecision was assessed using the optimal information size and the evidence was downgraded by 1 level if the total number of events (across both arms) was less than 300, and by 2 levels if the total number of events was less than 150. However, the GRADE handbook states that, if sample sizes are sufficiently large it is likely that prognostic balance will be achieved even if event rates are low. Therefore, if n was ≥4000, dichotomous outcomes were not downgraded, regardless of the number of events. For the outcome ‘complete abortion without the need for surgical intervention’ a statistically significant difference of at least 3 percentage points was considered a clinically significant effect. For this outcome imprecision was assessed using the absolute effect estimate and the evidence was downgraded by 1

level if the CI crossed 30 fewer (3% of 1000) or 30 more, and by 2 levels if it crossed both.

Qualitative reviews

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. Quality elements assessed using GRADE-CERQual are listed and defined in Table 6. Each element was graded using the levels of concern summarised in Table 7. 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 8

Table 6: Adaptation of GRADE quality elements for qualitative reviews

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

Table 7: GRADE-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 8: Overall confidence in the evidence in GRADE-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 10 domains summarised in Table 9.

Table 9: Risk of bias in qualitative studies

Aims of the research	This domain assesses whether the aims, importance and relevance of the study were described clearly
Appropriateness of using qualitative methodology	This domain assesses whether qualitative research methods were appropriate for investigating the research question, for example, does the study aim to interpret or illuminate actions or subjective experiences
Research design	This domain assesses whether the study approach has been documented clearly and if it was justified, for example, based on a theoretical framework
Recruitment strategy	This domain assesses the procedure and reasons for the method of selecting participants and whether reasons for non-participation are discussed
Data collection	This domain assesses the documentation and justification of the method of data collection (in-depth interviews, semi-structured interviews, focus groups or observations). It also assesses where interviews took place, what form the data took (e.g., tape recordings, written notes) and data saturation
Relationship between researcher and participants	This domain assesses who conducted any interviews, any potential biases they might have and how these might have influenced the research questions or data collection. The assessment should include consideration of how the researcher responded to events during the study

Ethical considerations	This domain assesses whether ethical approval was obtained and ethical standards maintained, including issues of informed consent, confidentiality and the effect of the study on participants
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 whether contradictory data are taken into account and whether the researcher considered their own biases during analysis and selection of data for presentation also forms part of this assessment
Findings	This domain assesses whether findings are credible, reported explicitly and discussed in the context of the original research question. It also assesses if findings for and against the researchers' arguments are discussed
Value of research	This domain assesses if the researchers discuss the generalisability of findings, the contribution they make to existing knowledge and directions for future research

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 importance in qualitative reviews

The committee discussed the context from which themes were derived, whether they were relevant to clinical practice in the UK and whether themes were sufficiently convincing to support or warrant a change in practice. The quality of the evidence was also considered. The committee discussion of the evidence sections explain if, and why, a recommendation was not made for a specific theme.

Evidence statements

Evidence statements are presented 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:

- the quality of the evidence
- the numbers of studies and participants for the outcome concerned or prognostic/risk factor (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

The aim of the health economic input to the guideline was to inform the committee of potential economic issues related to abortion and 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 and resource use for competing interventions. In addition, the health economic input aimed to identify areas of high resource impact. These are recommendations which might have a large impact on Clinical Commissioning Groups' or Trusts' finances and so requireattention.

Inclusion and exclusion of economic studies

The titles and abstracts of papers identified through the searches were independently assessed for inclusion using predefined eligibility criteria summarised in Table 10 .

Table 10: Inclusion and exclusion criteria for the systematic reviews of economic evaluations

Inclusion criteria
Economic evaluations that compare costs and health consequences of interventions (that is, true cost effectiveness analyses). Given the difficulty with using quality of life measures in this clinical area all outcome measures were considered
Population, interventions, comparators and outcomes which match those specified in the relevant review question

Inclusion criteria
Incremental results reported or enough information for incremental results to be derived
Conducted from a NHS and PSS perspective or from a OECD country with similar legislation for abortion
Exclusion criteria
Conference abstracts with insufficient methodological details for quality assessment
Non-English language papers

OECD Organisation for Economic Co-operation and Development; PSS: personal social services

Once the screening of titles and abstracts was complete, full versions of the selected papers were acquired for assessment. The quality of evidence was assessed using the economic evaluations checklist as specified in Developing NICE guidelines: the manual 2014.

Health economic modelling

As well as reviewing the published economic literature, as described above, new economic analysis was undertaken in selected areas prioritised by the committee in conjunction with the health economist. Topics were prioritised on the basis of the following criteria, in accordance with Developing NICE guidelines: the manual 2014:

- the overall importance of the recommendation, which may be a function of the number of people affected and the potential impact on costs and health outcomes per patient
- the current extent of uncertainty over cost effectiveness, and the likelihood that economic analysis will reduce this uncertainty
- the feasibility of building an economic model.

The full methods and results where bespoke economic analysis was prioritised are reported in appendix J of Evidence Report A: Accessibility and sustainability of abortion services, Evidence Report K: Medical versus surgical abortion between 13^{+0} and 24^{+0} weeks' gestation and Evidence Report P: Contraception after abortion. When new economic analysis was not prioritised, the committee made a qualitative judgement regarding cost effectiveness by considering expected differences in cost and resource 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 (given 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), or
- the intervention cost less than £20,000 per QALY gained compared with the next best strategy, or
- the intervention provided clinically significant benefits at an acceptable additional cost when compared with the next best strategy.

For some topics in the guideline it was not deemed appropriate to consider quality of life outcome measures in a quantitative way. Where this was the case the reasons

for not considering them are documented in the 'The committee's discussion of the evidence' sections and, where performed, in the report of the bespoke economic modelling (Appendix J in Evidence Report A: Accessibility and sustainability of abortion services, Evidence Report K: Medical versus surgical abortion between 13^{+0} and 24^{+0} weeks' gestation and Evidence Report P: Contraception after abortion.) In such cases a qualitative discussion of the evidence and issues relating to quality of life is presented.

The committee's considerations of cost effectiveness are discussed explicitly under the 'Cost effectiveness and resource use' headings of the relevant sections.

Developing recommendations

Guideline recommendations

Recommendations were drafted on the basis of the committee's interpretation of the available evidence, taking into account 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 the members' expert opinion. The considerations for making consensus-based recommendations include the balance between potential harms and benefits, the economic costs or implications compared with the economic benefits, current practices, recommendations made in other relevant guidelines, patient preferences and equality issues.

The main considerations specific to each recommendation are outlined under the 'The committee's discussion of the evidence' headings within each evidence report as well as the 'rationale and impact' section in the short guideline.

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

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 2014.

Validation process

This guideline is subject to a 6-week public consultation and feedback as part of the quality assurance and peer review of the guideline. All comments received from registered stakeholders were responded to in turn and posted on the NICE website at publication. For further details please refer to Developing NICE guidelines: the manual 2014.

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 2014.

Funding

The NGA was commissioned by NICE to develop this guideline.

References

Higgins 2011

Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from http://handbook.cochrane.org (accessed 13/02/19)

Lewin 2015

Lewin, S., Glenton, C., Munthe-Kaas, H., Carlsen, B., Colvin, C.J., Gülmezoglu, M., Noyes, J., Booth, A., Garside, R. and Rashidian, A., 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 Medicine, 12(10), p.e1001895.

NICE 2014

National Institute for Health and Care Excellence (NICE), Developing NICE guidelines: the manual (accessed 13/02/19)

NICE 2018

National Institute for Health and Care Excellence (NICE), Policy on declaring and managing interests for NICE advisory committees (accessed 13/02/19)

RMD 2018

Rehabilitation Measures Database, Rehabilitation Measures Database (accessed 13/02/19)

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.

Schünemann 2013

Schünemann, H., Brożek, J, Guyatt, G., Oxman, A., (eds.). Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach (updated October 2013). GRADE Working Group, 2013. Available from https://gdt.gradepro.org/app/handbook/handbook.html (accessed 13/02/19)

Wells 2008

Wells, G.A., Shea, B., O'Connel,l D., Peterson, J., Welch, V., Losos, M., Tugwel,l P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses, 2008. Available from http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (accessed 13/02/19)