National Institute for Health and Care Excellence

Final

Neonatal parenteral nutrition

Supplementary material C - Methods

NICE guideline NG154

Supplementary material C

February 2020

Final

These evidence reviews were developed by the National Guideline Alliance which is part of the Royal College of Obstetricians and Gynaecologists

Disclaimer

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

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

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 All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

ISBN: 978-1-4731-3673-1

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 for the management of parenteral nutrition in neonates.

What this guideline covers

Groups that are covered

- Babies born preterm, up to 28 days after their due birth date (preterm babies)
- Babies born at term, up to 28 days after their birth (term babies)

Specific consideration will be given to those babies who:

- Are critically ill **or**
- Need surgery

Clinical areas that are covered

1. Indications for, and approaches to, starting parenteral nutrition in preterm and term babies
2. Energy needs of preterm and term babies receiving parenteral nutrition
3. Individual constituents in parenteral nutrition for preterm and term babies:
    - macronutrients (amino acids, carbohydrates and lipids)
    - minerals and iron
    - chloride and acetate balance.
4. Venous access for parenteral nutrition in preterm and term babies
5. Monitoring parenteral nutrition in preterm and term babies
6. Stopping parenteral nutrition in preterm and term babies
7. Service design
8. Information and support for parents and carers

Methods

Introduction

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). A more up to date version is now available of the NICE manual; however development of this guideline was initiated before publication of the new manual and to ensure consistent methods the 2014 version was used throughout.

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 23 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). The review questions were based on the following frameworks:

- intervention reviews – using population, intervention, comparison and outcome (PICO)
- prognostic reviews – population, prognostic factor, outcome (PPO)
- qualitative reviews – using population, phenomenon of interest and context (PICo).
- formal consensus – using the nominal group consensus technique.

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

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

The review questions and corresponding review types for each evidence report are provided in Table 1.

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

Chapter/Evidence report	Subtopic	Review question	Type of review
[A1] Indications for, and approaches to, starting parenteral nutrition in preterm and term babies: predictors of feeding success	Predictors of feeding success	• What are the predictors for enteral feeding success?	Prognostic
[A2] Indications for, and approaches to, starting parenteral nutrition in preterm and term babies:	Optimal timeframe	• For those neonates where parenteral nutrition is required, what is the optimal timeframe for doing this?	Intervention

Chapter/Evidence report	Subtopic	Review question	Type of review
optimal timeframe to starting PN			
[B] Venous access	Venous access	• What overall osmolality, concentration of calcium, and glucose/dextrose in parenteral nutrition can determine whether to administer centrally or peripherally?	Intervention
[C] Energy needs of preterm and term babies receiving parenteral nutrition	Energy needs	• How many kcal/kg/day should be given to preterm and term babies receiving parenteral nutrition?	Intervention
[D1] Individual constituents in parenteral nutrition for preterm and term babies	Constituents of neonatal parenteral nutrition - glucose	• What is the optimal target dose for carbohydrates in preterm an term babies who are receiving parenteral nutrition? • What is the optimal way (starting dose and approach to increment, if employed) to achieve this target dose of carbohydrates?	Intervention
[D2] Individual constituents in parenteral nutrition for preterm and term babies	Constituents of neonatal parenteral nutrition – amino acids	• What is the optimal target dose for amino acids in preterm and term babies who are receiving parenteral nutrition? • What is the optimal way (starting dose and approach to increment, if employed) to achieve this target dose of amino acids?	Intervention
[D3] Individual constituents in parenteral nutrition for preterm and term babies	Constituents of neonatal parenteral nutrition - lipids	• What is the optimal target dose for lipids in preterm an term babies who are receiving parenteral nutrition? • What is the optimal way (starting dose and approach to increment, if employed) to achieve this target dose of lipid?	Intervention
[D4] Individual constituents in parenteral nutrition for preterm and term babies	Constituents of neonatal parenteral nutrition -lipid emulsions	• What is the clinical effectiveness, efficacy and safety of lipid formulations from different sources (for example soya, fish oil, or mixed sources)?	Intervention
[D5] Individual constituents in parenteral nutrition for preterm and term babies	Constituents of neonatal parenteral nutrition - Iron	• What is the effective and safe iron supplementation compared to no iron or different dosage/formulations in preterm and term babies who are receiving parenteral nutrition?	Intervention

Chapter/Evidence report	Subtopic	Review question	Type of review
[D6] Individual constituents in parenteral nutrition for preterm and term babies	Constituents of neonatal parenteral nutrition - Acetate	• How much (if any) intravenous acetate should be provided to preterm and term babies who are receiving parenteral nutrition?	Intervention
[D7] Individual constituents in parenteral nutrition for preterm and term babies	Constituents of neonatal parenteral nutrition – relative amounts of non-nitrogen energy to nitrogen	• What are the most effective relative amounts of nitrogen and non-nitrogen energy?	Intervention
[D8] Individual constituents in parenteral nutrition for preterm and term babies	Constituents of neonatal parenteral nutrition – relative amounts of carbohydrates and lipids	• What are the most effective relative amounts of carbohydrates and lipids?	Intervention
[D9] Individual constituents in parenteral nutrition for preterm and term babies	Constituents of neonatal parenteral nutrition - Calcium and phosphate	• What are the optimal target dosages for calcium and phosphate in preterm and term babies receiving parenteral nutrition?	Intervention
[D10] Individual constituents in parenteral nutrition for preterm and term babies	Relative amounts amino acids to phosphate	• What is the optimal ratio of phosphate to amino acid in preterm and term babies receiving parenteral nutrition?	Intervention
[E] Standardised neonatal parenteral nutrition	Standardised bags	• What is the effectiveness, efficacy and safety of standardised parenteral nutrition bags as compared to individualised bags?	Intervention
[F] Monitoring neonatal parenteral nutrition	Monitoring	• In babies on parenteral nutrition, what is the optimal frequency of blood sampling for monitoring glucose, calcium, phosphate, potassium and serum triglycerides?	Intervention
[G] Stopping neonatal parenteral nutrition	Stopping PN	• What amount of enteral nutrition (measured in terms of ml/kg/day or kcal/kg/day) should be given when starting enteral feeds?	Intervention
[H] Service design	Service design	• Are nutrition care teams (pharmacists and dietician) effective and safe in providing parenteral nutrition in preterm and term babies?	Intervention
[I] Information and support	Information and support	• What are the most effective methods of information provision about parenteral nutrition, and what information and support to parents/carers perceive as useful?	Mixed methods-intervention (quantitative) and qualitative

Chapter/Evidence report	Subtopic	Review question	Type of review
[J] General principles of neonatal parenteral nutrition	General principles	• What are the general principles of neonatal parenteral nutrition?	Formal nominal group consensus

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

- Supplementary material A (NGA staff list)
- Supplementary material B (Glossary and abbreviations)
- Supplementary material D (Economic study selection).

Searching for evidence

Clinical literature search

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

Databases were searched using medical subject headings, free-text terms and study type filters where appropriate. Where possible, searches were restricted to retrieve articles published in English. All searches were conducted in the following databases: Medline, Medline-in-Process, Cochrane Central Register of Controlled Trials (CCTR), Cochrane Database of Systematic Reviews (CDSR), Database of Abstracts of Reviews of Effects (DARE), Health Technology Assessments (HTA) and Embase. All searches, except the searches for 9.1 (the General principles chapter), were updated at least 6–8 weeks in advance of the final guideline committee meetings before consultation on the draft guideline; these updates were completed during April 2019. Any studies added to the databases after April 2019 (including those published before April 2019 but not yet indexed) were not considered for inclusion. The searches for 9.1 (the General Principles chapter), were not rerun as the evidence included for the formal consensus exercise were published guidelines, and the committee were not aware of any recently published guidelines which had been released following the undertaking of this review. .

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

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

Health economic literature search

Systematic literature searches were also undertaken to identify published health economic evidence. A broad search was conducted to identify health economic evidence related to parenteral nutrition in neonates in the following databases: NHS Economic Evaluation Database (NHS EED) and HTA. A broad search was also conducted to identify health economic evidence related to parenteral nutrition in neonates in the following databases with an economic search filter applied: Medline, CCTR and Embase. Where possible, the

searches were restricted to retrieve articles published in English; studies published in languages other than English were not eligible for inclusion.

The search strategies for the health economic literature search are included in Supplement D (Health economics). As for the clinical literature searches, economic literature searches were updated at least 6–8 weeks in advance of the final committee meeting before consultation on the draft guideline; these updates were completed during April 2019.

Reviewing evidence

Systematic review process

The evidence was reviewed in accordance with the following approach.

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

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

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

Type of studies and inclusion/exclusion criteria

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

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

For intervention reviews, randomised controlled trials (RCTs) were prioritised for inclusion because they are considered to be the most robust type of study design that could produce an unbiased estimate of intervention effects. Where there was limited evidence from RCT, non-randomised controlled trials and/or observational studies were considered for inclusion,

including cohort studies, case–control studies, cross-sectional studies and case series, and this was specified in each evidence review.

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

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

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

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. Conference abstracts were not considered for inclusion (see the review protocols for details).

Methods of combining evidence

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

Data synthesis for intervention reviews

Meta-analysis to pool results was conducted where possible using Cochrane Review Manager (RevMan5) software. The potential for bias and inconsistency across studies is higher in observational studies compared with RCTs. Therefore, evidence from RCTs and observational studies was analysed and presented to the committee separately. Differences and similarities between RCT and observational evidence were considered and discussed by the committee when making recommendations.

For dichotomous outcomes, such as mortality, the Mantel–Haenszel method with a fixed effect model was used to calculate risk ratios (RRs) with 95% confidence intervals (CIs). For all outcomes with zero events in both arms the risk difference (RD) was calculated and presented. For outcomes with zero events in only one arm, Peto odds ratios (PORs) was calculated as this method performs well when events are rare (Bradburn 2007). Both POR and RD were presented.

For continuous outcomes, measures of central tendency (mean) and variation (standard deviation; SD) are required for meta-analysis. Data for continuous outcomes, such as duration of hospital stay, were meta-analysed using an inverse-variance method for pooling weighted mean differences (WMDs). If a study reported only a summary statistic for a continuous outcome and a measure of variability (mean difference and 95% confidence interval (CI) or the mean difference and its SD or standard error) the generic-inverse variance method was used to enter data and conduct the meta-analysis.

Subgroups for stratified analyses were agreed a priori 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 reports). Forest plots were not presented for outcomes reported by single studies.

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

Data synthesis for prognostic reviews

Odds ratios or RRs with 95% CIs reported in published studies would be extracted or calculated by the NGA technical team to examine relationships between a given factor and the outcome of interest. Analyses adjusting for key confounders (such as gestational age) would have been included in preference to unadjusted analyses. Recognising variation across studies in terms of populations, risk factors, outcomes and statistical analysis methods (including adjustments for confounding factors), prognostic data would not usually be pooled, but results from individual studies would be presented in the evidence reports.

None of these methods were used because no evidence was identified for this topic.

Data synthesis for qualitative reviews

Where possible, a meta-synthesis was going to be conducted to combine evidence from qualitative studies. The main aim of qualitative data synthesis in this guideline was to determine what information parent or carers of preterm or term babies may receive, and how this may influence satisfaction with care, uptake of services and anxiety levels. The aim was also to determine the views of healthcare professionals regarding the provision of information to parents and carers. Whenever studies identified a qualitative theme, this would be extracted and the main characteristics would be summarised. When all themes would be extracted from studies, common concepts would be categorised and tabulated. This would include information on how many studies had contributed to each theme identified by the NGA technical team.

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

Themes from individual studies would be integrated into a wider context and overarching categories of themes with sub-themes may be identified, for example if there was a theme of ‘logistical support’ this may link into a wider theme of ‘satisfaction with overall care’. Themes would be derived from data presented in individual studies. When themes are extracted from one primary study only, theme names used in the guideline mirrored those in the source

study. However, when themes would be based on evidence from multiple studies, the theme names would be assigned by the NGA technical team. The names of overarching categories of themes would also be assigned by the NGA technical team.

Emerging themes would be placed into a thematic map. The purpose of such a map is to show relationships between overarching categories and associated themes.

None of these methods were used because no evidence was identified for this topic.

Appraising the quality of evidence

Intervention studies

GRADE methodology for intervention reviews

For intervention reviews, the evidence for outcomes from included RCTs and comparative observational studies was evaluated and presented using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology developed by the international GRADE working group. More information about this tool can be found on the developer's website.

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

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

Quality element	Description
Indirectness	This refers to differences in study populations, interventions, comparators or outcomes between the available evidence and inclusion criteria specified in the review protocol
Imprecision	This occurs when a study has relatively few participants or few events of interest, resulting in wide confidence intervals around estimates of effect that include clinically important thresholds
Publication bias	This refers to systematic under- or over-estimation of the underlying benefit or harm resulting from selective publication of study results

Table 3: GRADE quality ratings (by quality element)

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

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

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

Assessing risk of bias in intervention reviews

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

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

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

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

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 RCT the ROBIS checklist was used (see appendix H in Developing NICE guidelines: the manual; NICE 2014).

For observational studies the Cochrane risk of bias for non-randomised studies (ROBINS-I) was used (see appendix H in Developing NICE guidelines: the manual; NICE 2014).

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.

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 visually by inspecting forest plots and observing whether there was considerable heterogeneity in the results of the meta-analysis. Statistical heterogeneity was assessed by calculating the I-squared statistic for the meta-analysis with an I-squared value of more than 50% indicating considerable heterogeneity, and more than 80% indicating very serious heterogeneity. When considerable or very serious heterogeneity was observed, possible reasons were explored using sensitivity and subgroup analyses as pre-specified in the review protocol where possible. 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 and a random effects meta-analysis was 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 (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.

For outcomes with zero events in both arms, and where there were zero events in only one arm, the risk difference was calculated. However, there was no agreement on an equivalent to an MID for these cases. Due to the low event rate and usually associated wide CIs it was decided to downgrade these outcomes to 'serious' imprecision to prevent quality inflation.

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 committee was asked whether there were any recognised or acceptable MIDs in the clinical 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, RRs of 0.80 and 1.25 were considered the clinically important thresholds and 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 all outcomes with 0 events in both arms, or where there were 0 events in only one arm, the Peto odds ratio and risk difference was calculated. However, there was no agreement on the equivalent to an MID for these cases. Due to the low event rate which is usually associated with wide confidence intervals, it was decided to give those cases a 'serious' imprecision rating to prevent quality inflation for these outcomes.

For continuous outcomes, default MIDs are equal to half the mean SD of the control groups at baseline, or at follow-up if the SD is not available a baseline. If within one review question some outcomes have baseline SDs, yet other outcomes do not, then all MIDs will be calculated using follow-up SDs to ensure consistency.

Prognostic studies

Adapted GRADE methodology for prognostic reviews

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

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

RR, relative risk; OR, odds ratio

Assessing risk of bias in prognostic reviews

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

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

Assessing inconsistency in prognostic reviews

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

Assessing indirectness in prognostic reviews

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

Assessing imprecision and clinical importance in prognostic reviews

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

Qualitative reviews

GRADE-CERQual methodology for qualitative reviews

For qualitative reviews the GRADE Confidence in the Evidence from Reviews of Qualitative research (GRADE-CERQual) approach (Lewin 2015) was used. In this approach the quality of evidence is considered according to themes in the evidence. The themes may have been identified in the primary studies or they may have been identified by considering the reports of a number of studies. Quality elements assessed using GRADE-CERQual are listed and defined in Table 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: GRADE-CERQual 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

Quality element	Description
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: 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 CERQual (by review finding)

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

Assessing risk of bias in qualitative reviews

The risk of bias in qualitative studies was assessed using the Critical Appraisal Skills Programme (CASP) checklist for qualitative studies (see appendix H in Developing NICE guidelines: the manual; NICE 2014). The overall risk of bias was derived by assessing the risk of bias across the 6 domains summarised in Table 9.

Table 9: Risk of bias in qualitative studies

Aim and appropriateness of qualitative evidence	This domain assesses whether the aims and relevance of the study were described clearly and whether qualitative research methods were appropriate for investigating the research question
Rigour in study design or validity of theoretical approach	This domain assesses whether the study approach was documented clearly and whether it was based on a theoretical framework (such as ethnography or grounded theory). This does not necessarily mean that the framework has to be stated explicitly, but a detailed description ensuring transparency and reproducibility should be provided

Sample selection	This domain assesses the background, the procedure and reasons for the method of selecting participants. The assessment should include consideration of any relationship between the researcher and the participants, and how this might have influenced the findings
Data collection	This domain assesses the documentation of the method of data collection (in-depth interviews, semi-structured interviews, focus groups or observations). It also assesses who conducted any interviews, how long they lasted and where they took place
Data analysis	This domain assesses whether sufficient detail was documented for the analytical process and whether it was in accordance with the theoretical approach. For example, if a thematic analysis was used, the assessment would focus on the description of the approach used to generate themes. Consideration of data saturation would also form part of this assessment (it could be reported directly or it might be inferred from the citations documented that more themes could be found)
Results	This domain assesses any reasoning accompanying reporting of results (for example, whether a theoretical proposal or framework is provided)

Assessing relevance of evidence in qualitative reviews

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

Assessing coherence of findings in qualitative reviews

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

Assessing adequacy of data in qualitative reviews

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

Assessing clinical importance in qualitative reviews

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

Formal consensus reviews

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

Topics addressed with this method (evidence report J - general principles)

It was agreed that clinical evidence reviews would not be the most appropriate for topics where administration of care would be guided by physiological, pathophysiological and clinical principles; for example, vitamins, minerals, fluid volume and electrolytes. However, guidance on parenteral nutrition would not be complete without including general recommendations about these constituents. Therefore formal consensus was agreed as the most appropriate method to review this evidence. Agreed topics to be covered by formal consensus were:

- overall level of included vitamins
- general practice for fluid volume
- overall levels of blood and urinary electrolytes
- overall level of included minerals
- overall level of included trace elements
- delivery of lipids via syringe or bags
- filtration and protection from light.

Details of the nominal group technique as used in this guideline

A search was conducted for published guidelines on neonatal parenteral nutrition. Only international, national or regional guidelines/standard protocols were included, it was agreed that local guidelines would likely be based on these, lack wider applicability, and may not be generalisable; therefore, local guidelines were excluded. In order to identify most relevant literature, only those published in the last 10 years (January 2008 onwards) were included. All potentially relevant guidelines were identified and were assessed for quality using the Appraisal of Guidelines for Research and Evaluation (AGREE II) instrument (see assessing quality below).

Once the guidelines had been assessed and rated as high quality the NGA technical team extracted relevant recommendations from these guidelines and derived a set of statements for all included topics; guidelines rated as low quality were excluded (scores of below 70% in any two domains of the AGREE II instrument - see assessing quality of guidelines below).

All statements were checked for clinical content by the NGA clinical advisor and the committee chair. If no recommendations existed within the included guidelines for a particular review area then no statement was produced. The derived statements can be seen in the relevant evidence review (evidence review J – general principles).

The formal consensus exercise was conducted over two committee meetings. At the initial meeting the statements were presented to the committee in a questionnaire format (this questionnaire can be seen in the relevant evidence review (evidence review J – general principles, appendix M). All committee members were invited to take part in the formal consensus exercise (this did not include the chair as he had been involved in deriving the statements, nor co-opted members). Committee members were asked to rate each statement based on their personal opinion of what they believed 'best clinical practice' would be. The statements were rated using a 9-point Likert scale, where 1 represents 'strongly disagree', 5 represents 'neither agree nor disagree', and 9 represents 'strongly agree'. The participants were also able to state that they believed they had insufficient knowledge to provide a rating. There was also space for written comments about each statement. The questionnaire was completed anonymously with no prior discussion on the topic. Once this first round of consensus had been conducted, the NGA technical team calculated overall percentage agreement for each individual statement. The ratings were grouped into three categories: 1 to 3 (disagree), 4 to 6 (neither agree nor disagree), or 7 to 9 (agree). If a committee member indicated they had insufficient knowledge to provide a rating for a particular statement this was excluded from the calculation of agreement. Statements with 80% or greater agreement were kept, and were to be used to inform recommendations. Statements with less than 60% agreement were discarded, unless there were obvious and addressable issues identified from any comments. Those statements with between 60-80% agreement were re-drafted by the NGA technical team (using the written comments if provided).

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

At the following committee meeting, statements with 80% or greater agreement (from rounds 1 and 2) were presented as the evidence to inform the development of recommendations. For each topic (vitamins, fluid volume, electrolytes, minerals, trace elements, delivery of lipids, filtration and protection from light), the statements with >80% agreement were presented, the committee then discussed how these statements reflect practice, and how they could inform or drive best practice. The committee used these statements, along with their knowledge and experience to develop the recommendations for each topic. Statements below 80% were discarded in round 2 because they had already been redrafted after insufficient agreement in round 1 and further redrafting was not expected to lead to a higher level of agreement in a subsequent round of voting.

Assessing quality of guidelines

The identified potentially relevant guidelines were assessed for quality using the Appraisal of Guidelines for Research and Evaluation (AGREE II) instrument (Table 10). The tool assesses 6 domains: scope and purpose, stakeholder involvement, rigour of development, clarity of presentation, applicability and editorial independence.

Within each domain there is a set of questions, each of which is scored using a 7-point scale (1 – 'strongly disagree' to 7 – 'strongly agree'). Each section is rated and then an overall score for that domain is calculated. Two reviewers independently rated all identified guidelines using this method (see the AGREE II for detailed instructions). Only those guidelines which scored 70% or above in any two domains (and therefore rated as high quality) were included. Guidelines below this level were excluded.

Table 10: Assessing quality of guidelines

Domain	Aim of assessment
Scope and purpose	Assesses the aim of the guideline, the relevant health questions, and the population
Stakeholder involvement	Assesses the extent to which the guideline was developed by stakeholders, and whether it represents the views the guidelines intended users
Rigour of development	Assesses the processes used to collect, analyse and synthesise the evidence. It also assesses the methods used to generate and update the recommendations
Clarity of presentation	Assesses the language, and layout of the guideline
Applicability	Assesses potential barriers and facilitators of implementation, uptake of the guideline. It also assesses the potential resource implications.
Editorial independence	Assesses the possibility of the recommendations being biased and assesses potential conflict of interests

Collaboration with Cochrane (rapid update of a lipid formulation review – D4)

During the guideline scoping phase it was identified that Cochrane Neonatal were intending to conduct a rapid update of a review on different lipid formulations for parenteral nutrition. Cochrane Neonatal received funding from the National Institute of Health Research (NIHR) to update their published review to support development of the guideline. It was therefore agreed that the NGA technical team as well as the committee would work with Cochrane Neonatal to develop the review protocol and would have input into the structure of the evidence review to ensure that it would fit the purpose of the guideline. The Cochrane Neonatal team also agreed to conduct a second (new rather than updated) review to cover term babies. The NGA developed a standard protocol in accordance with NICE processes and then the Cochrane team used this to inform their full protocol, ensuring all critical and important outcomes as agreed by the committee were included. The draft Cochrane Neonatal protocols (developed using Cochrane processes) were then presented at a committee meeting to get committee input and sign off. The final review protocols were then peer reviewed in accordance with Cochrane processes.

Cochrane Neonatal agreed to conduct the review in accordance with the methods described above, including GRADE assessment of clinically relevant outcomes. Additionally, the NGA technical team reviewed the Cochrane reviews, including assessment of quality using the ROBIS tool, (ROBIS: Tool to assess risk of bias in systematic reviews), in line with those methods outlined in other sections of this document. The GRADE assessments completed by Cochrane Neonatal differed in several ways to those outlined in other sections of this document:

- Optimal information size was considered when judging imprecision:
  - Dichotomous outcomes were downgraded if the total number of events was less than 300.
  - Continuous outcomes were downgraded if the total number of participants was less than 400.
- GRADE default MIDs were not used when judging imprecision:
  - For dichotomous outcomes, RRs of 0.75 and 1.25 were considered the thresholds for appreciable benefit or harm, which is similar to the concept of MIDs, when judging imprecision
  - Continuous outcomes were downgraded for imprecision if the 95% confidence interval cross the null effect and were wide. If the confidence interval was wider than half a standardised mean difference, this was often considered wide. However, this was not

an absolute rule and the final decision was made based on consensus among the Cochrane authors, taking into account the clinical context.

- Some outcomes were further downgraded when only a single study contributed to the effect. This was dependent on sample size and consensus among the Cochrane authors.

Despite these minor differences in methods the NGA technical team was confident that it is unlikely any of the overall decisions would be significantly different. The draft Cochrane reviews were presented to the committee by Cochrane Neonatal and the committee and NGA technical team had an opportunity to review the draft and provide feedback. Additionally the NGA technical team summarised the Cochrane Neonatal reviews to develop a chapter for this guideline (Evidence Review - D4).

Evidence statements

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

- the quality of the evidence
- the numbers of studies and participants for the outcome concerned or prognostic/risk factor or prediction model (quantitative evidence) or that contributed to themes (qualitative evidence)
- where relevant, whether or not the estimate of effect is clinically important.
- where relevant, an indication of the direction of effect (for example, if a treatment is beneficial or harmful compared with another,
- a brief description of the participants
- the imprecision of the effect: for serious imprecision, the statement ‘there was uncertainty around the effect’ was included and for very serious imprecision, the statement ‘there was high uncertainty around the effect was used.

Economic evidence

The aim of the health economic input to the guideline was to inform the committee of potential economic issues related to preterm babies and term babies who require parenteral nutrition 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 of different care options. 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 need special attention.

Reviewing economic evidence

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

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

Inclusion criteria
Intervention or comparators in accordance with the guideline scope
Study population in accordance with the guideline scope

Inclusion criteria
Full economic evaluations (cost-utility, cost-effectiveness, cost-benefit or cost-consequence analyses) assessing both costs and outcomes associated with interventions of interest
Only studies from Organisation for Economic Co-operation and Development countries were included, because the aim of the review was to identify economic information transferable to the UK context
Exclusion criteria
Abstracts containing insufficient methodological details
Cost-of-illness type studies
Studies which adopted a very narrow perspective and included only intervention costs or reported only unit cost data

Once the screening of titles and abstracts was completed, full-text copies of potentially relevant articles were requested for detailed assessment. The economic evidence study selection is presented in supplementary material D. Existing economic evidence considered in the guideline is provided in the respective evidence reports, following presentation of the relevant clinical evidence. The references to included studies and the respective evidence tables with the study characteristics and results are provided in appendix H of the relevant evidence report, the economic evidence profiles are provided in the corresponding appendix I.

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 guideline committee prioritised the following review questions where it was thought that economic considerations would be particularly important in formulating recommendations.

- What are the predictors for enteral feeding success?
- For those neonates where parenteral nutrition is required, what is the optimal timeframe for doing this?
- What is the clinical effectiveness, efficacy and safety of lipid formulations from different sources (for example soya, fish oil, or mixed sources)?
- What is the effectiveness, efficacy and safety of standardised parenteral nutrition bags as compared to individualised bags?
- What overall osmolality, concentration of calcium, and glucose/dextrose in parenteral nutrition can determine whether to administer centrally or peripherally?
- In babies on parenteral nutrition, what is the optimal frequency of blood sampling for monitoring glucose, calcium, phosphate, potassium and serum triglycerides?
- What amount of enteral nutrition (measured in terms of ml/kg/day or kcal/kg/day) should be given when starting enteral feeds?
- Are nutrition care teams (pharmacists and dietician) effective and safe in providing parenteral nutrition in preterm and term babies?

Ultimately, original health economic modelling was undertaken only for the review question looking at the effectiveness, efficacy and safety of standardised parenteral nutrition bags as

compared to individualised bags. The clinical evidence was insufficient to allow de novo economic modelling in any of the other prioritised areas. Detail of the cost effectiveness analysis undertaken for the guideline is presented in the evidence chapter G with summary provided following the presentation of the relevant clinical evidence, with full methods and results provided in appendix J.

Cost effectiveness criteria

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

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

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

Developing recommendations

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

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

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

Research recommendations

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

Validation process

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

Updating the guideline

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

Funding

The NGA was commissioned by NICE to develop this guideline.

References

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.

Dixon-Woods 2005

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

Hayden 2013

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

Higgins 2011

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

Lewin 2015

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

Murphy 1998

Murphy, M,K,. Black, N,A,. Lamping, D,L,. McKee, C,M,. Standerson, C,F,B,. Askam, J,. Consensus development methods, and their use in clinical guideline development. Heath Technology Assessment, 2, 1998.

NICE 2014

National Institute for Health and Care Excellence (NICE) (2014) Developing NICE guidelines: the manual (updated 2017). Available from

https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview (accessed 15 June 2018)

NICE 2018

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

Santesso 2016

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