National Institute for Health and Care Excellence

Final

Specialist neonatal respiratory care for babies born preterm

Methods

NICE guideline NG124

Methods

April 2019

Final

This methods chapter was 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 All NICE guidance is subject to regular review and may be updated or withdrawn.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: 978-1-4731-3339-6

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 specialist neonatal respiratory care in babies born preterm.

Declarations of interest

Committee members' and developers' declarations of interest were recorded according to NICE's 2014 conflicts of interest policy until 31st March 2018, and thereafter in accordance with NICE's 2018 conflicts of interest policy.

What this guideline covers

Groups that are covered

- Babies born preterm who need respiratory support (for example oxygen supplementation or assisted ventilation) in hospital, beginning in the neonatal period.

Clinical areas that are covered

The guideline covers the following clinical issues:

- Early respiratory management (excluding resuscitation) after birth and before arrival in the neonatal unit. This includes oxygen supplementation and assisted ventilation with:
  - non-invasive techniques (for example high-flow therapy or continuous positive airway pressure [CPAP]) or
  - invasive techniques (for example conventional ventilation).
- Diagnosing bronchopulmonary dysplasia.
- Preventing and managing respiratory disorders on the neonatal unit, including with:
  - oxygen supplementation and assisted ventilation (including the techniques specified in key area 1 and high-frequency oscillatory ventilation)
  - medicines (for example, surfactants, corticosteroids, diuretics and caffeine)
  - treatment for patent ductus arteriosus
- Monitoring in the neonatal unit, including:
  - blood oxygen levels
  - blood carbon dioxide levels
  - blood pressure
- Sedation and analgesia (including morphine) in babies receiving respiratory support
- Involving and supporting parents and carers, communicating with them and providing them with information

- Discharge planning from hospital to home for babies who have had respiratory support in hospital (beginning in the neonatal period) and need continued support for chronic lung disease.

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

What this guideline does not cover

Groups that are not covered

The guideline does not cover the following groups:

- Babies born at term.
- Babies who need respiratory support because of congenital disorders, for example congenital diaphragmatic hernia.

Clinical areas that are not covered

This guideline does not cover the following areas:

- Resuscitating newborn babies (this is covered in the NICE-accredited Resuscitation Council UK guideline on the Resuscitation and support of transition of babies at birth)
- Technical aspects of airway management, such as intubation techniques
- Managing persistent pulmonary hypertension of the newborn
- Long-term management of chronic lung disease after discharge from the neonatal unit
- Neonatal feeding and nutrition
- Sepsis
- Neurological disorders
- Gastrointestinal disorders
- Congenital heart disease (apart from patent ductus arteriosus)
- Renal disorders
- Hypoglycaemia and hyperglycaemia
- Palliative care (this is covered in the NICE guideline on end of life care for infants, children and young people).

Methods

This chapter sets out in detail the methods used to review the evidence and to generate recommendations in the guideline. This guideline was developed using the methods described in Developing NICE guidelines: the manual 2014.

Developing the review questions and outcomes

The 21 review questions developed for this guideline were based on the key areas identified in the guideline scope. They were drafted by the NGA and refined and validated by the committee. They cover all areas of the scope and were signed-off by NICE (see Table 1).

The review questions were based on the following frameworks:

- intervention reviews: population, intervention, comparator and outcome (PICO)
- diagnostic test accuracy reviews: population, index test, reference standard and outcome (PIRO)
- prognostic reviews: population, presence or absence of a prognostic or predictive factor and outcome (PPO)
- qualitative reviews: sample, phenomenon of interest, design, evaluation, research type (SPIDER)

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 each review question.

Table 1: Description of review questions

Chapter or section	Type of review	Review question guideline[1]	Outcomes
A. Diagnosing respiratory disorders	Prognostic	2.1 What are the risk factors for bronchopulmonary dysplasia in preterm babies?	• Bronchopulmonary dysplasia
B. Respiratory support	Intervention	1.1 What respiratory support (excluding resuscitation) is the most effective for preterm babies before admission to the neonatal unit?	**Critical** • Mortality prior to discharge • Bronchopulmonary dysplasia • Neurodevelopmental outcomes at ≥18 months: ○ Cerebral palsy ○ Neurodevelopmental delay ○ Neurosensory impairment **Important** • Failed non-invasive ventilation • Pneumothorax/ pneumomediastinum

Chapter or section	Type of review	Review question guideline[1]	Outcomes
			• Severe intraventricular haemorrhage (grade 3 or 4)

B. Respiratory support	Intervention	3.3 What is the most effective way of using surfactant in managing respiratory distress syndrome?	**Critical**
			• Mortality prior to discharge
			• Bronchopulmonary dysplasia
			• Neurodevelopmental outcomes at ≥18 months:
			  ○ Cerebral palsy
			  ○ Neurodevelopmental delay
			  ○ Neurosensory impairment
			Important
			• Days on invasive ventilation
			• Severe intraventricular haemorrhage (grade 3 or 4)
			• Pneumothorax
			• Pulmonary haemorrhage

B. Respiratory support	Intervention	3.1 What is the most effective way to administer oxygen to preterm babies requiring respiratory support?	**Critical** • Bronchopulmonary dysplasia • Days of oxygen • Time spent within optimal target saturation limits **Important** • Retinopathy of prematurity • Nasal trauma • Comfort score/ pain score • Number of manual adjustments of titration

B. Respiratory support	Intervention	3.2 What is the effectiveness and safety of the different ventilation techniques in preterm babies needing respiratory support?	**Critical**
			• Mortality prior to discharge[2]
			• Bronchopulmonary dysplasia[2]
			• Neurodevelopmental outcomes at ≥18 months:
			  ○ Cerebral palsy
			  ○ Neurodevelopmental delay
			  ○ Neurosensory impairment
			Important
			• Number of days on invasive ventilation
			• Failed non-invasive ventilation
			• Pneumothorax
			• Parental satisfaction

B. Respiratory support	Intervention	3.7 What is the effectiveness of nitric oxide in preterm babies requiring invasive ventilation?	**Critical** • Mortality prior to discharge • Bronchopulmonary dysplasia

Chapter or section	Type of review	Review question guideline[1]	Outcomes

			• Neurodevelopmental outcomes at ≥18 months:
			  ○ Cerebral palsy
			  ○ Neurodevelopmental delay
			  ○ Neurosensory impairment
			Important
			• Days on ventilation
			• Severe intraventricular haemorrhage (grade 3 or 4)
			• Pulmonary haemorrhage
			• Methaemoglobinaemia

C. Managing respiratory disorders	Intervention	3.4 What is the effectiveness of corticosteroids in preterm babies requiring respiratory support?	**Critical**
			• Mortality prior to discharge
			• Bronchopulmonary dysplasia
			• Neurodevelopmental outcomes at ≥18 months:
			  ○ Cerebral palsy
			  ○ Neurodevelopmental delay
			  ○ Neurosensory impairment
			Important
			• Days on invasive ventilation
			• Gastro-intestinal perforation
			• Hypertension

C. Managing respiratory disorders	Intervention	3.5 What is the effectiveness of diuretics in preterm babies requiring respiratory support?	**Critical**
			• Mortality prior to discharge
			• Bronchopulmonary dysplasia
			• Neurodevelopmental outcomes at ≥18 months:
			  ○ Cerebral palsy
			  ○ Neurodevelopmental delay
			  ○ Neurosensory impairment
			Important
			• Days on invasive ventilation
			• Nephrocalcinosis
			• Ototoxicity
			• Hyponatraemia

C. Managing respiratory disorders	Intervention	3.6 What is the effectiveness of caffeine in preterm babies requiring respiratory support?	**Critical**
			• Mortality prior to discharge
			• Bronchopulmonary dysplasia
			• Neurodevelopmental outcomes at ≥18 months:
			  ○ Cerebral palsy
			  ○ Neurodevelopmental delay
			  ○ Neurosensory impairment

Chapter or section	Type of review	Review question guideline[1]	Outcomes
			Important • Continuing apnoea • Extubation failure • Tachycardia • Necrotising enterocolitis

C. Managing respiratory disorders	Intervention	3.8 What is the effectiveness of interventions for closing a patent ductus arteriosus in preterm babies requiring respiratory support?	**Critical**
			• Mortality prior to discharge
			• Bronchopulmonary dysplasia
			• Neurodevelopmental outcomes at ≥18 months:
			  ○ Cerebral palsy
			  ○ Neurodevelopmental delay
			  ○ Neurosensory impairment
			
			Important:
			• Failure of patent ductus arteriosus closure
			• Renal impairment
			• Gastrointestinal complications:
			  ○ Gastrointestinal perforation
			  ○ Gastrointestinal haemorrhage
			  ○ Necrotising enterocolitis

D. Monitoring	Intervention	4.1 What oxygen levels are optimal in the management of preterm babies requiring respiratory support?	**Critical**
			• Severe retinopathy of prematurity
			• Mortality prior to discharge
			• Neurodevelopmental outcomes at ≥18 months:
			  ○ Cerebral palsy
			  ○ Neurodevelopmental delay
			  ○ Neurosensory impairment
			
			Important
			• Bronchopulmonary dysplasia
			• Necrotising enterocolitis
			• Patent ductus arteriosus requiring medical or surgical treatment

D. Monitoring	Diagnostic accuracy	4.2 What is the best method for measuring oxygen levels in diagnosing hyperoxia or hypoxia in preterm babies?	**Critical** • Sensitivity • Specificity • Area Under the Receiver Operator Characteristic (ROC) Curve (AUC) • Positive likelihood ratio (LR+) • Negative likelihood ratio (LR-)

Chapter or section	Type of review	Review question guideline[1]	Outcomes
			Important • Adverse events • Infection • Burns • Ischaemic limbs • Emboli/thrombi • Blood loss due to excess sampling

D. Monitoring	Intervention	4.3 What carbon dioxide levels are optimal in the management of preterm babies requiring respiratory support?	**Critical**
			• Mortality prior to discharge
			• Bronchopulmonary dysplasia
			• Neurodevelopmental outcomes at ≥18 months:
			  ○ Cerebral palsy
			  ○ Neurodevelopmental delay
			  ○ Neurosensory impairment
			
			Important
			• Periventricular leukomalacia
			• Severe intraventricular haemorrhage
			• Days on invasive ventilation
			• Pneumothorax

D. Monitoring	Intervention	4.4 What blood pressure monitoring strategies are associated with improved outcomes in babies requiring respiratory support?	**Critical**
			• Mortality prior to discharge
			• Neurodevelopmental outcomes at ≥18 months:
			  ○ Cerebral palsy
			  ○ Neurodevelopmental delay
			  ○ Neurosensory impairment
			• Severe intraventricular haemorrhage (grade 3 or 4)
			
			Important
			• Periventricular leukomalacia
			• Necrotising enterocolitis
			• Renal impairment
			• Vascular complications associated with invasive monitoring

E. Sedation and analgesia	Intervention	5.1 What is the effectiveness of morphine during respiratory support?	**Critical** • Mortality prior to discharge • Severe intraventricular haemorrhage (grade 3 or 4) • Pain and comfort scores **Important**

Chapter or section	Type of review	Review question guideline[1]	Outcomes
			• Unplanned or accidental extubation • Days to achieve full enteral feeding • Hypotension which requires intervention • Parental satisfaction
E. Sedation and analgesia	Intervention	5.2 What is the effectiveness of using pre-medication for elective intubation in preterm babies?	**Critical** • Ease of intubation (e.g. number of intubation attempts, time to successful intubation, failed intubation) • Pain and comfort scores during intubation • Adverse physiological response during intubation **Important** • Neurodevelopmental outcomes at ≥18 months: ○ Cerebral palsy ○ Neurodevelopmental delay ○ Neurosensory impairment • Days on ventilation • Adverse drug reactions
F. Involving and supporting parents and carers	Intervention	6.1 What parent and carer involvement is effective in the care of preterm babies who are receiving respiratory support?	**Critical** • Days in hospital during initial admission • Bronchopulmonary dysplasia • Neurodevelopmental outcomes at ≥18 months: ○ Cerebral palsy ○ Neurodevelopmental delay ○ Neurosensory impairment **Important** • Number of episodes of confirmed or suspected sepsis during initial hospitalisation • Mortality prior to discharge • Infant growth defined as changes in z scores at 3, 6, 12 and 24 months of age: ○ Weight ○ Height ○ Head circumference • Parental/ carer satisfaction using validated scales

Chapter or section	Type of review	Review question guideline[1]	Outcomes
F. Involving and supporting parents and carers	Qualitative	6.2 What support is valued by parents and carers of preterm babies who are receiving respiratory support in the neonatal unit?	**Themes** • Psychological and social support: ○ Counselling ○ Crisis intervention ○ Emotional support ○ Stress management ○ Vulnerable families, safeguarding ○ Support groups • Support from staff: ○ Parental participation in decision-making, including participation in ward rounds ○ Parental presence and participation in care-giving • Hospital design and supportive spaces: ○ Facilities to support family presence in the neonatal unit e.g. comfortable reclining chairs ○ Accommodation, food ○ Parking and public transport links ○ Design of physical space that take into account infants', families', and staff members' needs • Financial support ○ Transportation to and from hospital, parking ○ Child care
F. Involving and supporting parents and carers	Qualitative	6.3 What information, and in what format, is valued by parents and carers of preterm babies who are receiving respiratory support in the neonatal unit?	**Themes** • Formats ○ In person ○ Print ○ Online ○ Internet resources ○ Technology • Qualities ○ Availability of different languages ○ Equality of access e.g. vision impairment ○ Timing of access ○ Frequency of accessibility • Types of information ○ Clinical Information

Chapter or section	Type of review	Review question guideline[1]	Outcomes
			○ Parent/carer-infant bonding information ○ Coping information
G. Discharge planning	Qualitative	7.1 What factors are important when planning for the safe transition from the neonatal unit of a baby born preterm, requiring respiratory support?	**Themes** • Access to the MDT, including medical, specialist nursing and therapy teams, and psychological support • Community team involvement • Training or qualifications of the care provider that will be providing care post neonatal unit discharge • Named discharge co-ordinator or key worker, such as named consultant or nurse • Training and completion of competencies of parents • Medication administration • Support to facilitate the confidence of parents • Equipment provision • Care package funded • Suitable discharge destination environment ○ Housing ○ Electricity • Follow Up Care including discharge summaries
G. Discharge planning	Qualitative	7.2 What are the support and information needs of parents and carers of preterm babies who are transitioning from the neonatal unit while receiving ongoing respiratory support?	• Access to the MDT including medical, specialist nursing and therapy teams and community team, continuity of carers, and psychological support for parents/carers and others who share the same household, including siblings • Community team involvement in discharge planning process • Involvement in decision making and care planning for their child. • Rooming-in; timing in relation to discharge, experience of stay • Experience of training and the support available • Experience of different types of training methods and resources available

Chapter or section	Type of review	Review question guideline[1]	Outcomes

			• Interventions which enabled/would have enabled parents or carers to feel confident in caring for their baby
			• Equipment provision
			• Care package funded
			• Including experiences of accessing funding package
			• Parent and carer feelings about suitability of the package their child has been awarded
			• Suitable discharge destination environment
			  ○ Housing
			  ○ Electricity
			• Follow-up care in place before discharge with timely provision of relevant documentation such as discharge summaries, paediatric passport, etc.
			• The format in which information is received
			• Equality considerations
			  ○ Accessibility of training and training information
			  ○ Advocacy services
			  ○ Conflict – process and resolution
			  ○ Flexibility of training

1. *Questions are listed in the order they appear in the final guideline which reflects the care pathway*
2. *These outcomes were included in the network meta-analysis performed within this review*

MDT – Multi-disciplinary team; NICU – Neonatal Intensive Care Unit

Searching for evidence

Clinical literature search

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. All searches were conducted in MEDLINE, Embase and The Cochrane Library, with some additional database searching in AMED, PsycINFO, CINAHL, MIDIRS, Health and Psychosocial Instruments and Web of Science Social Science Citation Index for certain topic areas (for example PsycINFO for question 3.3).

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 key or additional studies of which they were aware. The questions, the study types applied, the databases searched, the clinical search strategies and the years covered can be found in appendix B in each evidence review chapter.

A generalised exclusion filter was used in every review to exclude very low level study types, such as editorials, animal studies, etc. Elements from the protocol that were excluded, such as populations, interventions or comparisons, were not listed as specific exclusion criteria due to the risk that the search strategy would then exclude relevant studies that happened to mention excluded items, but otherwise would have been relevant.

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.

Health economics literature search

A global search of economic evidence was undertaken on 13 December 2016. The following databases were searched:

- MEDLINE (Ovid)
- EMBASE (Ovid)
- Health Technology Assessment database (HTA)
- National Health Service Economic Evaluations 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 specialist neonatal respiratory care for babies born preterm that had not been already identified via the database searches.

Individual searches for questions were also undertaken alongside the clinical searches. Databases were searched using relevant medical subject headings, free-text terms and, for searches undertaken in MEDLINE and EMBASE, a search filter to capture economic evaluations. Studies published in languages other than English were not reviewed. No restrictions on setting were applied to any of the searches, but a standard exclusions filter was applied (letters, animals, etc.). Full details of the search strategies are presented in appendix B in each evidence review chapter.

Re-runs of literature searches

The committee reviewed the list of questions at guideline committee meeting 7 (January 2018) and identified 4 questions where re-runs of literature searches were not required as adequate evidence had been identified in the initial review to inform the recommendations and the committee agreed that it was unlikely that additional

evidence would be found that would change the recommendations, or data saturation had been achieved in the initial qualitative review. These 4 questions were:

- 2.1 What are the risk factors for bronchopulmonary dysplasia in preterm babies?
- 6.2 What support is valued by parents and carers of preterm babies who are receiving respiratory support in the neonatal unit?
- 6.3 What information, and in what format, is valued by parents and carers of preterm babies who are receiving respiratory support in the neonatal unit?
- 7.2 What are the support and information needs of parents and carers of preterm babies who are transitioning from the neonatal unit while receiving ongoing respiratory support?

Re-run searches of the clinical and health economics literature were done for the remaining 17 review questions. Re-runs were carried out in 2 stages with initial re-runs starting 1st May. For review questions where the initial re-run was conducted during May, a second re-run was carried out between the 18th June and 2nd July. In this way final re-runs were carried out for all the 17 questions not more than 8 weeks before the final committee meeting, in accordance with Developing NICE guidelines: the manual 2014.

Call for evidence

No call for evidence was made.

Reviewing clinical evidence

Systematic review process

The evidence was reviewed following these steps.

- Potentially relevant studies were identified for each review question from the relevant search results by reviewing titles and abstracts. Full papers were then obtained.
- Full papers were reviewed against pre-specified inclusion and exclusion criteria in the review protocols (in appendix A of each evidence review chapter).
- Key information was extracted on the study's methods, according to the factors specified in the protocols and results. These were presented in summary tables (in each review chapter) and evidence tables (in appendix D of each evidence review chapter).
- Relevant studies were critically appraised using the appropriate checklist as specified in Developing NICE guidelines: the manual 2014.
- Summaries of evidence were generated by outcome (included in the relevant review chapters) and were presented in committee meetings.
- Results were summarised and reported in GRADE profiles (for intervention reviews) or their equivalent (for diagnostic test accuracy and qualitative reviews)

All drafts of reviews were checked by a senior reviewer.

Type of studies and inclusion/exclusion criteria

Systematic reviews (SRs) with meta-analyses were considered the highest quality evidence to be selected for inclusion.

For intervention reviews, randomised controlled trials (RCTs) were included because they are considered the most robust study design for unbiased estimation of intervention effects. Based on their judgement, if the committee believed RCT data were not appropriate or there was limited evidence from RCTs, they agreed to include cohort studies with a comparative group. Only RCTs were included in the network meta-analyses performed. Only RCTs with greater than 15 participants in each arm were included, as smaller RCTs would be unlikely to produce a meaningful effect.

For diagnostic test accuracy reviews cross-sectional or cohort studies of diagnostic test accuracy were considered for inclusion.

For prognostic reviews, systematic reviews/meta-analyses of cohort studies were prioritised for inclusion. No such reviews were identified so in the absence of such studies, prospective population-based cohort studies and prospective multicentre cohort studies with sample sizes of greater than 100 participants were considered for inclusion. This minimum sample size was chosen in order overcome the issue of multiple comparisons in the selection of prognostic factors and comparison of prognostic models.

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

The committee was consulted about any uncertainty regarding inclusion or exclusion of studies. Excluded studies by review question with the reasons for their exclusion are listed in appendix K in each evidence review.

Posters, letters, editorials, comment articles, unpublished studies and studies not in the English language were excluded. Narrative reviews were also excluded, but individual references were checked for inclusion. Conference abstracts and studies performed in non-OECD countries were not included, as the committee agreed that the standards of neonatal care in non-OECD countries were likely to be very different than those in more developed nations. Studies where less than 2/3 of the population were preterm were also not included.

For quality assurance of study identification, a 10% random sample of the literature search results was sifted by a second reviewer for the following review questions:

3.2 What is the effectiveness and safety of the different assisted ventilation techniques in preterm babies?

This question was selected because it was a review question for which a network meta-analysis was carried out by the NGA, in addition to a pair-wise comparison, so it was important to ensure a high quality of study identification as it would be very difficult to add studies in later if omitted in error by the first reviewer.

5.1 What is the effectiveness of morphine during respiratory support?

This question was selected because it was the first review carried out by a reviewer who was new to the guideline.

Possible discrepancies were resolved by discussion between the two reviewers and with a third (senior) reviewer if necessary.

The inclusion and exclusion of studies was based on the review protocols, which can be found in appendix A of each evidence review chapter. Excluded studies and the reasons for their exclusion are listed in appendix K of each evidence review. In addition, the committee was consulted to resolve any uncertainty about inclusion or exclusion.

Methods of combining evidence

Data synthesis for intervention reviews

Pairwise meta-analysis

Pairwise meta-analysis of randomised trials reporting the same outcomes of interest was done using Review Manager 5 (RevMan 5) software. For binary outcomes, such as occurrence of adverse events, the Mantel-Haenszel method of statistical analysis was used to calculate risk ratios (relative risks, RRs) with 95% confidence intervals (CIs).

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 initial hospital admission stay) were analysed using an inverse-variance method for pooling weighted mean differences.

Subgroups for stratified analyses were decided for some review questions a priori at the protocol stage if the committee identified biological or clinical characteristics which would affect the effectiveness of the intervention.

Statistical heterogeneity was assessed by visually examining the forest plots, and by considering the chi-squared test for significance with heterogeneity defined as a $p<0.1$ or an I-squared inconsistency statistic value of 50% or more. Where heterogeneity was present, subgroup analyses were performed for any confounders defined in the review protocol. If the heterogeneity remained, a random effects (DerSimonian 2015) model was used to provide a more conservative estimate of the effect.

Results from multiple observational studies of the same comparison were not pooled but presented as a range of effects due to the high risk of selection bias in observational studies whereby differences in participant characteristics between treatment arms leads to a biased estimate of treatment effect. Forest plots were generated to present the results of meta-analyses and stratified for subgroup analyses (please see appendix E of each intervention evidence review).

Network meta-analysis

In the review looking at the effectiveness and safety of the different assisted ventilation techniques in preterm babies requiring respiratory support, bronchopulmonary dysplasia (BPD) at 36 weeks postmenstrual age (PMA) and mortality prior to discharge outcomes were synthesised using network meta-analytic techniques with the NMA review protocol presented in the relevant chapter (C), appendix N.

As is the case for ordinary pairwise meta-analysis, network meta-analysis (NMA) may be conducted using either fixed or random effect models. A fixed effect model typically assumes that there is no variation in relative effects across trials for a particular pairwise comparison and any observed differences are solely due to

chance. For a random effects model, it is assumed that the relative effects are different in each trial but that they are from a single common distribution. The variance reflecting heterogeneity is often assumed to be constant across trials.

In a Bayesian analysis, for each parameter the evidence distribution is weighted by a distribution of prior beliefs. The Markov chain Monte Carlo (MCMC) algorithm was used to generate a sequence of samples from a joint posterior distribution of 2 or more random variables and is particularly well adapted to sampling the treatment effects (known as a posterior distribution) of a Bayesian network. A prior distribution was used to maximise the weighting given to the data and to generate the posterior distribution of the results.

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

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

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

We adapted standard fixed and random effects models available from NICE Decision Support Unit (DSU) technical support document number 2: http://nicedsu.org.uk/wp-content/uploads/2017/05/TSD2-General-meta-analysis-corrected-2Sep2016v2.pdf

For further description of the model used, specific methods, outcomes and the results of the NMA please see the evidence review for question 3.2, chapter B.

The quality assurance of all the NMA work was undertaken by the NICE Guidelines Technical Support Unit, University of Bristol (TSU).

Data synthesis for diagnostic test accuracy reviews

Meta-analysis of diagnostic test accuracy was not performed because there were no reviews with multiple studies reporting the same test that could be combined. Results were presented individually for each study.

Sensitivity, specificity, positive and negative likelihood ratios and area under the receiver operator characteristic (ROC) curve (AUC) with 95% CIs were used as outcomes for diagnostic test accuracy. These diagnostic accuracy parameters were

obtained from the studies or calculated by the technical team using data from the studies.

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

The following cut-offs were used when summarising the levels of sensitivity or specificity for the committee:

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

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

The following thresholds (see for example Jaeschke 2002) were used when summarising the likelihood ratios for the committee:

- very useful test: LR+ higher than 10.0, LR- lower than 0.1
- moderately useful test: LR+ 5.0 to 10.0, LR- 0.1 to 0.2
- not a useful test: LR+ lower than 5.0, LR- higher than 0.2.

AUC shows true positive rate (sensitivity) as a function of false positive rate (1 minus specificity). The following cut-offs for AUC were used when determining the discriminative value of a test:

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

Data synthesis for prognostic reviews

Identification of risk factors for bronchial pulmonary dysplasia could aid early identification and management strategies. Adjusted odds ratios (ORs) or RRs with 95% CIs reported by the studies were extracted to study the relationship between a given factor and the outcome of interest. Studies reporting multivariable analyses adjusted for key confounders as specified in the review protocol (such as gestational age) were considered. Because of variation across the studies in terms of population, the risk factor, outcome and statistical methods (including adjustments for confounding factors), the prognostic data were not pooled but results from individual studies were reported.

Data synthesis for qualitative reviews

The most relevant evidence for this guideline originated from studies set in the target context of an OECD country setting, and ideally a UK NHS setting. The main aim of the synthesis of the qualitative data was to produce a description of the topics that may influence the experience of parents and carers of babies born preterm who are receiving or have received specialist neonatal respiratory care and healthcare professionals involved in their care, rather than build new theories or reconceptualise the topic under review.

Themes were derived from data presented in individual studies based on the study authors' interpretations in combination with direct quotes from interviewees. When themes were extracted, theme names were derived from the studies that provided them. Overarching themes (when synthesising the information from all studies as a whole), were named by the systematic reviewers.

Whenever studies identified a qualitative theme related to the review's phenomenon of interest, this data was extracted and the main characteristics were summarised. When all the relevant themes were extracted, meta-syntheses were performed where appropriate to identify an overarching framework of themes and their subthemes. Information on the number studies that contributed to each theme and subtheme was also tabulated.

The synthesis of the themes was drafted by a member of the technical team, but the final framework was further developed and, when necessary, re-classified through discussion with at least one other member of the technical team. The committee could then draw conclusions from each theme in each setting or country and how they may help in forming recommendations.

Themes with their accompanying subheadings were then organised into a thematic map that visually depicted the interconnecting relationships.

Appraising the quality of evidence

Intervention reviews

Pairwise meta-analysis

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

For intervention reviews, the evidence for outcomes from the included studies was evaluated and presented using GRADE, which was developed by the international GRADE working group.

The software developed by the GRADE working group (GRADEpro) was used to assess the quality of each outcome, taking into account individual study quality factors and the meta-analysis results. The clinical evidence profile tables include details of the quality assessment and pooled outcome data, where appropriate, 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 summary measures of effect and measures of dispersion (such as mean and SD or median and range) 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 binary outcomes. Reporting or publication bias was taken into consideration in the quality assessment and reported in the clinical evidence profile tables if it was apparent on visual inspection of funnel plots.

The selection of outcomes for each review question was decided when each review protocol was discussed with the committee, and was informed by committee discussion and by key papers.

The evidence for each outcome in the intervention reviews was examined separately for the quality elements listed and defined in Table 2. Each element was graded using the quality levels listed in Table 3.

The main criteria considered in the rating of these elements are discussed below. Footnotes were used in the GRADE profiles to describe reasons for grading a quality element as having serious or very serious limitations. The ratings for each component were combined to obtain an overall assessment for each outcome (Table 4).

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

Quality element	Description
Risk of bias	Limitations in the study design and implementation may bias the estimates of the treatment effect. High risk of bias for the majority of the evidence decreases confidence in the estimate of the effect.
Inconsistency	Inconsistency refers to an unexplained heterogeneity of results or findings.
Indirectness	Indirectness refers to differences in study population, intervention, comparator and outcomes between the available evidence and the review question, such that the effect estimate is likely to be changed. This is also related to applicability or generalisability of findings.
Imprecision	Results are imprecise when studies include relatively few patients and / or few events and thus have wide confidence intervals around the estimate of the effect. Imprecision results if the confidence interval includes one or more of the clinically important thresholds (minimally important difference – see below).
Publication bias	Publication bias is a systematic underestimate or an overestimate of the underlying beneficial or harmful effect due to selective publication of studies.

Table 3: Levels of quality elements in GRADE

Levels of quality elements in GRADE	Description
None/ no serious	There are no serious issues with the evidence.
Serious	The issues are serious enough to downgrade the outcome evidence by 1 level.
Very serious	The issues are serious enough to downgrade the outcome evidence by 2 levels.

Table 4: Levels of overall quality of outcome evidence in GRADE

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.

It should be noted that 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 this poor design will impact on the estimation of the intervention effect.

For systematic reviews of RCTs the AMSTAR checklist was used to assess risk of bias and for systematic reviews of other study types the Cochrane ROBIS checklist was used. For RCTs the Cochrane risk of bias tool for RCTs was used and for observational studies the Cochrane risk of bias tool for non-randomised studies (ROBINS-I) was used (see Appendix H in Developing NICE guidelines: the manual 2014).

Assessing inconsistency in intervention reviews

Inconsistency refers to unexplained heterogeneity of results of meta-analysis. When estimates of the 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 applicable when statistical meta-analysis is conducted (that is, results from different studies are pooled). For outcomes derived from a single study 'no serious inconsistency' was used when assessing this domain, as per GRADE methodology (Santesso 2016).

Statistical heterogeneity was assessed by visually examining the forest plots, and by considering the chi-squared test for significance at $p<0.1$ and the I-squared inconsistency statistic (with an I-squared value of 50 to 80% indicating potentially serious inconsistency and I-squared value of over 80% indicating very serious inconsistency). When no plausible explanation for the heterogeneity could be found, the quality of the evidence was downgraded in GRADE by 1 or 2 levels for the domain of inconsistency, depending on the extent of heterogeneity in the results and the random effects model was applied to the data.

Assessing indirectness in intervention reviews

Directness refers to the extent to which the populations, intervention, comparisons and outcome measures are similar to those defined in the inclusion criteria for the

reviews. Evidence is downgraded for indirectness by 1 or 2 levels if there are serious or very serious issues with the directness of the evidence. Indirectness is important when these differences are expected to contribute to a difference in effect size, or may affect the balance of harms and benefits considered for an intervention.

Assessing imprecision and clinical significance in intervention reviews

Imprecision in guidelines concerns whether the uncertainty (CI) around the effect estimate means that it is not clear whether there is a clinically important difference between interventions or not (that is, whether the evidence would clearly support one recommendation or appear to be consistent with several different types of recommendations). Therefore, imprecision differs from the other aspects of evidence quality because it is not really concerned with whether the point estimate is accurate or correct (has internal or external validity). Instead, it is concerned with what the uncertainty around the point estimate actually is. This uncertainty is reflected in the width of the CI.

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

Imprecision in the evidence reviews is assessed by considering whether the width of the 95% CI of the effect estimate is relevant to decision-making, taking each outcome in isolation. This assessment also involves effect size thresholds for clinical importance (the minimally important difference, MID) for benefit and for harm.

If the effect estimate CI includes clinically important benefit (or harm) there is uncertainty over which decision to make (based on this outcome alone). The CI is consistent with 2 possible decisions and so this is considered to be imprecise in the GRADE analysis and the evidence is downgraded by 1 level (‘serious imprecision’).

An effect CI including clinically important benefit, clinically important harm and no effect is consistent with 3 possible decisions. This is considered to be very imprecise in the GRADE analysis and the evidence is downgraded by 2 levels (‘very serious imprecision’).

Minimally important differences

The literature was searched for established MIDs for the selected outcomes in the evidence reviews. In addition, the committee was asked whether they were aware of any acceptable MIDs in the clinical community.

If no published or acceptable MIDs were identified, the committee considered whether it was clinically acceptable to use the GRADE default MIDs to assess imprecision. For binary outcomes, the GRADE default MIDs of RRs of 0.75 and 1.25 were adapted and RRs of 0.8 and 1.25 were used instead (due to the statistical distribution of the RR the adapted values are symmetrical on a log [RR] scale). For continuous outcomes, GRADE default MIDs are half of the median SD of the control group. As no published MID values were identified, the committee agreed that GRADE default MID values were to be used as a starting point for all outcomes and any exception to their application based on the committee’s consideration of clinical acceptability were noted and explained in the evidence review. On the rare occasions when outcomes were reported medians, in the absence of both published and GRADE default MIDs the evidence was downgraded by 1 level (‘serious

imprecision'). Outcomes reported as ORs were transformed to RRs to assess imprecision.

Network meta-analysis

For the NMAs, quality was assessed by looking at risk of bias across the included evidence using the Cochrane Risk of Bias Tool for Randomized Controlled Trials, as well as heterogeneity and consistency (also called incoherence). In addition, a threshold analysis was undertaken to assess the robustness of any conclusions based on the NMAs of the invasive ventilation techniques to potential biases in the included evidence. Quality was assessed for each NMA outcome including BPD at 36 weeks PMA and mortality prior to discharge.

The following limits of the upper 95% credible interval (CrI) for between-study standard deviation were used to assess heterogeneity for NMAs in which a random effects model was used:

- less than 0.3 – low heterogeneity
- 0.3 to 0.6 – moderate heterogeneity
- more than 0.6 to 0.9 – high heterogeneity
- more than 0.9 to 1.2 – very high heterogeneity.

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

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

For fixed-effect NMAs that did not model heterogeneity, or for networks in which inconsistency could not be assessed as no closed treatment loops existed, these criteria were not considered to impact the quality of evidence.

Threshold analysis was undertaken by the TSU to assess the robustness of the conclusions from the invasive ventilation NMAs. If studies included in a NMA are assessed to have flaws in their conduct or reporting, the reliability of results from the NMA can be in doubt. Therefore, analysts and decision makers need to assess the

robustness of any conclusions based on the NMA to potential biases in the included evidence. Suppose that we ask, “how much would the evidence have to change before the recommendation changes?” This is the motivation behind threshold analysis. The results of a threshold analysis describe how much each data point could change (or be adjusted for bias) before the recommendation changes and what the revised recommendation would be. Threshold analysis may be carried out at two levels: (i) at a study level, assessing the influence of individual study estimates on the recommendation and (ii) at a contrast level, where the influence of the combined evidence on each treatment contrast is considered. Full Methods and findings are summarised in the chapter B, appendix T.

Modified GRADE methodology for diagnostic test accuracy reviews

The GRADE approach was modified to assess the quality of evidence about diagnostic test accuracy by adapting the principles of GRADE for intervention reviews as described below. Four domains were considered: risk of bias, indirectness, inconsistency and imprecision. Each domain was rated as ‘no serious’, ‘serious’ or ‘very serious’. These domains were then combined to give the overall certainty in the body of evidence, rated as ‘very low’, ‘low’, ‘moderate’ or ‘high’.

Assessing risk of bias in diagnostic test accuracy reviews

Risk of bias in diagnostic test accuracy studies was assessed using the risk of bias items from the QUADAS-2 checklist (see appendix H in Developing NICE guidelines: the manual 2014). An overall risk of bias judgement was for each study was reached by considering the QUADAS-2 bias domains together. The risk of bias for the body of diagnostic test accuracy evidence was based on the risk of bias from the individual studies but with consideration of how much each study contributed to the overall evidence base.

Assessing indirectness in diagnostic test accuracy reviews

Indirectness was assessed using the applicability items from the QUADAS-2 checklist.. The indirectness for the body of diagnostic test accuracy evidence was based on the indirectness of the individual studies but with consideration of how much each study contributed to the overall evidence base.

Assessing inconsistency in diagnostic test accuracy reviews

Where there were multiple studies, the body of evidence was downgraded for serious inconsistency if there was unexplained variability between studies, when viewed on a forest plot or ROC curve. If there was only one study then inconsistency was rated as ‘no serious inconsistency.

Assessing imprecision in diagnostic test accuracy reviews

Imprecision was judged by comparing the CI of the estimate of sensitivity or specificity to clinical decision thresholds agreed beforehand by the committee. The committee decided whether sensitivity or specificity was the most important for decision making and agreed two threshold values. First a threshold for high sensitivity/specificity (above which the test would be definitely recommended) and second a threshold for low sensitivity/specificity (below which the test would not be recommended). If the CI of the estimate of sensitivity or specificity included one of these thresholds then the evidence was downgraded for serious imprecision, because it was consistent with two possible decisions. If the CI included both these

thresholds then the evidence was downgraded for very serious imprecision because it was consistent with three possible decisions.

Prognostic reviews

GRADE methodology for prognostic reviews

The GRADE approach was not used to assess the quality of evidence prognostic test for prognostic reviews. Quality assessment of outcomes was based upon risk of bias assessment for outcomes from individual studies.

Assessing risk of bias in prognostic reviews

Risk of bias in individual prognostic studies was assessed using the risk of bias items from the QUIPS checklist (see appendix H in Developing NICE guidelines: the manual 2014). An overall risk of bias judgement for each study was reached by considering the QUIPS bias domains together.

Qualitative reviews

GRADE CERQual methodology for qualitative reviews

The GRADE-CERQual (Confidence in the Evidence from Reviews of Qualitative research; Lewin 2015) approach was used to summarise the confidence in qualitative evidence. The overall confidence in evidence about each theme or sub-theme was rated on four dimensions: methodological limitations, applicability, coherence and adequacy of data.

Methodological limitations refer to the extent to which there were problems in the design or conduct of the studies and was assessed with the Critical Appraisal Skills Programme (CASP) checklist for qualitative studies. Applicability of evidence was assessed by determining the extent to which the body of evidence from the primary studies were applicable to the context of the review question. Coherence of findings was assessed by examining the clarity of the data. Adequacy of data was assessed by looking at the degree of richness and quantity of findings.

After assessing each of the components separately, a judgment of the overall confidence in each of the review findings was made.

Confidence could be assessed as high, moderate, low or very low:

- High confidence: it is highly likely that the review finding is a reasonable representation of the phenomenon of interest
- Moderate confidence: it is likely that the review finding is a reasonable representation of the phenomenon of interest
- Low confidence: it is possible that the review finding is a reasonable representation of the phenomenon of interest
- Very low confidence: it is not clear whether the review finding is a reasonable representation of the phenomenon of interest

Assessing risk of bias in qualitative reviews

For qualitative studies, quality was assessed using a checklist for qualitative studies (as suggested in appendix H in Developing NICE guidelines: the manual 2014). This

was based on the Critical Appraisal Skills Programme (CASP) checklist for qualitative studies.

Evidence statements

Evidence statements are summary statements presented in each evidence review highlighting the key features of the clinical evidence presented. The evidence statements are presented by outcome or theme and encompass the following key features of the evidence:

- the quality of the evidence
- the number of studies and the number of participants for a particular outcome or a particular risk factor or theme
- a brief description of the participants
- the clinical significance of the effect and an indication of its direction (for example, if a treatment is clinically significant (beneficial or harmful) compared with another, or whether there is no clinically significant difference between the tested treatments) or a summary of the effect size of the prognostic factor or accuracy of a diagnostic test, or findings within a theme for a qualitative review. If there is uncertainty about the estimate this wording is included in the evidence statement to show this.

Economic evidence

The aim of the health economic input to the guideline was to inform the committee of potential economic issues related to specialist neonatal respiratory care of babies born preterm 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

The titles and abstracts of papers identified through the searches were independently assessed for inclusion using predefined eligibility criteria summarised in Table 5.

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

Inclusion criteria
Only studies from Organisation for Economic Co-operation and Development countries were included, as the aim of the review was to identify economic information transferable to the UK context.
Selection criteria based on types of clinical conditions and population as well as interventions assessed were identical to the clinical review.
Study population according to the scope.
Only studies published from 2007 onwards were included in the review. This date restriction was imposed so that retrieved economic evidence was relevant to current healthcare settings and costs.

Inclusion criteria
Studies were included provided that sufficient details regarding methods and results were available to enable the methodological quality of the study to be assessed, and provided that the study's data and results were extractable. Conference abstracts, poster presentations or dissertation abstracts were excluded.
Full economic evaluations (cost utility, cost effectiveness, cost benefit or cost consequence analyses) that assess both the costs and outcomes associated with the interventions of interest. Cost studies were also considered for the inclusion.

Once the screening of titles and abstracts was complete, full versions of the selected papers were acquired for assessment. The applicability and quality of evidence was assessed using the economic evaluations checklist as specified in Developing NICE guidelines: the manual 2014. The completed checklists of existing economic evidence are provided in appendix M. The economic evidence study selection for each question is presented in appendix G. Existing economic evidence considered in the guideline is provided in the respective evidence chapters, 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.

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

- Question 3.2 What is the effectiveness and safety of the different ventilation techniques in preterm babies needing respiratory support?
- Question 3.7 What is the effectiveness of nitric oxide in preterm babies requiring invasive ventilation?
- Question 4.2 What is the best method for measuring oxygen levels in diagnosing hyperoxia or hypoxia in preterm babies?
- Question 6.1 What parent and carer involvement is effective in the care of preterm babies who are receiving respiratory support?

For the effectiveness and safety of the different assisted ventilation techniques, only costings of non-invasive ventilation techniques were undertaken. For the effectiveness of nitric oxide in preterm babies requiring invasive respiratory support there was directly applicable UK-based economic evidence and as a result, de novo modelling was not required by the committee. Also, the clinical evidence was insufficient to allow de novo economic modelling to assess the cost-effectiveness of

methods for measuring oxygen levels. For question 6.1 an economic analysis was undertaken to assess the cost-utility of interventions with a focus on involving parents, carers and family members in the care of babies who are receiving respiratory support.

The full methods and results of de novo economic analyses are reported in appendix J of each evidence review that was modelled. When new economic analysis was not prioritised, the committee made a qualitative judgement regarding cost effectiveness by considering existing economic evidence, expected differences in resource and cost use between options, alongside clinical effectiveness evidence identified from the clinical evidence review.

Characteristics and results of all economic studies considered during the guideline development process (including modelling studies conducted for this guideline) are summarised in economic evidence profiles in appendix I.

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.

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 chapter 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 agreed that there was no accepted best practice, or they could not reach a consensus on recommendations based on their knowledge and clinical experience, then the committee considered making recommendations for future research. For further details please refer to Developing NICE guidelines: the manual 2014.

Validation process

This guidance is subject to a 6-week public consultation and feedback as part of the quality assurance and peer review of the document. All comments received from registered stakeholders are 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, or other factors such as clinical practice have changed, which means the recommendations are out of date. For further details please refer to Developing NICE guidelines: the manual 2014.

Funding

The NGA was commissioned by NICE to develop this guideline.

References

DerSimonian2015

DerSimonian, R., Laird, N. Meta-analysis in clinical trials revisited. Contemporary clinical trials. 2015; 45:139-45.

Jaeschke 2002

Jaeschke R, Guyatt G, Lijmer J. Diagnostic tests. In: Guyatt G, Rennie D, eds. Users' guides to the medical literature. Chicago: AMA Press, 121-140, 2002

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., 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. 2015

NICE 2014

National Institute for Health and Care Excellence (NICE), Developing NICE guidelines: the manual (https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview)

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 gdt.guidelinedevelopment.org/app/handbook/handbook.html (accessed 26 November 2017)