National Institute for Health and Care Excellence

Consultation

Perioperative care in adults

NICE guideline: methods

NICE guideline

Methods

November 2019

Draft for Consultation

Developed by the National Guideline Centre, hosted by the Royal College of Physicians

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, where appropriate, their carer or guardian.

Local commissioners and 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 All NICE guidance is subject to regular review and may be updated or withdrawn.

Contents

1 Development of the guideline

1.1 What is a NICE guideline?

NICE guidelines are recommendations for the care of individuals in specific clinical conditions or circumstances within the NHS – from prevention and self-care through primary and secondary care to more specialised services. These may also include elements of social care or public health measures. We base our guidelines on the best available research evidence, with the aim of improving the quality of healthcare. We use predetermined and systematic methods to identify and evaluate the evidence relating to specific review questions.

NICE guidelines can:

- provide recommendations for the treatment and care of people by health professionals
- be used to develop standards to assess the clinical practice of individual health professionals
- be used in the education and training of health professionals
- help patients to make informed decisions
- improve communication between patient and health professional.

While guidelines assist the practice of healthcare professionals, they do not replace their knowledge and skills.

We produce our guidelines using the following steps:

- A guideline topic is referred to NICE from NHS England.
- Stakeholders register an interest in the guideline and are consulted throughout the development process.
- The scope is prepared by the National Guideline Centre (NGC).
- The NGC establishes a guideline committee.
- A draft guideline is produced after the group assesses the available evidence and makes recommendations.
- There is a consultation on the draft guideline.
- The final guideline is produced.

The guideline is made up of a collection of documents including this Methods report and a number of evidence reports covering each of the review questions included in the guideline. These can all be downloaded from NICE at www.nice.org.uk.

NICE also publishes a summary of the recommendation in this guideline, known as ‘the NICE guideline’.

NICE Pathways brings together all connected NICE guidance.

1.2 Remit

The Department of Health and Social Care England asked NICE to develop a clinical guideline on perioperative care. NICE commissioned the NGC to produce the guideline. This guideline will also be used to develop the NICE quality standard for perioperative care.

1.3 Who developed this guideline?

A multidisciplinary guideline committee comprising health professionals and researchers as well as lay members developed this guideline (see the list of guideline committee members and the acknowledgements).

The National Institute for Health and Care Excellence (NICE) funds the National Guideline Centre (NGC) and thus supported the development of this guideline. The committee was convened by the NGC and chaired by Paul Wallman in accordance with guidance from NICE.

The group met approximately every 6 weeks during the development of the guideline. At the start of the guideline development process all committee members declared interests including consultancies, fee-paid work, shareholdings, fellowships and support from the healthcare industry. At all subsequent committee meetings, members declared arising conflicts of interest.

Members were either required to withdraw completely or for part of the discussion if their declared interest made it appropriate. The details of declared interests and the actions taken are shown in the declaration of interest register for this guideline published on the NICE website.

Staff from the NGC provided methodological support and guidance for the development process. The team working on the guideline included a project manager, systematic reviewers (research fellows), health economists and information specialists. They undertook systematic searches of the literature, appraised the evidence, conducted meta-analysis and cost-effectiveness analysis where appropriate and drafted the guideline in collaboration with the committee.

1.3.1 What this guideline covers

The guideline will cover adults (18 and older) undergoing surgery in secondary and tertiary healthcare (whether in a conventional hospital setting or elsewhere) and in general dental practices. Specific consideration will be given to the needs of older people. It updates areas from the Surgical site infections: prevention and treatment NICE guidance CG74 and CardioQ-ODM oesophageal doppler monitor NICE guidance MTG3. The areas included in this guideline are:

- Information and support needs of adults undergoing surgery, and their families and carers.
- Preoperative assessment.
- Preoperative optimisation.
- Intraoperative management.
- Postoperative management and recovery.

For further details please refer to the scope for this guideline (published on the NICE website) and the review questions in section 2.1.

1.3.2 What this guideline does not cover

This guideline does not cover:

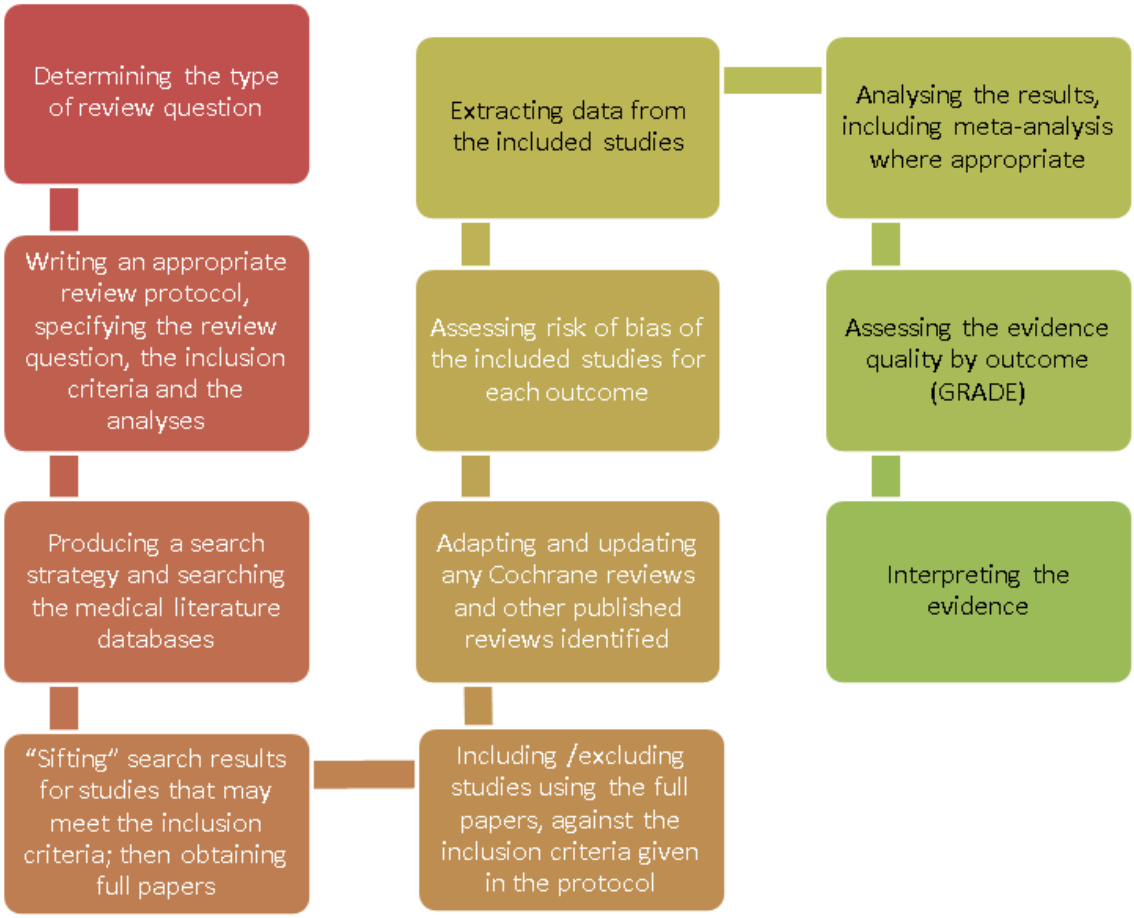
- Minor surgery in outpatient clinics.
- Routine preoperative tests for elective surgery.
- Blood products and blood transfusion.

- Preventing infection (apart from effectiveness of perioperative perfusion and hydration, and strict blood glucose control, on the prevention of surgical site infection).
- Surgery for burns.
- Surgery for traumatic brain injury and neurosurgery.
- Aspect of perioperative care that apply only to specific types of surgery (for example, caesarean section).
- Lifestyle modification prior to surgery, as this guideline will cross refer to other appropriate guidelines.

2 Methods

This report sets out in detail the methods used to review the evidence and to develop the recommendations that are presented in each of the evidence reviews for this guideline. This guidance was developed in accordance with the methods outlined in the NICE guidelines manual, 2014 version.

Figure 1: Step-by-step process of review of evidence in the guideline



2.1 Developing the review questions and outcomes

Review questions were developed using a PICO framework (population, intervention, comparison and outcome) for intervention reviews; using a framework of population, index tests, reference standard and target condition for reviews of diagnostic test accuracy; using population, presence or absence of factors under investigation (for example prognostic factors) and outcomes for prognostic reviews; and using a framework of population, setting and context for qualitative reviews.

This use of a framework guided the literature searching process, critical appraisal and synthesis of evidence, and facilitated the development of recommendations by the guideline committee. The review questions were drafted by the NGC technical team and refined and validated by the committee.

A total of 22 review questions were identified.

Full literature searches, critical appraisals and evidence reviews were completed for all the specified review questions.

Table 1: Review questions

Evidence report	Type of review	Review questions	Outcomes
1	Qualitative	What information and support is useful for adults undergoing surgery, and their families and carers, before, during and after an operation?	Themes will be derived from the evidence identified for this review and not pre-specified. However for information to guide the technical team, relevant themes may include: • Decision making (including involvement in discharge planning) • Preferred format of information provision (plain English, web-based for older people) • Content of information (including ongoing care needs) • Impact of treatment on lifestyle • Information sources other than healthcare professionals (e.g. support groups, online resources) • Psychological support • Financial support • Employment rights
2	Risk prediction	Which validated preoperative risk stratification tools best identify increased risk of mortality and morbidity in adults who will be undergoing surgery?	Outcomes: • Sensitivity, specificity, predictive values; • Area under the ROC curve (c-statistic); • Predicted risk versus observed risk (calibration) of • Mortality • Morbidity
3	Intervention	In older people (>60 years) who will be undergoing surgery, what is the clinical and cost effectiveness of pre-operative optimisation clinics?	Critical outcomes: • health-related quality of life • mortality • patient, family and carer experience of care • adverse events and complications (Clavien-Dindo, postoperative morbidity score (POMS)) • length of hospital stay (total pre and postoperative) Important outcomes: • unplanned intensive care unit admission • length of stay in intensive care unit

Evidence report	Type of review	Review questions	Outcomes
			• hospital readmission
4	Intervention	What is the most clinically and cost-effective type of intraoperative intravenous fluid for adults undergoing surgery?	Critical outcomes: • health-related quality of life • mortality • adverse events and complications (Clavien-Dindo; postoperative morbidity score (POMS); acute kidney injury; coagulopathy; nausea and vomiting; pulmonary complications, surgical site infections) Important outcomes: • length of hospital stay • unplanned ICU admission • ICU length of stay (planned and unplanned)
5	Intervention	Does nutritional screening in preoperative assessment improve surgical outcome for adults?	Critical outcomes: • health-related quality of life • mortality • patient, family and carer experience of care • adverse events and complications (Clavien-Dindo, postoperative morbidity score (POMS), respiratory complications, infection and sepsis, postoperative cardiac complications) Important outcomes: • length of hospital stay • unplanned ICU admission • ICU length of stay (planned and unplanned)
6	Intervention	What is the most clinically and cost effective preoperative fasting strategy for adults?	Critical outcomes: • health-related quality of life • mortality • patient, family and carer experience of care • adverse events and complications (Clavien-Dindo, postoperative morbidity score (POMS), aspiration – pulmonary complications, acute kidney injury) Important outcomes: • unplanned ICU admission • thirst • headache

Evidence report	Type of review	Review questions	Outcomes
			• cancellation of surgery
7	Intervention	What is the most clinically and cost effective oral iron supplementation strategy for the preoperative management of iron deficiency anaemia?	Critical outcomes: • mortality • health-related quality of life • preoperative Hb level • transfusion (pre-, intra- and post-surgery) • postoperative morbidity score (POMS) • change in healthcare management (for example, delayed surgery or surgery cancellation) Important outcomes: • length of hospital stay • unplanned ICU admission • ICU length of stay (planned and unplanned) • adherence • adverse events from iron tablets (e.g. constipation, nausea)
8	Intervention	What is the most clinically and cost effective management strategy for the preoperative management of iron deficiency anaemia?	Critical outcomes: • mortality • health-related quality of life • preoperative Hb level • blood transfusion (pre-, intra- and post-surgery) • postoperative morbidity score (POMS) • change in healthcare management (for example, delayed surgery or surgery cancellation) Important outcomes: • length of hospital stay • unplanned ICU admission • ICU length of stay (planned and unplanned) • adverse events from iron infusion(e.g. constipation, nausea) • adverse events from transfusion (e.g. infections, reactions (compatibility), hypersensitivity)
9	Intervention	What is the most clinically and cost effective strategy for perioperative management of managing anticoagulant medication in patients taking warfarin with target INR >3?	Critical outcomes: • health-related quality of life • mortality • bleeding • thromboembolism • stroke

Evidence report	Type of review	Review questions	Outcomes
			Important outcomes: • length of hospital stay (pre and postoperative)
10	Intervention	What is the clinical and cost effectiveness of blood glucose control management in adults undergoing surgery?	Critical outcomes: • health-related quality of life • mortality • adverse events and complications (Clavien-Dindo, postoperative morbidity score (POMS), cardiovascular, respiratory and neurological complications) • infections (including surgical site) • hypoglycaemia Important outcomes: • length of hospital stay • unplanned ICU admission • ICU length of stay (planned and unplanned) • hospital readmission
11	Intervention	What is the clinical and cost effectiveness of non-invasive cardiac output monitoring during major, complex or high risk surgery in adults?	Critical outcomes: • health-related quality of life • mortality • perioperative complications (such as surgical site infections) Important outcomes: • length of hospital stay • length of stay in intensive care unit • hospital readmission
12	Intervention	What is the clinical and cost effectiveness of management systems to promote safety in operating theatres?	Critical outcomes: • health-related quality of life • mortality • patient, family and carer experience of care • adverse events and complications (such as Clavien-Dindo, postoperative morbidity score (POMS)) • never events • serious incidents • compliance Important outcomes: • length of hospital stay • hospital readmission • unplanned ICU admission • ICU length of stay (planned and unplanned)

Evidence report	Type of review	Review questions	Outcomes
13	Intervention	What is the clinical and cost effectiveness of IV paracetamol compared to oral paracetamol given post operatively in managing acute post-operative pain?	Critical outcomes: • health-related quality of life • pain reduction ○ ≤ 6 hours post op ○ >6 to 24 hours post op • amount of additional medication use ○ ≤ 6 hours post op ○ >6 to 24 hours post op • adverse events (including respiratory depression, nausea, vomiting) Important outcomes: • psychological distress and mental well-being • symptom scores • functional measures • length of stay in intensive care • length of stay in hospital • hospital readmission
14	Intervention	What is the clinical and cost effectiveness of adding IV paracetamol to IV opioids given intraoperatively in managing acute post-operative pain?	Critical outcomes: • health-related quality of life • pain reduction ○ ≤ 6 hours post op ○ >6 to 24 hours post op • amount of additional medication use ○ ≤ 6 hours post op ○ >6 to 24 hours post op • adverse events (including respiratory depression, nausea, vomiting) Important outcomes: • psychological distress and mental well-being • symptom scores • functional measures • length of stay in intensive care • length of stay in hospital • hospital readmission
15	Intervention	What is the clinical and cost effectiveness of NSAIDs for managing acute postoperative pain?	Critical outcomes: • health-related quality of life • pain reduction ○ ≤ 6 hours post op ○ >6 to 24 hours post op • amount of additional medication use ○ ≤ 6 hours post op

Evidence report	Type of review	Review questions	Outcomes
			○ >6 to 24 hours post op • adverse events (including respiratory depression, nausea, vomiting, cardiac events, acute kidney injury, gastrointestinal complications, bone healing complications) Important outcomes: • psychological distress and mental well-being • symptom scores • functional measures (including time to mobilisation) • length of stay in intensive care • length of stay or hospital • hospital readmission
16	Intervention	What is the clinical and cost effectiveness of IV opioid compared to oral opioid given post operatively in managing acute post-operative pain?	Critical outcomes: • health-related quality of life • pain reduction ○ ≤ 6 hours post op ○ >6 to 24 hours post op • amount of additional medication use ○ ≤ 6 hours post op ○ >6 to 24 hours post op • adverse events (including respiratory depression, nausea, vomiting) Important outcomes: • psychological distress and mental well-being • symptom scores • functional measures • length of stay in intensive care • length of stay in hospital • hospital readmission
17	Intervention	What is the most clinically and cost effective opioid administration strategy?	Critical outcomes: • health-related quality of life • pain reduction ○ ≤ 6 hours post op ○ >6 to 24 hours post op • amount of additional medication use ○ ≤ 6 hours post op ○ >6 to 24 hours post op • adverse events (including respiratory depression, nausea, vomiting)

Evidence report	Type of review	Review questions	Outcomes
			Important outcomes: • psychological distress and mental well-being • symptom scores • functional measures • length of stay in intensive care • length of stay in hospital • hospital readmission
18	Intervention	What is the clinical and cost effectiveness of adding IV ketamine to IV opioids in managing acute post-operative pain?	Critical outcomes: • health-related quality of life • pain reduction ○ ≤ 6 hours post op ○ >6 to 24 hours post op • amount of additional medication use ○ ≤ 6 hours post op ○ >6 to 24 hours post op • adverse events (including respiratory depression, nausea, vomiting, delirium) Important outcomes: • psychological distress and mental well-being • symptom scores • functional measures • length of stay in intensive care • length of stay in hospital • hospital readmission
19	Intervention	What is the clinical and cost effectiveness of neuropathic nerve stabilisers in managing acute post-operative pain?	Critical outcomes: • health-related quality of life • pain reduction ○ ≤ 6 hours post op ○ >6 to 24 hours post op • amount of additional medication use ○ ≤ 6 hours post op ○ >6 to 24 hours post op • adverse events (including respiratory depression, nausea, vomiting, sedation, postural hypotension, antimuscarinic/anticholinergic side effects) Important outcomes: • psychological distress and mental well-being • symptom scores • functional measures • length of stay in intensive care

Evidence report	Type of review	Review questions	Outcomes
			• length of stay in hospital • hospital readmission
20	Intervention	What is the clinical and cost effectiveness of enhanced recovery programmes for adults having major surgery?	Critical outcomes: • health-related quality of life • mortality • patient, family and carer experience of care • adverse events and complications (Clavien-Dindo, postoperative morbidity score (POMS)) • patient and staff adherence Important outcomes: • length of hospital stay • unplanned intensive unit admission • length of stay in intensive care unit • hospital readmission • psychological distress and mental wellbeing (hospital anxiety and depression scale (HADS)) • pain
21	Intervention	What is the clinical and cost effectiveness of postoperative recovery in specialist areas, including intensive care, for adults?	Critical outcomes: • health-related quality of life • mortality • adverse events and complications (Clavien-Dindo, postoperative morbidity score (POMS)) • unplanned ICU admission/readmission Important outcomes: • length of hospital stay • hospital readmission • postponed/cancelled surgery • patient, family and carer experience of care

2.2 Searching for evidence

2.2.1 Clinical and health economics literature searches

The full search strategy including population terms, intervention terms, study types applied, the databases searched and the years covered can be found in Appendix B of the evidence review report.

Systematic literature searches were undertaken to identify all published clinical and health economics evidence relevant to the review questions. Searches were undertaken according

to the parameters stipulated within the NICE guidelines manual 2014[3] (https://www.nice.org.uk/media/default/about/what-we-do/our-programmes/developing-nice-guidelines-the-manual.pdf). Databases were searched using relevant medical subject headings, free-text terms and study-type filters applied where appropriate. Studies published in languages other than English were not reviewed, and where possible, searches were restricted to English Language. All searches were updated on 30th May 2019. If new evidence falls outside of the timeframe for the guideline searches e.g. from stakeholder comments, the impact on the guideline will be considered, and any further action agreed between the developer and NICE staff with a quality assurance role.

Prior to running, searches were quality assured using different approaches. Checking key papers were retrieved and Medline search strategies were peer reviewed by a second information specialist using the PRESS checklist[2]. Additional studies were added by checking reference lists of relevant systematic reviews, and those highlighted by committee members.

During the scoping stage, a search was conducted for guidelines and reports on the websites listed below. Web sites searched include:

- Guidelines International Network database (www.g-i-n.net)
- National Guideline Clearing House (www.guideline.gov)
- National Institute for Health and Care Excellence (NICE) (www.nice.org.uk)
- TRIP (www.tripdatabase.com)
- Scottish Intercollegiate **Guidelines** Network (SIGN) (www.sign.ac.uk)
- Evidence Search (www.evidence.nhs.uk).

Searching for unpublished literature was not undertaken. The NGC and NICE do not have access to drug manufacturers' unpublished clinical trial results, so the clinical evidence considered by the committee for pharmaceutical interventions may be different from that considered by the MHRA and European Medicines Agency for the purposes of licensing and safety regulation.

2.3 Identifying and analysing evidence of effectiveness

Research fellows conducted the tasks listed below, which are described in further detail in the rest of this section:

- Identified potentially relevant studies for each review question from the relevant search results by reviewing titles and abstracts. Full papers were then obtained.
- Reviewed full papers against prespecified inclusion and exclusion criteria to identify studies that addressed the review question in the appropriate population, and reported on outcomes of interest (review protocols are included in an appendix to each of the evidence reports).
- Critically appraised relevant studies using the appropriate study design checklist as specified in the NICE guidelines manual. Qualitative studies and systematic reviews were critically appraised using the GRADE CERQual approach for rating confidence in the body of evidence as a whole and using the respective CASP checklists for the methodological limitations section of the quality assessment.
- Extracted key information about interventional study methods and results using 'Evibase', NGC's purpose-built software. Evibase produces summary evidence tables, including critical appraisal ratings. Key information about non-interventional study methods and results was manually extracted onto standard evidence tables and critically appraised separately (evidence tables are included in an appendix to each of the evidence reports).
- Generated summaries of the evidence by outcome. Outcome data were combined, analysed and reported according to study design:

  - Randomised data were meta-analysed where appropriate and reported in GRADE profile tables.
  - Data from non-randomised studies were presented as a range of values in GRADE profile tables or meta-analysed if appropriate.
  - Qualitative data were synthesised across studies and presented as summary statements with accompanying GRADE CERQual ratings for each review finding.
- A sample of a minimum of 10% of the abstract lists of the first 3 sifts by new reviewers and those for complex review questions (for example, prognostic reviews) were double-sifted by a senior research fellow and any discrepancies were rectified. All of the evidence reviews were quality assured by a senior research fellow. This included checking:
  - papers were included or excluded appropriately
  - a sample of the data extractions
  - correct methods were used to synthesise data
  - a sample of the risk of bias assessments.

2.3.1 Inclusion and exclusion criteria

The inclusion and exclusion of studies was based on the criteria defined in the review protocols, which can be found in an appendix to each of the evidence reports. Excluded studies (with the reasons for their exclusion) are listed in another appendix to each of the evidence reports. The committee was consulted about any uncertainty regarding inclusion or exclusion.

Conference abstracts were not automatically excluded from any review. The abstracts were initially assessed against the inclusion criteria for the review question and further processed when a full publication was not available for that review question. If the abstracts were included the authors were contacted for further information. No relevant conference abstracts were identified for this guideline. Literature reviews, posters, letters, editorials, comment articles, unpublished studies and studies not in English were excluded.

2.3.1.1 Saturation of qualitative studies

Data extraction in qualitative reviews is a thorough process and may require more time compared to intervention reviews. It is common practice to stop extracting data once saturation has been reached. This is the point when no new information emerges from studies that match the review protocol. The remaining identified studies are, however, not directly excluded from the review as they nevertheless fit the criteria defined in the review protocol. Any studies for which data were not extracted due to saturation having been reached, but that fit the inclusion criteria of the protocol, were listed in the table for studies 'identified but not included due to saturation' in an appendix to the qualitative evidence review. During the literature search for qualitative reviews all potentially relevant papers will be reviewed, however saturation 'within themes' can occur when no new evidence adds to those already identified themes. At this point studies will only be included if they contribute towards the development of existing themes or to the development of new themes.

2.3.2 Type of studies

Systematic reviews of randomised trials, randomised trials, non-randomised intervention studies, and other observational studies (including risk prediction studies) were included in the evidence reviews as appropriate.

For most intervention reviews in this guideline, parallel randomised controlled trials (RCTs) were included because they are considered the most robust type of study design that can produce an unbiased estimate of the intervention effects. If non-randomised intervention studies were considered appropriate for inclusion (for example, where no randomised

evidence was available for critical outcomes) the committee stated a priori in the protocol that either certain identified variables must be equivalent at baseline or else the analysis had to adjust for any baseline differences. If the study did not fulfil either criterion it was excluded. Please refer to the review protocols in each evidence report for full details on the study design of studies selected for each review question.

Where data from non-randomised studies were included, the results for each outcome were presented separately for each study or meta-analysed if appropriate.

2.3.3 Methods of combining clinical studies

2.3.3.1 Data synthesis for intervention reviews

Where possible, meta-analyses were conducted using Cochrane Review Manager (RevMan5)[6] software to combine the data given in all studies for each of the outcomes of interest for the review question.

2.3.3.1.1 Analysis of different types of data

Dichotomous outcomes

Fixed-effects (Mantel–Haenszel) techniques (using an inverse variance method for pooling) were used to calculate risk ratios (relative risk, RR) for the binary outcomes, which included:

- mortality
- complication
- ICU admission
- rehospitalisation

The absolute risk difference was also calculated using GRADEpro[1] software, using the median event rate in the control arm of the pooled results.

For binary variables where there were zero events in either arm or a less than 1% event rate, Peto odds ratios, rather than risk ratios, were calculated. Peto odds ratios are more appropriate for data with a low number of events.

In cases where relative risk could not be calculated due to the event rate being greater than the total population number (i.e. 12 events in 10 people), results were reported narratively.

Continuous outcomes

Continuous outcomes were analysed using an inverse variance method for pooling weighted mean differences. These outcomes included:

- heath-related quality of life (HRQoL)
- length of stay in hospital
- pain scores
- symptom scores and functional measures
- psychological distress and mental wellbeing

Where the studies within a single meta-analysis had different scales of measurement, standardised mean differences were used (providing all studies reported either change from baseline or final values rather than a mixture of both); each different measure in each study was ‘normalised’ to the standard deviation value pooled between the intervention and comparator groups in that same study.

The means and standard deviations of continuous outcomes are required for meta-analysis. However, in cases where standard deviations were not reported, the standard error was calculated if the p values or 95% confidence intervals (95% CI) were reported, and meta-analysis was undertaken with the mean and standard error using the generic inverse variance method in Cochrane Review Manager (RevMan5) [6] software. Where p values were reported as 'less than', a conservative approach was undertaken. For example, if a p value was reported as 'p≤0.001', the calculations for standard deviations were based on a p value of 0.001. If these statistical measures were not available then the methods described in section 16.1.3 of the Cochrane Handbook (version 5.1.0, updated March 2011) were applied.

2.3.3.1.2 Generic inverse variance

If a study reported only the summary statistic and 95% CI the generic-inverse variance method was used to enter data into RevMan5[6]. If the control event rate was reported this was used to generate the absolute risk difference in GRADEpro[1]. If multivariate analysis was used to derive the summary statistic but no adjusted control event rate was reported no absolute risk difference was calculated.

2.3.3.1.3 Heterogeneity

Statistical heterogeneity was assessed for each meta-analysis estimate by considering the chi-squared test for significance at $p<0.1$ or an I-squared (I^2) inconsistency statistic (with an I-squared value of more than 50% indicating significant heterogeneity) as well as the distribution of effects. Where significant heterogeneity was present, studies were sub-grouped based on predefined categories specified by the committee at the protocol setting stage.

If the subgroup analysis resolved heterogeneity within all of the derived subgroups, then each of the derived subgroups were adopted as separate outcomes (providing at least 1 study remained in each subgroup. Assessments of potential differences in effect between subgroups were based on the chi-squared tests for heterogeneity statistics between subgroups. Any subgroup differences were interpreted with caution as separating the groups breaks the study randomisation and as such is subject to uncontrolled confounding.

For some questions additional subgrouping was applied, and this is documented in the individual review question protocols. These additional subgrouping strategies were applied independently, so subunits of subgroups were not created, unlike the situation with strata.

If all predefined strategies of subgrouping were unable to explain statistical heterogeneity within each derived subgroup, then a random effects (DerSimonian and Laird) model was employed to the entire group of studies in the meta-analysis. A random-effects model assumes a distribution of populations, rather than a single population. This leads to a widening of the confidence interval around the overall estimate, thus providing a more realistic interpretation of the true distribution of effects across more than 1 population. If, however, the committee considered the heterogeneity was so large that meta-analysis was inappropriate, then the results were described narratively.

2.3.3.2 Data synthesis for risk prediction rules

Evidence reviews on risk prediction rules or risk prediction tool results were presented separately for discrimination and calibration.

The discrimination data were analysed according to model accuracy. Model accuracy measures used in the analysis were: c-statistic or area under the receiver operating

characteristics (ROC) curve (AUC), and, if appropriate, sensitivity and specificity. Heterogeneity or inconsistency amongst studies was visually inspected. The c-statistic data for each study were also plotted on a graph, for each prediction model. The c-statistic describes the overall discriminative performance across the full range of thresholds. The following criteria were used for evaluating

c-statistic:

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

Calibration data such as r-squared (R2) or expected/observed event ratio were presented separately to the discrimination data. The results were presented for each study separately along with the quality rating for the study. As the majority of calibration data was reported as a single outcome of the total number of observed events (morbidity or mortality) compared to the expected event rate, a simple observed/expected event ratio was reported. Hosmer-Lemeshow statistical analysis could not be performed.

2.3.3.3 Data synthesis for qualitative study reviews

The main findings for each included paper were identified and thematic analysis methods were used to synthesise this information into broad overarching themes which were summarised into the main review findings. The evidence was presented in the form of a narrative summary detailing the evidence from the relevant papers and how this informed the overall review finding plus a statement on the level of confidence for that review finding. Considerable limitations and issues around relevance were listed. A summary evidence table with the succinct summary statements for each review finding was produced including the associated quality assessment.

2.3.4 Appraising the quality of evidence by outcomes

2.3.4.1 Intervention reviews

The evidence for outcomes from the included RCTs and, where appropriate, non-randomised intervention studies, were evaluated and presented using an adaptation of the 'Grading of Recommendations Assessment, Development and Evaluation (GRADE) toolbox' developed by the international GRADE working group (http://www.gradeworkinggroup.org/). The software (GRADEpro[1]) developed by the GRADE working group was used to assess the quality of each outcome, taking into account individual study quality and the meta-analysis results.

Each outcome was first examined for each of the quality elements listed and defined in Table 2.

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

Quality element	Description
Risk of bias	Limitations in the study design and implementation may bias the estimates of the treatment effect. Major limitations in studies decrease the confidence in the estimate of the effect. Examples of such limitations are selection bias (often due to poor allocation concealment), performance and detection bias (often due to a lack of blinding of the patient, healthcare professional or assessor) and attrition

Quality element	Description
	bias (due to missing data causing systematic bias in the analysis).
Indirectness	Indirectness refers to differences in study population, intervention, comparator and outcomes between the available evidence and the review question.
Inconsistency	Inconsistency refers to an unexplained heterogeneity of effect estimates between studies in the same meta-analysis.
Imprecision	Results are imprecise when studies include relatively few patients and few events (or highly variable measures) and thus have wide confidence intervals around the estimate of the effect relative to clinically important thresholds. 95% confidence intervals denote the possible range of locations of the true population effect at a 95% probability, and so wide confidence intervals may denote a result that is consistent with conflicting interpretations (for example a result may be consistent with both clinical benefit AND clinical harm) and thus be imprecise.
Publication bias	Publication bias is a systematic underestimate or overestimate of the underlying beneficial or harmful effect due to the selective publication of studies. A closely related phenomenon is where some papers fail to report an outcome that is inconclusive, thus leading to an overestimate of the effectiveness of that outcome.
Other issues	Sometimes randomisation may not adequately lead to group equivalence of confounders, and if so this may lead to bias, which should be taken into account. Potential conflicts of interest, often caused by excessive pharmaceutical company involvement in the publication of a study, should also be noted.

Details of how the 4 main quality elements (risk of bias, indirectness, inconsistency and imprecision) were appraised for each outcome are given below. Publication or other bias was only taken into consideration in the quality assessment if it was apparent.

2.3.4.1.1 Risk of bias

The main domains of bias for RCTs are listed in Table 3. Each outcome had its risk of bias assessed within each study first. For each study, if there were no risks of bias in any domain, the risk of bias was given a rating of 0. If there was risk of bias in just 1 domain, the risk of bias was given a 'serious' rating of −1, but if there was risk of bias in 2 or more domains the risk of bias was given a 'very serious' rating of −2. A weighted average score was then calculated across all studies contributing to the outcome, by taking into account the weighting of studies according to study precision. For example if the most precise studies tended to each have a score of −1 for that outcome, the overall score for that outcome would tend towards −1.

Table 3: Principle domains of bias in randomised controlled trials

Limitation	Explanation
Selection bias (sequence generation and allocation concealment)	If those enrolling patients are aware of the group to which the next enrolled patient will be allocated, either because of a non-random sequence that is predictable, or because a truly random sequence was not concealed from the researcher, this may translate into systematic selection bias. This may occur if the researcher chooses not to recruit a participant into that specific group because of: • knowledge of that participant's likely prognostic characteristics, and • a desire for one group to do better than the other.
Performance and detection bias (lack of blinding of patients and healthcare	Patients, caregivers, those adjudicating or recording outcomes, and data analysts should not be aware of the arm to which patients are allocated. Knowledge of the group can influence: • the experience of the placebo effect

Limitation	Explanation
professionals)	• performance in outcome measures • the level of care and attention received, and • the methods of measurement or analysis all of which can contribute to systematic bias.
Attrition bias	Attrition bias results from an unaccounted for loss of data beyond a certain level (a differential of 10% between groups). Loss of data can occur when participants are compulsorily withdrawn from a group by the researchers (for example, when a per-protocol approach is used) or when participants do not attend assessment sessions. If the missing data are likely to be different from the data of those remaining in the groups, and there is a differential rate of such missing data from groups, systematic attrition bias may result.
Selective outcome reporting	Reporting of some outcomes and not others on the basis of the results can also lead to bias, as this may distort the overall impression of efficacy.
Other limitations	For example: • Stopping early for benefit observed in randomised trials, in particular in the absence of adequate stopping rules. • Use of unvalidated patient-reported outcome measures. • Lack of washout periods to avoid carry-over effects in crossover trials. • Recruitment bias in cluster-randomised trials.

2.3.4.1.2 The assessment of risk of bias differs for non-randomised intervention studies, as they are inherently at high risk of selection bias. For this reason, GRADE requires that non-randomised evidence is initially downgraded on the basis of study design, starting with a rating of −2. This accounts for selection bias and so non-randomised intervention studies are not downgraded any further on that domain. Non-randomised evidence was assessed against the remaining domains used for RCTs in ***Table 3*** and downgraded further as appropriate.

Where systematic reviews were included, risk of bias was assessed using the ROBIS checklist. ROBIS considers four domains of potential bias with systematic reviews: study eligibility; identification and selection of studies; data collection and study appraisal; synthesis and findings. Details on each of these domains are given in Table 5.

Table 4: Overall quality of outcome evidence in GRADE

Domain	Signalling question
Study eligibility	Did the review adhere to pre-defined objectives and eligibility criteria?
	Were the eligibility criteria appropriate for the review question?
	Were eligibility criteria unambiguous?
	Were all restrictions in eligibility criteria based on study characteristics appropriate?
	Were any restrictions in eligibility criteria based on sources of information appropriate?
Identification and selection of studies	Did the search include an appropriate range of databases/ electronic sources for published and unpublished reports?
	Were methods additional to database searching used to identify relevant reports?
	Were the terms and structure of the search strategy likely to retrieve as many eligible studies as possible?
	Were restrictions based on date, publication format, or language appropriate?
	Were efforts made to minimise errors in selection of studies?
Data collection and study	Were efforts made to minimise error in data collection?
	Were sufficient study characteristics available for both review authors and

Domain	Signalling question
appraisal	readers to be able to interpret the results?
	Were all relevant study results collected for use in the synthesis?
	Was risk of bias (or methodological quality) formally assessed using appropriate criteria?
	Were efforts made to minimise error in risk of bias assessment?
Synthesis and findings	Did the synthesis include all studies that it should?
	Were all predefined analyses followed or departures explained?
	Was the synthesis appropriate given the nature and similarity in the research questions, study designs and outcomes across included studies?
	Was between-studies variation (heterogeneity) minimal or addressed in the synthesis?
	Were the findings robust, e.g. as demonstrated through funnel plot or sensitivity analyses?
	Were biases in primary studies minimal or addressed in the synthesis?

2.3.4.1.3 Indirectness

Indirectness refers to the extent to which the populations, interventions, comparisons and outcome measures are dissimilar to those defined in the inclusion criteria for the reviews. 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. As for the risk of bias, each outcome had its indirectness assessed within each study first. For each study, if there were no sources of indirectness, indirectness was given a rating of 0. If there was indirectness in just 1 source (for example in terms of population), indirectness was given a 'serious' rating of −1, but if there was indirectness in 2 or more sources (for example, in terms of population and treatment) the indirectness was given a 'very serious' rating of −2. A weighted average score was then calculated across all studies contributing to the outcome by taking into account study precision. For example, if the most precise studies tended to have an indirectness score of −1 each for that outcome, the overall score for that outcome would tend towards −1.

2.3.4.1.4 Inconsistency

Inconsistency refers to an unexplained heterogeneity of results for an outcome across different studies. When estimates of the treatment effect across studies differ widely, this suggests true differences in the underlying treatment effect, which may be due to differences in populations, settings or doses. When heterogeneity existed within an outcome (chi-squared $p<0.1$, or $I^2>50\%$), but no plausible explanation could be found, the quality of evidence for that outcome was downgraded. Inconsistency for that outcome was given a 'serious' score of −1 if the I^2 was 50–74%, and a 'very serious' score of −2 if the I^2 was 75% or more.

If inconsistency could be explained based on prespecified subgroup analysis (that is, each subgroup had an $I^2<50\%$), the committee took this into account and considered whether to make separate recommendations on new outcomes based on the subgroups defined by the assumed explanatory factors. In such a situation the quality of evidence was not downgraded for those emergent outcomes.

Since the inconsistency score was based on the meta-analysis results, the score represented the whole outcome and so weighted averaging across studies was not necessary.

2.3.4.1.5 Imprecision

The criteria applied for imprecision were based on the 95% CIs for the pooled estimate of effect, and the minimal important differences (MID) for the outcome. The MIDs are the threshold for appreciable benefits and harms, separated by a zone either side of the line of no effect where there is assumed to be no clinically important effect. If either end of the 95% CI of the overall estimate of effect crossed 1 of the MID lines, imprecision was regarded as serious and a 'serious' score of −1 was given. This was because the overall result, as represented by the span of the confidence interval, was consistent with 2 interpretations as defined by the MID (for example, both no clinically important effect and clinical benefit were possible interpretations). If both MID lines were crossed by either or both ends of the 95% CI then imprecision was regarded as very serious and a 'very serious' score of −2 was given. This was because the overall result was consistent with all 3 interpretations defined by the MID (no clinically important effect, clinical benefit and clinical harm). This is illustrated in Figure 2. As for inconsistency, since the imprecision score was based on the meta-analysis results, the score represented the whole outcome and so weighted averaging across studies was not necessary.

The position of the MID lines is ideally determined by values reported in the literature. 'Anchor-based' methods aim to establish clinically meaningful changes in a continuous outcome variable by relating or 'anchoring' them to patient-centred measures of clinical effectiveness that could be regarded as gold standards with a high level of face validity. For example, a MID for an outcome could be defined by the minimum amount of change in that outcome necessary to make patients feel their quality of life had 'significantly improved'. MIDs in the literature may also be based on expert clinician or consensus opinion concerning the minimum amount of change in a variable deemed to affect quality of life or health. For binary variables, any MIDs reported in the literature will inevitably be based on expert consensus, as such MIDs relate to all-or-nothing population effects rather than measurable effects on an individual, and so are not amenable to patient-centred 'anchor' methods.

In the absence of values identified in the literature, the alternative approach to deciding on MID levels is the 'default' method, as follows:

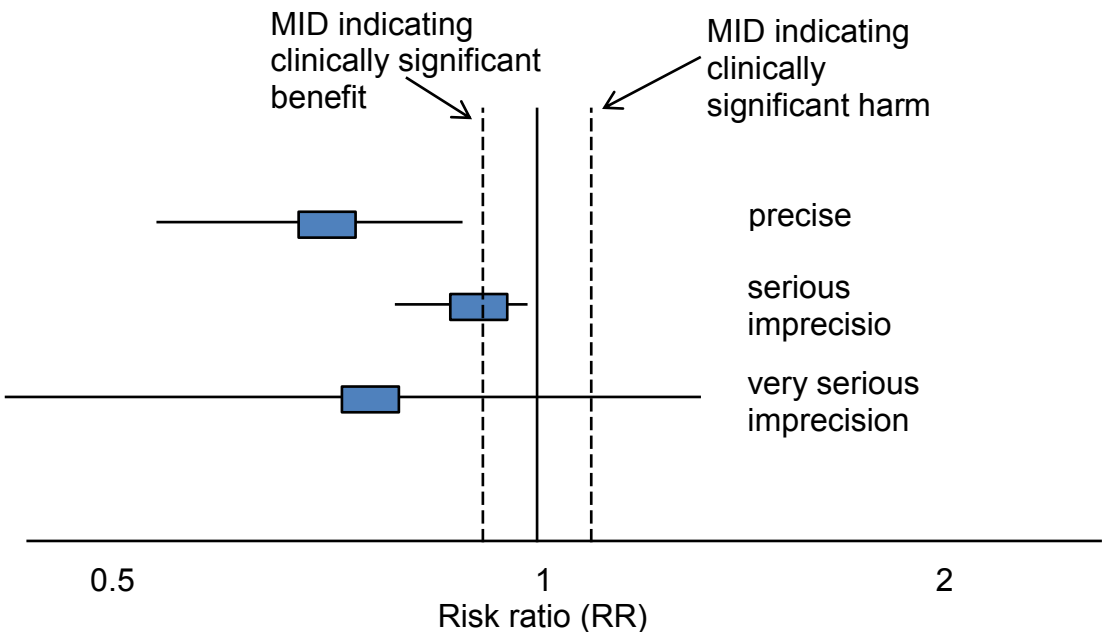
- For categorical outcomes the MIDs were taken to be RRs or Peto OR of 0.80 and 1.25. For 'positive' outcomes such as 'patient satisfaction', the point of 0.80 is taken as the line denoting the boundary between no clinically important effect and a clinically significant harm, whilst the point of 1.25 is taken as the line denoting the boundary between no clinically important effect and a clinically significant benefit. For 'negative' outcomes such as 'bleeding', the opposite occurs, so the point of 0.80 is taken as the line denoting the boundary between no clinically important effect and a clinically significant benefit, whilst the point of 1.25 is taken as the line denoting the boundary between no clinically important effect and a clinically significant harm.
- For continuous outcome variables the MID was taken as half the median baseline standard deviation of that variable, across all studies in the meta-analysis. Hence the MID denoting the minimum clinically significant benefit was positive for a 'positive' outcome (for example, a quality of life measure where a higher score denotes better health), and negative for a 'negative' outcome (for example, a visual analogue scale [VAS] pain score). Clinically significant harms will be the converse of these. If baseline values are unavailable, then half the median comparator group standard deviation of that variable will be taken as the MID.
- If standardised mean differences have been used, then the MID will be set at the absolute value of +0.5. This follows because standardised mean differences are mean differences normalised to the pooled standard deviation of the 2 groups, and are thus effectively expressed in units of 'numbers of standard deviations'. The 0.5 MID value in this context therefore indicates half a standard deviation, the same definition of MID as used for non-standardised mean differences.

- If Risk Difference has been used, then imprecision will be assessed by inference for proportions and the calculation of the optimal information size to provide desired power. Values of <80% represent very serious imprecision, 80-90% serious imprecision, and >90% power – no imprecision.

The default MID value was subject to amendment after discussion with the committee. If the committee decided that the MID level should be altered, after consideration of absolute as well as relative effects, this was allowed, provided that any such decision was not influenced by any bias towards making stronger or weaker recommendations for specific outcomes.

For this guideline, the default method was adopted except for mortality and modality failure where the committee agreed, based on their consensus, that the 0.8 to 1.25 MIDs were too wide and decided on the use of 0.9 to 1.11 instead.

Figure 2: Illustration of precise and imprecise outcomes based on the 95% CI of dichotomous outcomes in a forest plot (Note that all 3 results would be pooled estimates, and would not, in practice, be placed on the same forest plot)



2.3.4.1.6 Overall grading of the quality of clinical evidence

Once an outcome had been appraised for the main quality elements, as above, an overall quality grade was calculated for that outcome. The scores (0, −1 or −2) from each of the main quality elements were summed to give a score that could be anything from 0 (the best possible) to −8 (the worst possible). However scores were capped at −3. This final score was then applied to the starting grade that had originally been applied to the outcome by default, based on study design. All RCTs started as High and the overall quality became Moderate, Low or Very Low if the overall score was −1, −2 or −3 points respectively. The significance of these overall ratings is explained in Table 5**Error! Not a valid bookmark self-reference.**. The reasons for downgrading in each case were specified in the footnotes of the GRADE tables.

Non-randomised intervention studies started at Low, and so a score of −1 would be enough to take the grade to the lowest level of Very Low. Non-randomised intervention studies could, however, be upgraded if there was a large magnitude of effect or a dose-response gradient.

Table 5: Overall quality of outcome evidence in GRADE

Level	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

2.3.4.2 Prognostic reviews

The quality of evidence for prognostic studies was evaluated according to the criteria given in Table 6. If data were meta-analysed, the quality for pooled studies was presented. If the data were not pooled, then a quality rating was presented for each study.

2.3.5 Risk prediction studies

The outcomes from evidence for the included risk tool studies were assessed against the same principles of quality listed for intervention reviews. These were considered against criteria specifically for risk prediction evidence reviews (outlined below).

2.3.5.1.1 Risk of bias

Risk of bias and applicability was assessed for each study using PROBAST (Table 6:). Bias occurs if systematic flaws or limitations in the design, conduct or analysis of a primary study distort the results. Risk of bias refers to the likelihood that a prediction model leads to distorted predictive performance for its intended use and targeted individuals. The predictive performance is typically evaluated using calibration, discrimination, and (re)classification. Applicability refers to the extent to which the prediction model from the primary study matches your systematic review question, for example in terms of the population or outcomes of interest.

Table 6: Summary of PROBAST risk of bias and applicability questions.

Domain	Risk of bias/Applicability
Participant selection	1. Were appropriate data sources used, for example, cohort, RCT or nested case-control study data? 2. Were all inclusions and exclusions of participants appropriate? 3. Were participants enrolled at a similar state of health, or were predictors considered to account for differences?
Predictors	1. Were predictors defined and assessed in a similar way for all participants? 2. Were predictor assessments made without knowledge of outcome data? 3. Are all predictors available at the time the model is intended to be used? 4. Were all relevant predictors analysed?
Outcome	1. Was a pre-specified outcome definition used? 2. Were predictors excluded from the outcome definition? 3. Was the outcome defined and determined in a similar way for all participants? 4. Was the outcome determined without knowledge of predictor information?

Domain	Risk of bias/Applicability
Sample size and participant flow	1. Were there a reasonable number of outcome events? 2. Was the time interval between predictor assessment and outcome determination appropriate? 3. Were all enrolled participants included in the analysis? 4. Were participants with missing data handled appropriately?
Analysis	1. Were non-binary predictors handled appropriately? 2. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately? 3. For the model or any simplified score, were relevant performance measures evaluated, e.g. calibration, discrimination, (re)classification and net benefit? 4. Was the model recalibrated or was it likely (based on the evidence presented, e.g. calibration plot) that recalibration was not needed?
Applicability	Is there concern that the included participants and setting do not match the review question?
	Is there concern that the definition, assessment or timing of assessment of predictors in the model do not match the review question?
	Is there concern that the outcome, its definition, timing or determination do not match the review question?

2.3.5.1.2 Inconsistency

Inconsistency was assessed by visual inspection of a plotted summary of c-statistics and calibration data, and for overlap of confidence intervals where possible.

2.3.5.1.3 Imprecision

Imprecision was assessed by visual inspection of the position of the 95% CIs for each study in relation to the threshold criteria for c-statistic accuracy (e.g. ≤0.50: worse than chance; 0.50–0.60: very poor; 0.61–0.70: poor; 0.71–0.80: moderate; 0.81–0.92: good; 0.91–1.00: excellent or perfect test). If the 95% CI crossed the threshold for ≥2 categories then serious imprecision was recorded. If no information was given on the CI of the point estimate then the study was downgraded for having an unclear level of imprecision. A weighted average of imprecision across all studies contributing evidence for each risk tool was taken.

The judgement of precision was not possible for O/E ratio data with no means of assessing data variance (such as 95% CI). The overall summary outcomes were downgraded due to this limitation and the possible risk of imprecise outcomes.

2.3.5.1.4 Overall grading

Quality rating started at High for prospective studies, and each major limitation brought the rating down by 1 increment to a minimum grade of Very Low, as explained for interventional reviews.

2.3.6 Qualitative reviews

Review findings from the included qualitative studies were evaluated and presented using the 'Confidence in the Evidence from Reviews of Qualitative Research' (CERQual) Approach developed by the GRADE-CERQual Project Group, a subgroup of the GRADE Working Group.

The CERQual Approach assesses the extent to which a review finding is a reasonable representation of the phenomenon of interest (the focus of the review question). Each review finding was assessed for each of the 4 quality elements listed and defined below in Table 7.

Table 7: Description of quality elements in GRADE-CERQual for qualitative studies

Quality element	Description
Methodological limitations	The extent of problems in the design or conduct of the included studies that could decrease the confidence that the review finding is a reasonable representation of the phenomenon of interest. Assessed at the study level using an CASP checklist.
Coherence	The extent to which the reviewer is able to identify a clear pattern across the studies included in the review.
Relevance	The extent to which the body of evidence from the included studies is applicable to the context (study population, phenomenon of interest, setting) specified in the protocol.
Adequacy	The degree of the confidence that the review finding is being supported by sufficient data. This is an overall determination of the richness (depth of analysis) and quantity of the evidence supporting a review finding or theme.

Details of how the 4 quality elements (methodological limitations, coherence, relevance and adequacy) were appraised for each review finding are given below.

2.3.6.1.1 Methodological limitations

Each review finding had its methodological limitations assessed within each study first using an CASP checklist. Based on the degree of methodological limitations studies were evaluated as having minor, moderate or severe limitations. The questions to be answered in the checklist below included:

- Was qualitative design an appropriate approach?
- Was the study approved by an ethics committee?
- Was the study clear in what it sought to do?
- Is the context clearly described?
- Is the role of the researcher clearly described?
- Are the research design and methods rigorous?
- Was the data collection rigorous?
- Was the data analysis rigorous?
- Are the data rich?
- Are the findings relevant to the aims of the study?
- Are the findings and conclusions convincing?

The overall assessment of the methodological limitations of the evidence was based on the primary studies contributing to the review finding. The relative contribution of each study to the overall review finding and of the type of methodological limitation(s) were taken into account when giving an overall rating.

2.3.6.1.2 Coherence

Coherence is the extent to which the reviewer is able to identify a clear pattern across the studies included in the review, and if there is variation present (contrasting or disconfirming data) whether this variation is explained by the contributing study authors. If a review finding in 1 study does not support the main finding and there is no plausible explanation for this variation, then the confidence that the main finding reasonably reflects the phenomenon of interest is decreased. Each review finding was given a rating of minor, moderate or major concerns about coherence.

2.3.6.1.3 Relevance

Relevance is the extent to which the body of evidence from the included studies is applicable to the context (study population, phenomenon of interest, setting) specified in the protocol. As such, relevance is dependent on the individual review and discussed with the guideline committee. Relevance is categorised in 3 ways: partial relevance, indirect relevance and no concerns about relevance.

2.3.6.1.4 Adequacy

The judgement of adequacy is based on the confidence of the finding being supported by sufficient data. This is an overall determination of the richness (depth of analysis) and quantity of the evidence supporting a review finding or theme. Rich data provide sufficient detail to gain an understanding of the theme or review finding, whereas thin data do not provide enough detail for an adequate understanding. Quantity of data is the second pillar of the assessment of adequacy. For review findings that are only supported by 1 study or data from only a small number of participants, the confidence that the review finding reasonable represents the phenomenon of interest might be decreased. As with richness of data, quantity of data is review dependent. Based on the overall judgement of adequacy, a rating of no concerns, minor concerns, or substantial concerns about adequacy was given.

2.3.6.1.5 Overall judgement of the level of confidence for a review finding

GRADE-CERQual is used to assess the body of evidence as a whole through a confidence rating representing the extent to which a review finding is a reasonable representation of the phenomenon of interest. The 4 components (methodological limitations, coherence, relevance and adequacy) are used in combination to form an overall judgement. GRADE-CERQual uses 4 levels of confidence: high, moderate, low and very low confidence. The significance of these overall ratings is explained in Table 8. Each review finding starts at a high level of confidence and is downgraded based on the concerns identified in any 1 or more of the 4 components. Quality assessment of qualitative reviews is a subjective judgement by the reviewer based on the concerns that have been noted. A detailed explanation of how such a judgement had been made was included in the narrative summary.

Table 8: Overall level of confidence for a review finding in GRADE-CERQual

Level	Description
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.

2.4 Assessing clinical importance

The committee assessed the evidence by outcome in order to determine if there was, or potentially was, a clinically important benefit, a clinically important harm or no clinically important difference between interventions. To facilitate this, binary outcomes were converted into absolute risk differences (ARDs) using GRADEpro[1] software: the median control group risk across studies was used to calculate the ARD and its 95% CI from the pooled risk ratio.

The assessment of clinical benefit, harm, or no benefit or harm was based on the point estimate of absolute effect for intervention studies, which was standardised across the reviews. The committee considered for most of the outcomes in the intervention reviews that if at least 100 more participants per 1000 (10%) achieved the outcome of interest in the intervention group compared to the comparison group for a positive outcome then this intervention was considered beneficial. The same point estimate but in the opposite direction applied for a negative outcome. For the critical outcomes of mortality a change of at least 10 participants per 1000 (1%) represented a clinical important effect. For continuous outcomes if the mean difference was greater than the minimally important difference (MID) then this represented a clinical benefit or harm. For outcomes where validated MIDs were not available, a MID rule of half the standard deviation of the control group was used. The relative effect MIDs were used as the starting point for the committee's consideration of the importance of each outcome.

Outcomes reported in ways that prevented their analysis through GRADE (e.g. median values or mean values with no variance data) and subsequent conversion to ARDs were assessed for statistical significance. Outcome where no statistical significance was reported were otherwise assessed to review if a notable difference between groups was viable, as decided by the guideline committee,

Clinical evidence statements

Clinical evidence statements are summary statements that are included in each evidence report, and which summarise the key features of the clinical effectiveness evidence presented. The wording of the evidence statements reflects the certainty or uncertainty in the estimate of effect. The evidence statements are presented by outcome and encompass the following key features of the evidence:

- The number of studies for a particular outcome.
- An indication of the direction of clinical importance (if one treatment is beneficial or harmful compared to the other, or whether there is no difference between the 2 tested treatments).
- A description of the overall quality of the evidence (GRADE overall quality).

2.5 Reviewing economic evidence

The committee is required to make decisions based on the best available evidence of both clinical effectiveness and cost effectiveness. Guideline recommendations should be based on the expected costs of the different options in relation to their expected health benefits (that is, their 'cost effectiveness') rather than the total implementation cost. However, the committee will also need to be increasingly confident in the cost effectiveness of a recommendation as the cost of implementation increases. Therefore, the committee may require more robust evidence on the effectiveness and cost effectiveness of any recommendations that are expected to have a substantial impact on resources; any uncertainties must be offset by a compelling argument in favour of the recommendation. The cost impact or savings potential of a recommendation should not be the sole reason for the committee's decision.[3]

Health economic evidence was sought relating to the key clinical issues being addressed in the guideline. Health economists:

- Undertook a systematic review of the published economic literature.
- Undertook new cost-effectiveness analysis in priority areas.

2.5.1 Literature review

The health economists:

- Identified potentially relevant studies for each review question from the health economic search results by reviewing titles and abstracts. Full papers were then obtained.
- Reviewed full papers against prespecified inclusion and exclusion criteria to identify relevant studies (see below for details).
- Critically appraised relevant studies using economic evaluations checklists as specified in the NICE guidelines manual.[3]
- Extracted key information about the studies' methods and results into health economic evidence tables (which can be found in appendices to the relevant evidence reports).
- Generated summaries of the evidence in NICE health economic evidence profile tables (included in the relevant evidence report for each review question) – see below for details.

2.5.2 Inclusion and exclusion of economic studies

Full economic evaluations (studies comparing costs and health consequences of alternative courses of action: cost–utility, cost-effectiveness, cost–benefit and cost–consequences analyses) and comparative costing studies that addressed the review question in the relevant population were considered potentially includable as health economic evidence.

Studies that only reported cost per hospital (not per patient), or only reported average cost effectiveness without disaggregated costs and effects were excluded. Literature reviews, abstracts, posters, letters, editorials, comment articles, unpublished studies and studies not in English were excluded. Studies published before 2003 and studies from non-OECD countries or the USA were also excluded, on the basis that the applicability of such studies to the present UK NHS context is likely to be too low for them to be helpful for decision-making.

Remaining health economic studies were prioritised for inclusion based on their relative applicability to the development of this guideline and the study limitations. For example, if a high quality, directly applicable UK analysis was available, then other less relevant studies may not have been included. However, in this guideline, no economic studies were excluded on the basis that more applicable evidence was available.

For more details about the assessment of applicability and methodological quality see Table 9 below and the economic evaluation checklist (appendix H of the NICE guidelines manual[3]) and the health economics review protocol, which can be found in each of the evidence reports.

When no relevant health economic studies were found from the economic literature review, relevant UK NHS unit costs related to the compared interventions were presented to the committee to inform the possible economic implications of the recommendations.

2.5.2.1 NICE health economic evidence profiles

NICE health economic evidence profile tables were used to summarise cost and cost-effectiveness estimates for the included health economic studies in each evidence review report. The health economic evidence profile shows an assessment of applicability and methodological quality for each economic study, with footnotes indicating the reasons for the assessment. These assessments were made by the health economist using the economic evaluation checklist from the NICE guidelines manual.[3] It also shows the incremental costs, incremental effects (for example, quality-adjusted life years [QALYs]) and incremental cost-effectiveness ratio (ICER) for the base case analysis in the study, as well as information about the assessment of uncertainty in the analysis. See Table 9 for more details.

When a non-UK study was included in the profile, the results were converted into pounds sterling using the appropriate purchasing power parity.[5]

Table 9: Content of NICE health economic evidence profile

Item	Description
Study	Surname of first author, date of study publication and country perspective with a reference to full information on the study.
Applicability	An assessment of applicability of the study to this guideline, the current NHS situation and NICE decision-making:[(a)] • Directly applicable – the study meets all applicability criteria, or fails to meet 1 or more applicability criteria but this is unlikely to change the conclusions about cost effectiveness. • Partially applicable – the study fails to meet 1 or more applicability criteria, and this could change the conclusions about cost effectiveness. • Not applicable – the study fails to meet 1 or more of the applicability criteria, and this is likely to change the conclusions about cost effectiveness. Such studies would usually be excluded from the review.
Limitations	An assessment of methodological quality of the study:[(a)] • Minor limitations – the study meets all quality criteria, or fails to meet 1 or more quality criteria, but this is unlikely to change the conclusions about cost effectiveness. • Potentially serious limitations – the study fails to meet 1 or more quality criteria, and this could change the conclusions about cost effectiveness. • Very serious limitations – the study fails to meet 1 or more quality criteria, and this is highly likely to change the conclusions about cost effectiveness. Such studies would usually be excluded from the review.
Other comments	Information about the design of the study and particular issues that should be considered when interpreting it.
Incremental cost	The mean cost associated with one strategy minus the mean cost of a comparator strategy.
Incremental effects	The mean QALYs (or other selected measure of health outcome) associated with one strategy minus the mean QALYs of a comparator strategy.
Cost effectiveness	Incremental cost-effectiveness ratio (ICER): the incremental cost divided by the incremental effects (usually in £ per QALY gained).
Uncertainty	A summary of the extent of uncertainty about the ICER reflecting the results of deterministic or probabilistic sensitivity analyses, or stochastic analyses of trial data, as appropriate.

(a) Applicability and limitations were assessed using the economic evaluation checklist in appendix H of the NICE guidelines manual[3]

2.5.3 Undertaking new health economic analysis

As well as reviewing the published health economic literature for each review question, as described above, new health economic analysis was undertaken by the health economist in selected areas. Priority areas for new analysis were agreed by the committee after formation of the review questions and consideration of the existing health economic evidence.

The committee identified cardiac output monitoring as the highest priority area for original health economic modelling. This area was prioritised for various reasons. Firstly, previous medical technologies guidance on the CardioQ-ODM monitor (MTG3) showed recommended the use of CardioQ-ODM for patients during surgery on the basis it led to cost savings. Since the publication of this medical technologies guidance, practice has improved and the cost savings may no longer be relevant. Also, there is variation in current practice regarding whether the monitors are used during surgery, and the monitors themselves are costly and a positive recommendation would likely result in a cost impact. In addition, although economic studies were identified and included in this review, their date and methodological limitations meant that uncertainty still remained around the cost effectiveness of cardiac output monitoring. The rationale and details of the analysis undertaken is described in full in the technical report.

The following general principles were adhered to in developing the cost-effectiveness analysis:

- Methods were consistent with the NICE reference case for interventions with health outcomes in NHS settings.[3]
- The committee was involved in the design of the model, selection of inputs and interpretation of the results.
- Model inputs were based on the systematic review of the clinical literature supplemented with other published data sources where possible.
- When published data were not available committee expert opinion was used to populate the model.
- Model inputs and assumptions were reported fully and transparently.
- The results were subject to sensitivity analysis and limitations were discussed.
- The model was peer-reviewed by another health economist at the NGC.

Full methods and results of the cost-effectiveness analysis for cardiac output monitoring versus conventional clinical assessment during surgery are described in a separate economic analysis report.

2.5.4 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.[4] In general, an intervention was considered to be cost effective (given that the estimate was considered plausible) if either of the following criteria applied:

- 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.

If the committee recommended an intervention that was estimated to cost more than £20,000 per QALY gained, or did not recommend one that was estimated to cost less than £20,000 per QALY gained, the reasons for this decision are discussed explicitly in 'The committee's discussion of the evidence' section of the relevant evidence report, with reference to issues regarding the plausibility of the estimate or to the factors set out in 'Social value judgements: principles for the development of NICE guidance'.[4]

If a study reported the cost per life year gained but not QALYs, the cost per QALY gained was estimated by multiplying by an appropriate utility estimate to aid interpretation. The estimated cost per QALY gained is reported in the health economic evidence profile with a footnote detailing the life-years gained and the utility value used. When QALYs or life years gained are not used in the analysis, results are difficult to interpret unless one strategy dominates the others with respect to every relevant health outcome and cost.

2.5.5 In the absence of health economic evidence

When no relevant published health economic studies were found, and a new analysis was not prioritised, the committee made a qualitative judgement about cost effectiveness by considering expected differences in resource use between options and relevant UK NHS unit costs, alongside the results of the review of clinical effectiveness evidence.

The UK NHS costs reported in the guideline are those that were presented to the committee and were correct at the time recommendations were drafted. They may have changed

subsequently before the time of publication. However, we have no reason to believe they have changed substantially.

2.6 Developing recommendations

Over the course of the guideline development process, the committee was presented with:

- Summaries of clinical and health economic evidence and quality (as presented in evidence reports).
- Evidence tables of the clinical and health economic evidence reviewed from the literature. All evidence tables can be found in appendices to the relevant evidence reports.
- Forest plots (in appendices to the relevant evidence reports).
- A description of the methods and results of the cost-effectiveness analyses undertaken for the guideline.

Recommendations were drafted based on the committee's interpretation of the available evidence, taking into account the balance of benefits, harms and costs between different courses of action. This was either done formally in an economic model, or informally. Firstly, the net clinical benefit over harm (clinical effectiveness) was considered, focusing on the critical outcomes. When this was done informally, the committee took into account the clinical benefits and harms when one intervention was compared with another. The assessment of net clinical benefit was moderated by the importance placed on the outcomes (the committee's values and preferences), and the confidence the committee had in the evidence (evidence quality). Secondly, the committee assessed whether the net clinical benefit justified any differences in costs between the alternative interventions.

When clinical and health economic evidence was of poor quality, conflicting or absent, the committee drafted recommendations based on its expert opinion. The considerations for making consensus-based recommendations include the balance between potential harms and benefits, the economic costs compared to the economic benefits, current practices, recommendations made in other relevant guidelines, patient preferences and equality issues. The committee also considered whether the uncertainty was sufficient to justify delaying making a recommendation to await further research, taking into account the potential harm of failing to make a clear recommendation (see section below).

The committee considered the appropriate 'strength' of each recommendation. This takes into account the quality of the evidence but is conceptually different. Some recommendations are 'strong' in that the committee believes that the vast majority of healthcare and other professionals and patients would choose a particular intervention if they considered the evidence in the same way that the committee has. This is generally the case if the benefits clearly outweigh the harms for most people and the intervention is likely to be cost effective. However, there is often a closer balance between benefits and harms, and some patients would not choose an intervention whereas others would. This may happen, for example, if some patients are particularly averse to some side effect and others are not. In these circumstances, the recommendation is generally weaker, although it may be possible to make stronger recommendations about specific groups of patients.

The committee focused on the following factors in agreeing the wording of the recommendations:

- The actions health professionals need to take.
- The information readers need to know.
- The strength of the recommendation (for example the word 'offer' was used for strong recommendations and 'consider' for weaker recommendations).
- The involvement of patients (and their carers if needed) in decisions on treatment and care.

- Consistency with NICE's standard advice on recommendations about drugs, waiting times and ineffective interventions (see section 9.2 in the NICE guidelines manual[3]).

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

2.6.1 Research recommendations

When areas were identified for which good evidence was lacking, the committee considered making recommendations for future research. Decisions about the inclusion of a research recommendation were based on factors such as:

- the importance to patients or the population
- national priorities
- potential impact on the NHS and future NICE guidance
- ethical and technical feasibility.

2.6.2 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.

2.6.3 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.

2.6.4 Disclaimer

Healthcare providers need to use clinical judgement, knowledge and expertise when deciding whether it is appropriate to apply guidelines. The recommendations cited here are a guide and may not be appropriate for use in all situations. The decision to adopt any of the recommendations cited here must be made by practitioners in light of individual patient circumstances, the wishes of the patient, clinical expertise and resources.

The National Guideline Centre disclaims any responsibility for damages arising out of the use or non-use of this guideline and the literature used in support of this guideline.

2.6.5 Funding

The National Guideline Centre was commissioned by the National Institute for Health and Care Excellence to undertake the work on this guideline.

3 Acronyms and abbreviations

Acronym or abbreviation	Description
ACS	American college of surgeons
AKI	Acute kidney injury
ASA	American society of Anaesthesiologists
AU ROC	Area under receiver operating characteristic curve
BMI	Body mass index
CCA	conventional clinical assessment
CCI	Charlston comorbidity index
CEA	Cost-effectiveness analysis
CHO	Carbohydrate drink
CI	Confidence interval
COM	Cardiac output monitoring
CUA	Cost-utility analysis
CVP	central venous pressure
DOAC	Direct oral anticoagulant
EORTC QLQ-30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire
E-PASS	surgical-specific risk scoring system estimating the physiologic ability and surgical stress
ERAS	Enhanced recovery after surgery
ERP	Enhanced recovery program
EQ-5D	Quality of life scale
FID	Functional Iron Deficiency
FTS	Fast track surgery
GA	General anaesthetic
GI	Gastrointestinal
GRADE	Grading of Recommendations, Assessment, Development and Evaluations
HADS	Hospital anxiety and depression scale
Hb	Haemoglobin
HES	Heta-starch
HDU	High dependency unit
ICER	Incremental cost-effectiveness ratio
ICU	Intensive Care Unit
IM	Intramuscular
INR	International normalized ratio
IQR	Interquartile range
IV	Intravenous
LMWH	Low molecular weight heparin
M:F	Male to female ratio
mRS	Modified Rankin scale
NGC	National Guideline Centre
NICE	National Institute for Health and Care Excellence
NRS	Numerical rating scale

Acronym or abbreviation	Description
NSAIDs	Non-steroidal anti-inflammatory drugs
NSQIP	National Surgical Quality Improvement Program
OECD	Organisation for Economic Co-operation and Development
ODM	oesophageal Doppler monitoring
OR	Odds ratio
PACU	Post anaesthetic care unit
POD	Post-operative day
POP	Preoperative optimisation
POPS	Perioperative medicine for Older Patients undergoing surgery
POMS	postoperative morbidity score
PCA	Patient controlled analgesia
PPWA	pulse pressure waveform analysis
PCA	Patient controlled analgesia
PCA	pulse contour analysis
P-POSSUM	Physiological and Operative Severity Score for the enumeration of Mortality and Morbidity
QALY	Quality adjusted life year
QoR-40	Quality of recovery scale
RR	Risk ratio
SD	Standard deviation
SF-12	Quality of life scale
SORT	Surgical Outcome Risk Tool
UFH	Unfractionated heparin
VAS	Visual analogue scale
VKA	Vitamin K antagonists
WHO SSC	World Health Organization surgical safety checklist

4 Glossary

The NICE Glossary can be found at www.nice.org.uk/glossary.

4.1 Guideline-specific terms

Term	Definition
Anaemia	A condition in which there are not enough red blood cells or haemoglobin to meet the body's needs.
Anticoagulants	Medicines that help prevent blood clots.
Blood glucose	The amount of sugar in the blood.
Cardiac output monitoring	The measurement of blood delivery through the circularity system. Minimally invasive cardiac output monitors collectively describes all devices that calculate cardiac output without requiring insertion of a pulmonary artery catheter.
Cleveland Global Quality of Life scale	A patient reported measure of quality of life. Each domain score ranges from 0 to 10 where high scores indicate a positive outcome.
Central venous pressure	Venous pressure is a term that represents the average blood pressure within the venous compartment.
Carbohydrate drinks	Sugary (carbohydrate-rich) drinks
Crystalloid	Crystalloid solutions are isotonic plasma volume expanders that contain electrolytes. Crystalloids are a plasma volume expander used to increase a depleted circulating volume.
Colloids	Colloids are gelatinous solutions that maintain a high osmotic pressure in the blood. Colloids are a plasma volume expander used to increase a depleted circulating volume.
Enhanced recovery programmes	An evidence-based approach that helps people recover more quickly after having major surgery.
Epidural	Injection in the back to stop feeling pain in parts of the body.
EORTC QLQ-30 scale	A patient reported measure of quality of life in cancer patients. Scores range from 0 to 100 where high scores indicate a positive outcome.
EQ-5D scale	A patient reported measure of quality of life in cancer patients. Summary scores range from -0.59 to 1 where high scores indicate a positive outcome.
Fasting	Abstaining from all food or drink.
Functional Iron Deficiency	A state in which there is insufficient iron incorporation into erythroid precursors in the face of apparently adequate body iron stores, as defined by the presence of stainable iron in the bone marrow together with a serum ferritin value within normal limits.
Haemoglobin	A red protein responsible for transporting oxygen in the blood of vertebrates.
Heparin	A medication and naturally occurring glycosaminoglycan commonly used as an anticoagulant.
Hypoglycaemia	Where the level of sugar (glucose) in the blood drops too low.
Intraoperative Care	Health care that is given during a surgical procedure.
Intravenous iron	A procedure in which iron is delivered to the body intravenously, into a vein through a needle.
Intravenous fluids	Liquids given to replace water, sugar and salt.

Term	Definition
Iron deficiency	A state where the body has insufficient iron, often resulting in anaemia.
Ketamine	A drug to relieve pain.
Neuropathic nerve stabilisers	A class of drug with pain relieving properties.
Never events	A term used to describe certain serious patient safety incidents which can occur in hospital. What sets Never Events apart from other types of serious incidents is that they are regarded as being preventable when appropriate safety protocols are followed by healthcare professionals.
oesophageal Doppler monitoring	Oesophageal Doppler monitoring is undertaken with a single-use probe which is placed in the oesophagus via the mouth or nose. The device generates a low-frequency ultrasound signal, which is reflected by red blood cells travelling down the aorta. By applying the Doppler principle, the reflected signal can be used to determine flow velocity. A validated nomogram is used to derive volumetric data such as stoke volume and cardiac output from the directly measured flow velocity.
Oral iron therapy	Iron supplementation taken by mouth
Opioids	A drug to relieve pain.
Orthopaedic	The branch of surgery concerned with conditions involving the musculoskeletal system
Oedema	A condition characterized by an excess of watery fluid collecting in the cavities or tissues of the body
Paracetamol	A drug to relieve and reduce pain and fever.
Perioperative Care	Health care that is given before, during and after a surgical procedure.
Preoperative Care	Health care that is given before a surgical procedure.
Postoperative Care	Health care that is given after a surgical procedure.
Pre-operative optimisation	A process of preoperative monitoring and subsequent management of patients scheduled to undergo surgery to reduce the risk of the operation.
POPs clinics	Specialist clinics providing perioperative medicine for Older Patients undergoing surgery.
pulse pressure waveform analysis	Pulse pressure wave analysis is a technique that allows the recording of peripheral pressure waveforms and generation of the corresponding central waveform, from which the augmentation index and central pressure can be derived.
pulse contour analysis	Pulse contour analysis provides continuous cardiac output and stroke volume variation.
Risk tools	A surgical preoperative risk prediction tool that intends to predict the chances of morbidity and mortality with surgery.
Safety management systems	A series of defined, organization-wide processes that provide for risk-based decision-making related to the perioperative process.
SF-36 score	A patient reported measure of quality of life. Scores range from 0 to 100 where high scores indicate a positive outcome.
Warfarin	A type of medicine known as an anticoagulant, or blood thinner.

4.2 General terms

Term	Definition
Abstract	Summary of a study, which may be published alone or as an introduction to a full scientific paper.
Algorithm (in guidelines)	A flow chart of the clinical decision pathway described in the guideline, where decision points are represented with boxes, linked with arrows.
Allocation concealment	The process used to prevent advance knowledge of group assignment in an RCT. The allocation process should be impervious to any influence by the individual making the allocation, by being administered by someone who is not responsible for recruiting participants.
Applicability	How well the results of a study or NICE evidence review can answer a clinical question or be applied to the population being considered.
Arm (of a clinical study)	Subsection of individuals within a study who receive one particular intervention, for example placebo arm.
Association	Statistical relationship between 2 or more events, characteristics or other variables. The relationship may or may not be causal.
Base case analysis	In an economic evaluation, this is the main analysis based on the most plausible estimate of each input. In contrast, see Sensitivity analysis.
Baseline	The initial set of measurements at the beginning of a study (after run-in period where applicable), with which subsequent results are compared.
Bayesian analysis	A method of statistics, where a statistic is estimated by combining established information or belief (the 'prior') with new evidence (the 'likelihood') to give a revised estimate (the 'posterior').
Before-and-after study	A study that investigates the effects of an intervention by measuring particular characteristics of a population both before and after taking the intervention, and assessing any change that occurs.
Bias	Influences on a study that can make the results look better or worse than they really are. (Bias can even make it look as if a treatment works when it does not.) Bias can occur by chance, deliberately or as a result of systematic errors in the design and execution of a study. It can also occur at different stages in the research process, for example, during the collection, analysis, interpretation, publication or review of research data. For examples see selection bias, performance bias, information bias, confounding factor, and publication bias.
Blinding	A way to prevent researchers, doctors and patients in a clinical trial from knowing which study group each patient is in so they cannot influence the results. The best way to do this is by sorting patients into study groups randomly. The purpose of 'blinding' or 'masking' is to protect against bias. A single-blinded study is one in which patients do not know which study group they are in (for example whether they are taking the experimental drug or a placebo). A double-blinded study is one in which neither patients nor the researchers and doctors know which study group the patients are in. A triple blind study is one in which neither the patients, clinicians or the people carrying out the statistical analysis know which treatment patients received.
Carer (caregiver)	Someone who looks after family, partners or friends in need of help because they are ill, frail or have a disability.
Case–control study	A study to find out the cause(s) of a disease or condition. This is

Term	Definition
	done by comparing a group of patients who have the disease or condition (cases) with a group of people who do not have it (controls) but who are otherwise as similar as possible (in characteristics thought to be unrelated to the causes of the disease or condition). This means the researcher can look for aspects of their lives that differ to see if they may cause the condition. For example, a group of people with lung cancer might be compared with a group of people the same age that do not have lung cancer. The researcher could compare how long both groups had been exposed to tobacco smoke. Such studies are retrospective because they look back in time from the outcome to the possible causes of a disease or condition.
Case series	Report of a number of cases of a given disease, usually covering the course of the disease and the response to treatment. There is no comparison (control) group of patients.
Clinical efficacy	The extent to which an intervention is active when studied under controlled research conditions.
Clinical effectiveness	How well a specific test or treatment works when used in the 'real world' (for example, when used by a doctor with a patient at home), rather than in a carefully controlled clinical trial. Trials that assess clinical effectiveness are sometimes called management trials. Clinical effectiveness is not the same as efficacy.
Clinician	A healthcare professional who provides patient care. For example, a doctor, nurse or physiotherapist.
Cochrane Review	The Cochrane Library consists of a regularly updated collection of evidence-based medicine databases including the Cochrane Database of Systematic Reviews (reviews of randomised controlled trials prepared by the Cochrane Collaboration).
Cohort study	A study with 2 or more groups of people – cohorts – with similar characteristics. One group receives a treatment, is exposed to a risk factor or has a particular symptom and the other group does not. The study follows their progress over time and records what happens. See also observational study.
Comorbidity	A disease or condition that someone has in addition to the health problem being studied or treated.
Comparability	Similarity of the groups in characteristics likely to affect the study results (such as health status or age).
Concordance	This is a recent term whose meaning has changed. It was initially applied to the consultation process in which doctor and patient agree therapeutic decisions that incorporate their respective views, but now includes patient support in medicine taking as well as prescribing communication. Concordance reflects social values but does not address medicine-taking and may not lead to improved adherence.
Confidence interval (CI)	There is always some uncertainty in research. This is because a small group of patients is studied to predict the effects of a treatment on the wider population. The confidence interval is a way of expressing how certain we are about the findings from a study, using statistics. It gives a range of results that is likely to include the 'true' value for the population. The CI is usually stated as '95% CI', which means that the range of values has a 95 in a 100 chance of including the 'true' value. For example, a study may state that "based on our sample findings, we are 95% certain that the 'true' population blood pressure is not higher than 150 and not lower than 110". In such a case the 95% CI would be 110 to 150. A wide confidence interval indicates a lack of certainty about the true

Term	Definition
	effect of the test or treatment – often because a small group of patients has been studied. A narrow confidence interval indicates a more precise estimate (for example, if a large number of patients have been studied).
Confounding factor	Something that influences a study and can result in misleading findings if it is not understood or appropriately dealt with. For example, a study of heart disease may look at a group of people that exercises regularly and a group that does not exercise. If the ages of the people in the 2 groups are different, then any difference in heart disease rates between the 2 groups could be because of age rather than exercise. Therefore age is a confounding factor.
Consensus methods	Techniques used to reach agreement on a particular issue. Consensus methods may be used to develop NICE guidance if there is not enough good quality research evidence to give a clear answer to a question. Formal consensus methods include Delphi and nominal group techniques.
Control group	A group of people in a study who do not receive the treatment or test being studied. Instead, they may receive the standard treatment (sometimes called 'usual care') or a dummy treatment (placebo). The results for the control group are compared with those for a group receiving the treatment being tested. The aim is to check for any differences. Ideally, the people in the control group should be as similar as possible to those in the treatment group, to make it as easy as possible to detect any effects due to the treatment.
Cost–benefit analysis (CBA)	Cost–benefit analysis is one of the tools used to carry out an economic evaluation. The costs and benefits are measured using the same monetary units (for example, pounds sterling) to see whether the benefits exceed the costs.
Cost–consequences analysis (CCA)	Cost–consequences analysis is one of the tools used to carry out an economic evaluation. This compares the costs (such as treatment and hospital care) and the consequences (such as health outcomes) of a test or treatment with a suitable alternative. Unlike cost–benefit analysis or cost-effectiveness analysis, it does not attempt to summarise outcomes in a single measure (like the quality-adjusted life year) or in financial terms. Instead, outcomes are shown in their natural units (some of which may be monetary) and it is left to decision-makers to determine whether, overall, the treatment is worth carrying out.
Cost-effectiveness analysis (CEA)	Cost-effectiveness analysis is one of the tools used to carry out an economic evaluation. The benefits are expressed in non-monetary terms related to health, such as symptom-free days, heart attacks avoided, deaths avoided or life years gained (that is, the number of years by which life is extended as a result of the intervention).
Cost-effectiveness model	An explicit mathematical framework, which is used to represent clinical decision problems and incorporate evidence from a variety of sources in order to estimate the costs and health outcomes.
Cost–utility analysis (CUA)	Cost–utility analysis is one of the tools used to carry out an economic evaluation. The benefits are assessed in terms of both quality and duration of life, and expressed as quality-adjusted life years (QALYs). See also utility.
Credible interval (CrI)	The Bayesian equivalent of a confidence interval.
Decision analysis	An explicit quantitative approach to decision-making under uncertainty, based on evidence from research. This evidence is translated into probabilities, and then into diagrams or decision trees which direct the clinician through a succession of possible scenarios,

Term	Definition
	actions and outcomes.
Deterministic analysis	In economic evaluation, this is an analysis that uses a point estimate for each input. In contrast, see Probabilistic analysis
Diagnostic odds ratio	The diagnostic odds ratio is a measure of the effectiveness of a diagnostic test. It is defined as the ratio of the odds of the test being positive if the subject has a disease relative to the odds of the test being positive if the subject does not have the disease.
Discounting	Costs and perhaps benefits incurred today have a higher value than costs and benefits occurring in the future. Discounting health benefits reflects individual preference for benefits to be experienced in the present rather than the future. Discounting costs reflects individual preference for costs to be experienced in the future rather than the present.
Disutility	The loss of quality of life associated with having a disease or condition. See Utility
Dominance	A health economics term. When comparing tests or treatments, an option that is both less effective and costs more is said to be 'dominated' by the alternative.
Drop-out	A participant who withdraws from a trial before the end.
Economic evaluation	An economic evaluation is used to assess the cost effectiveness of healthcare interventions (that is, to compare the costs and benefits of a healthcare intervention to assess whether it is worth doing). The aim of an economic evaluation is to maximise the level of benefits – health effects – relative to the resources available. It should be used to inform and support the decision-making process; it is not supposed to replace the judgement of healthcare professionals. There are several types of economic evaluation: cost–benefit analysis, cost–consequences analysis, cost-effectiveness analysis, cost-minimisation analysis and cost–utility analysis. They use similar methods to define and evaluate costs, but differ in the way they estimate the benefits of a particular drug, programme or intervention.
Effect (as in effect measure, treatment effect, estimate of effect, effect size)	A measure that shows the magnitude of the outcome in one group compared with that in a control group. For example, if the absolute risk reduction is shown to be 5% and it is the outcome of interest, the effect size is 5%. The effect size is usually tested, using statistics, to find out how likely it is that the effect is a result of the treatment and has not just happened by chance (that is, to see if it is statistically significant).
Effectiveness	How beneficial a test or treatment is under usual or everyday conditions, compared with doing nothing or opting for another type of care.
Efficacy	How beneficial a test, treatment or public health intervention is under ideal conditions (for example, in a laboratory), compared with doing nothing or opting for another type of care.
Epidemiological study	The study of a disease within a population, defining its incidence and prevalence and examining the roles of external influences (for example, infection, diet) and interventions.
EQ-5D (EuroQol 5 dimensions)	A standardised instrument used to measure health-related quality of life. It provides a single index value for health status.
Evidence	Information on which a decision or guidance is based. Evidence is obtained from a range of sources including randomised controlled trials, observational studies, expert opinion (of clinical professionals or patients).
Exclusion criteria (literature review)	Explicit standards used to decide which studies should be excluded from consideration as potential sources of evidence.

Term	Definition
Exclusion criteria (clinical study)	Criteria that define who is not eligible to participate in a clinical study.
Extended dominance	If Option A is both more clinically effective than Option B and has a lower cost per unit of effect, when both are compared with a do-nothing alternative then Option A is said to have extended dominance over Option B. Option A is therefore cost effective and should be preferred, other things remaining equal.
Extrapolation	An assumption that the results of studies of a specific population will also hold true for another population with similar characteristics.
Follow-up	Observation over a period of time of an individual, group or initially defined population whose appropriate characteristics have been assessed in order to observe changes in health status or health-related variables.
Generalisability	The extent to which the results of a study hold true for groups that did not participate in the research. See also external validity.
Gold standard	A method, procedure or measurement that is widely accepted as being the best available to test for or treat a disease.
GRADE, GRADE profile	A system developed by the GRADE Working Group to address the shortcomings of present grading systems in healthcare. The GRADE system uses a common, sensible and transparent approach to grading the quality of evidence. The results of applying the GRADE system to clinical trial data are displayed in a table known as a GRADE profile.
Harms	Adverse effects of an intervention.
Health economics	Study or analysis of the cost of using and distributing healthcare resources.
Health-related quality of life (HRQoL)	A measure of the effects of an illness to see how it affects someone's day-to-day life.
Heterogeneity or Lack of homogeneity	The term is used in meta-analyses and systematic reviews to describe when the results of a test or treatment (or estimates of its effect) differ significantly in different studies. Such differences may occur as a result of differences in the populations studied, the outcome measures used or because of different definitions of the variables involved. It is the opposite of homogeneity.
Imprecision	Results are imprecise when studies include relatively few patients and few events and thus have wide confidence intervals around the estimate of effect.
Inclusion criteria (literature review)	Explicit criteria used to decide which studies should be considered as potential sources of evidence.
Incremental analysis	The analysis of additional costs and additional clinical outcomes with different interventions.
Incremental cost	The extra cost linked to using one test or treatment rather than another. Or the additional cost of doing a test or providing a treatment more frequently.
Incremental cost-effectiveness ratio (ICER)	The difference in the mean costs in the population of interest divided by the differences in the mean outcomes in the population of interest for one treatment compared with another.
Incremental net benefit (INB)	The value (usually in monetary terms) of an intervention net of its cost compared with a comparator intervention. The INB can be calculated for a given cost-effectiveness (willingness to pay) threshold. If the threshold is £20,000 per QALY gained then the INB is calculated as: (£20,000 × QALYs gained) − Incremental cost.
Indirectness	The available evidence is different to the review question being addressed, in terms of PICO (population, intervention, comparison

Term	Definition
	and outcome).
Intention-to-treat analysis (ITT)	An assessment of the people taking part in a clinical trial, based on the group they were initially (and randomly) allocated to. This is regardless of whether or not they dropped out, fully complied with the treatment or switched to an alternative treatment. Intention-to-treat analyses are often used to assess clinical effectiveness because they mirror actual practice: that is, not everyone complies with treatment and the treatment people receive may be changed according to how they respond to it.
Intervention	In medical terms this could be a drug treatment, surgical procedure, diagnostic or psychological therapy. Examples of public health interventions could include action to help someone to be physically active or to eat a more healthy diet.
Intraoperative	The period of time during a surgical procedure.
Kappa statistic	A statistical measure of inter-rater agreement that takes into account the agreement occurring by chance.
Length of stay	The total number of days a participant stays in hospital.
Licence	See 'Product licence'.
Life years gained	Mean average years of life gained per person as a result of the intervention compared with an alternative intervention.
Likelihood ratio	The likelihood ratio combines information about the sensitivity and specificity. It tells you how much a positive or negative result changes the likelihood that a patient would have the disease. The likelihood ratio of a positive test result (LR+) is sensitivity divided by (1 minus specificity).
Long-term care	Residential care in a home that may include skilled nursing care and help with everyday activities. This includes nursing homes and residential homes.
Logistic regression or Logit model	In statistics, logistic regression is a type of analysis used for predicting the outcome of a binary dependent variable based on one or more predictor variables. It can be used to estimate the log of the odds (known as the 'logit').
Loss to follow-up	A patient, or the proportion of patients, actively participating in a clinical trial at the beginning, but whom the researchers were unable to trace or contact by the point of follow-up in the trial
Markov model	A method for estimating long-term costs and effects for recurrent or chronic conditions, based on health states and the probability of transition between them within a given time period (cycle).
Meta-analysis	A method often used in systematic reviews. Results from several studies of the same test or treatment are combined to estimate the overall effect of the treatment.
Multivariate model	A statistical model for analysis of the relationship between 2 or more predictor (independent) variables and the outcome (dependent) variable.
Negative predictive value (NPV)	In screening or diagnostic tests: A measure of the usefulness of a screening or diagnostic test. It is the proportion of those with a negative test result who do not have the disease, and can be interpreted as the probability that a negative test result is correct. It is calculated as follows: TN/(TN+FN)
Net monetary benefit (NMB)	The value in monetary terms of an intervention net of its cost. The NMB can be calculated for a given cost-effectiveness threshold. If the threshold is £20,000 per QALY gained then the NMB for an intervention is calculated as: (£20,000 × mean QALYs) − mean cost. The most preferable option (that is, the most clinically effective option to have an ICER below the threshold selected) will be the treatment

Term	Definition
	with the highest NMB.
Non-randomised intervention study	A quantitative study investigating the effectiveness of an intervention that does not use randomisation to allocate patients (or units) to treatment groups. Non-randomised studies include observational studies, where allocation to groups occurs through usual treatment decisions or people's preferences. Non-randomised studies can also be experimental, where the investigator has some degree of control over the allocation of treatments. Non-randomised intervention studies can use a number of different study designs, and include cohort studies, case–control studies, controlled before-and-after studies, interrupted-time-series studies and quasi-randomised controlled trials.
Number needed to treat (NNT)	The average number of patients who need to be treated to get a positive outcome. For example, if the NNT is 4, then 4 patients would have to be treated to ensure 1 of them gets better. The closer the NNT is to 1, the better the treatment. For example, if you give a stroke prevention drug to 20 people before 1 stroke is prevented, the number needed to treat is 20. See also number needed to harm, absolute risk reduction.
Observational study	Individuals or groups are observed or certain factors are measured. No attempt is made to affect the outcome. For example, an observational study of a disease or treatment would allow 'nature' or usual medical care to take its course. Changes or differences in one characteristic (for example, whether or not people received a specific treatment or intervention) are studied without intervening. There is a greater risk of selection bias than in experimental studies.
Odds ratio	Odds are a way to represent how likely it is that something will happen (the probability). An odds ratio compares the probability of something in one group with the probability of the same thing in another. An odds ratio of 1 between 2 groups would show that the probability of the event (for example a person developing a disease, or a treatment working) is the same for both. An odds ratio greater than 1 means the event is more likely in the first group. An odds ratio less than 1 means that the event is less likely in the first group. Sometimes probability can be compared across more than 2 groups – in this case, one of the groups is chosen as the 'reference category', and the odds ratio is calculated for each group compared with the reference category. For example, to compare the risk of dying from lung cancer for non-smokers, occasional smokers and regular smokers, non-smokers could be used as the reference category. Odds ratios would be worked out for occasional smokers compared with non-smokers and for regular smokers compared with non-smokers. See also confidence interval, risk ratio.
Opportunity cost	The loss of other healthcare programmes displaced by investment in or introduction of another intervention. This may be best measured by the health benefits that could have been achieved had the money been spent on the next best alternative healthcare intervention.
Outcome	The impact that a test, treatment, policy, programme or other intervention has on a person, group or population. Outcomes from interventions to improve the public's health could include changes in knowledge and behaviour related to health, societal changes (for example, a reduction in crime rates) and a change in people's health and wellbeing or health status. In clinical terms, outcomes could include the number of patients who fully recover from an illness or the number of hospital admissions, and an improvement or deterioration in someone's health, functional ability, symptoms or situation.

Term	Definition
	Researchers should decide what outcomes to measure before a study begins.
P value	The p value is a statistical measure that indicates whether or not an effect is statistically significant. For example, if a study comparing 2 treatments found that one seems more effective than the other, the p value is the probability of obtaining these results by chance. By convention, if the p value is below 0.05 (that is, there is less than a 5% probability that the results occurred by chance) it is considered that there probably is a real difference between treatments. If the p value is 0.001 or less (less than a 1% probability that the results occurred by chance), the result is seen as highly significant. If the p value shows that there is likely to be a difference between treatments, the confidence interval describes how big the difference in effect might be.
Perioperative	The period from admission through surgery until discharge, encompassing the preoperative and postoperative periods.
Placebo	A fake (or dummy) treatment given to participants in the control group of a clinical trial. It is indistinguishable from the actual treatment (which is given to participants in the experimental group). The aim is to determine what effect the experimental treatment has had – over and above any placebo effect caused because someone has received (or thinks they have received) care or attention.
Polypharmacy	The use or prescription of multiple medications.
Posterior distribution	In Bayesian statistics this is the probability distribution for a statistic based after combining established information or belief (the prior) with new evidence (the likelihood).
Positive predictive value (PPV)	In screening or diagnostic tests: A measure of the usefulness of a screening or diagnostic test. It is the proportion of those with a positive test result who have the disease, and can be interpreted as the probability that a positive test result is correct. It is calculated as follows: TP/(TP+FP)
Postoperative	Pertaining to the period after patients leave the operating theatre, following surgery.
Power (statistical)	The ability to demonstrate an association when one exists. Power is related to sample size; the larger the sample size, the greater the power and the lower the risk that a possible association could be missed.
Preoperative	The period before surgery commences.
Prior distribution	In Bayesian statistics this is the probability distribution for a statistic based on previous evidence or belief.
Primary care	Healthcare delivered outside hospitals. It includes a range of services provided by GPs, nurses, health visitors, midwives and other healthcare professionals and allied health professionals such as dentists, pharmacists and opticians.
Primary outcome	The outcome of greatest importance, usually the one in a study that the power calculation is based on.
Probabilistic analysis	In economic evaluation, this is an analysis that uses a probability distribution for each input. In contrast, see Deterministic analysis.
Product licence	An authorisation from the MHRA to market a medicinal product.
Prognosis	A probable course or outcome of a disease. Prognostic factors are patient or disease characteristics that influence the course. Good prognosis is associated with low rate of undesirable outcomes; poor prognosis is associated with a high rate of undesirable outcomes.
Prospective study	A research study in which the health or other characteristic of

Term	Definition
	participants is monitored (or 'followed up') for a period of time, with events recorded as they happen. This contrasts with retrospective studies.
Publication bias	Publication bias occurs when researchers publish the results of studies showing that a treatment works well and don't publish those showing it did not have any effect. If this happens, analysis of the published results will not give an accurate idea of how well the treatment works. This type of bias can be assessed by a funnel plot.
Quality of life	See 'Health-related quality of life'.
Quality-adjusted life year (QALY)	A measure of the state of health of a person or group in which the benefits, in terms of length of life, are adjusted to reflect the quality of life. One QALY is equal to 1 year of life in perfect health. QALYS are calculated by estimating the years of life remaining for a patient following a particular treatment or intervention and weighting each year with a quality of life score (on a scale of 0 to 1). It is often measured in terms of the person's ability to perform the activities of daily life, freedom from pain and mental disturbance.
Randomisation	Assigning participants in a research study to different groups without taking any similarities or differences between them into account. For example, it could involve using a random numbers table or a computer-generated random sequence. It means that each individual (or each group in the case of cluster randomisation) has the same chance of receiving each intervention.
Randomised controlled trial (RCT)	A study in which a number of similar people are randomly assigned to 2 (or more) groups to test a specific drug or treatment. One group (the experimental group) receives the treatment being tested, the other (the comparison or control group) receives an alternative treatment, a dummy treatment (placebo) or no treatment at all. The groups are followed up to see how effective the experimental treatment was. Outcomes are measured at specific times and any difference in response between the groups is assessed statistically. This method is also used to reduce bias.
RCT	See 'Randomised controlled trial'.
Receiver operated characteristic (ROC) curve	A graphical method of assessing the accuracy of a diagnostic test. Sensitivity is plotted against 1 minus specificity. A perfect test will have a positive, vertical linear slope starting at the origin. A good test will be somewhere close to this ideal.
Reference standard	The test that is considered to be the best available method to establish the presence or absence of the outcome – this may not be the one that is routinely used in practice.
Reporting bias	See 'Publication bias'.
Resource implication	The likely impact in terms of finance, workforce or other NHS resources.
Retrospective study	A research study that focuses on the past and present. The study examines past exposure to suspected risk factors for the disease or condition. Unlike prospective studies, it does not cover events that occur after the study group is selected.
Review question	In guideline development, this term refers to the questions about treatment and care that are formulated to guide the development of evidence-based recommendations.
Risk ratio (RR)	The ratio of the risk of disease or death among those exposed to certain conditions compared with the risk for those who are not exposed to the same conditions (for example, the risk of people who smoke getting lung cancer compared with the risk for people who do not smoke). If both groups face the same level of risk, the risk ratio is 1. If the first

Term	Definition
	group had a risk ratio of 2, subjects in that group would be twice as likely to have the event happen. A risk ratio of less than 1 means the outcome is less likely in the first group. The risk ratio is sometimes referred to as relative risk.
Secondary outcome	An outcome used to evaluate additional effects of the intervention deemed a priori as being less important than the primary outcomes.
Selection bias	Selection bias occurs if: a) The characteristics of the people selected for a study differ from the wider population from which they have been drawn, or b) There are differences between groups of participants in a study in terms of how likely they are to get better.
Sensitivity	How well a test detects the thing it is testing for. If a diagnostic test for a disease has high sensitivity, it is likely to pick up all cases of the disease in people who have it (that is, give a 'true positive' result). But if a test is too sensitive it will sometimes also give a positive result in people who don't have the disease (that is, give a 'false positive'). For example, if a test were developed to detect if a woman is 6 months pregnant, a very sensitive test would detect everyone who was 6 months pregnant, but would probably also include those who are 5 and 7 months pregnant. If the same test were more specific (sometimes referred to as having higher specificity), it would detect only those who are 6 months pregnant, and someone who was 5 months pregnant would get a negative result (a 'true negative'). But it would probably also miss some people who were 6 months pregnant (that is, give a 'false negative'). Breast screening is a 'real-life' example. The number of women who are recalled for a second breast screening test is relatively high because the test is very sensitive. If it were made more specific, people who don't have the disease would be less likely to be called back for a second test but more women who have the disease would be missed.
Sensitivity analysis	A means of representing uncertainty in the results of economic evaluations. Uncertainty may arise from missing data, imprecise estimates or methodological controversy. Sensitivity analysis also allows for exploring the generalisability of results to other settings. The analysis is repeated using different assumptions to examine the effect on the results. One-way simple sensitivity analysis (univariate analysis): each parameter is varied individually in order to isolate the consequences of each parameter on the results of the study. Multi-way simple sensitivity analysis (scenario analysis): 2 or more parameters are varied at the same time and the overall effect on the results is evaluated. Threshold sensitivity analysis: the critical value of parameters above or below which the conclusions of the study will change are identified. Probabilistic sensitivity analysis: probability distributions are assigned to the uncertain parameters and are incorporated into evaluation models based on decision analytical techniques (for example, Monte Carlo simulation).
Significance (statistical)	A result is deemed statistically significant if the probability of the result occurring by chance is less than 1 in 20 ($p<0.05$).
Specificity	The proportion of true negatives that are correctly identified as such. For example in diagnostic testing the specificity is the proportion of non-cases correctly diagnosed as non-cases.

Term	Definition
	See related term ‘Sensitivity’. In terms of literature searching a highly specific search is generally narrow and aimed at picking up the key papers in a field and avoiding a wide range of papers.
Stakeholder	An organisation with an interest in a topic that NICE is developing a guideline or piece of public health guidance on. Organisations that register as stakeholders can comment on the draft scope and the draft guidance. Stakeholders may be: • manufacturers of drugs or equipment • national patient and carer organisations • NHS organisations • organisations representing healthcare professionals.
State transition model	See Markov model
Systematic review	A review in which evidence from scientific studies has been identified, appraised and synthesised in a methodical way according to predetermined criteria. It may include a meta-analysis.
Time horizon	The time span over which costs and health outcomes are considered in a decision analysis or economic evaluation.
Transition probability	In a state transition model (Markov model), this is the probability of moving from one health state to another over a specific period of time.
Treatment allocation	Assigning a participant to a particular arm of a trial.
Univariate	Analysis which separately explores each variable in a data set.
Utility	In health economics, a 'utility' is the measure of the preference or value that an individual or society places upon a particular health state. It is generally a number between 0 (representing death) and 1 (perfect health). The most widely used measure of benefit in cost–utility analysis is the quality-adjusted life year, but other measures include disability-adjusted life years (DALYs) and healthy year equivalents (HYEs).

References

1. GRADE Working Group. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group website. 2011. Available from: http://www.gradeworkinggroup.org/ Last accessed: 10/09/2019

2. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. Journal of Clinical Epidemiology. 2016; 75:40-46

3. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual, updated 2018. London. National Institute for Health and Care Excellence, 2014. Available from: https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview

4. National Institute for Health and Clinical Excellence. Social value judgements: principles for the development of NICE guidance. London. National Institute for Health and Clinical Excellence, 2008. Available from: https://www.nice.org.uk/media/default/about/what-we-do/research-and-development/social-value-judgements-principles-for-the-development-of-nice-guidance.pdf

5. Organisation for Economic Co-operation and Development (OECD). Purchasing power parities (PPP). Available from: http://www.oecd.org/sdd/prices-ppp/ Last accessed: 17/06/2019

6. Review Manager (RevMan) [Computer program]. Version 5. Copenhagen. The Nordic Cochrane Centre, The Cochrane Collaboration, 2015. Available from: http://tech.cochrane.org/Revman