National Institute for Health and Care Excellence

Draft for consultation

# Meningitis (bacterial) and meningococcal disease: recognition, diagnosis and management

*NICE guideline number TBC*

*Methods*

*August 2023*

**NICE guideline: methods**

*Draft for Consultation*

*Commissioned by the National Institute for Health and Care Excellence*

**Disclaimer**

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

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

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 All NICE guidance is subject to regular review and may be updated or withdrawn.

**Copyright**



ISBN:

# Contents

# Development of the guideline

## Remit

This guideline will update the following National Institute for Health and Care Excellence (NICE) clinical guideline: Meningitis (bacterial) and meningococcal septicaemia: recognition, diagnosis and management (NICE CG102). The guideline will be extended to cover people aged over 16 years.

## What this guideline covers

### Groups that are covered

- All adults, young people, children and babies (aged 29 days old and over, using corrected age for pre-term babies) with suspected or confirmed bacterial meningitis or meningococcal disease.
- Parents or carers of babies, children and young people who have suspected or confirmed bacterial meningitis or meningococcal disease.

Specific consideration will be given to babies between 29 days and 1 year old.

Babies aged up to 28 days old (using corrected age for pre-term babies) are generally not included in the guideline as the NICE guideline on neonatal infection includes recommendations for this population. However, for some evidence reviews in this guideline, where the review questions were not covered by the neonatal infection guideline, babies aged up to 28 days were included (evidence reviews B3, G1, G4, I1, and J1).

### Settings that are covered

Primary, secondary and tertiary healthcare settings (including the ambulance service, accident and emergency departments, inpatient care and transitions between departments and services). This includes remote contact (for example NHS 111) and face-to-face contact. Community facing services such as community child health will be included where relevant.

### Key areas that are covered

- Recognising suspected bacterial meningitis and meningococcal disease, including 'safety netting'
- Investigations used in cases of suspected bacterial meningitis and meningococcal disease to support diagnosis
- Antibiotics for bacterial meningitis and meningococcal disease
- Non-antibiotic management of bacterial meningitis
- Non-antibiotic management of meningococcal disease
- Long-term complications and follow-up for bacterial meningitis and meningococcal disease
- Further investigation

- Information and support

# What this guideline does not cover

**Groups that are not covered**

- People:
  - with known immunodeficiency.
  - who have brain tumours, pre-existing hydrocephalus, intracranial shunts, previous neurosurgical procedures, or known cranial or spinal anomalies that increase the risk of bacterial meningitis.
  - with confirmed viral meningitis or viral encephalitis.
  - with confirmed tuberculous meningitis.
  - with confirmed fungal meningitis.

# Methods

This guideline was developed using the methods described in the 2018 NICE guidelines manual.

Declarations of interest were recorded according to the NICE conflicts of interest policy.

# Developing the review questions and outcomes

The review questions developed for this guideline were based on the key areas identified in the guideline scope. They were drafted by the technical team and refined and validated by the guideline committee.

The review questions were based on the following frameworks:

- population, intervention, comparator and outcome (PICO) for reviews of interventions
- diagnostic reviews and reviews of prediction model accuracy – using population, diagnostic test (index test), reference standard and target condition
- prognostic reviews – using population, presence or absence of a prognostic, risk or predictive factor and outcome
- qualitative reviews – using population, phenomenon of interest and context

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

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

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

| Evidence review | Review question | Type of review |
|---|---|---|
| [A1] Symptoms and signs associated with bacterial meningitis | What symptoms and signs, individually or in combination, are associated with bacterial meningitis? | Diagnostic |
| [A2] Risk factors associated with bacterial meningitis | What factors are associated with an increased risk of bacterial meningitis? | Prognostic |
| [A3] Symptoms and signs associated with meningococcal disease | What symptoms and signs, individually or in combination, are associated with meningococcal disease? | Diagnostic |
| [A4] Risk factors associated with meningococcal disease | What factors are associated with an increased risk of meningococcal disease? | Prognostic |
| [B1] Investigating and diagnosing suspected bacterial meningitis with blood and urine investigations | What is the accuracy and effectiveness of blood and urine investigations in diagnosing bacterial meningitis? | Diagnostic |
| [B2] Investigating and diagnosing suspected meningococcal disease with blood and urine investigations | What is the accuracy and effectiveness of blood and urine investigations in diagnosing meningococcal disease? | Diagnostic |
| [B3] Investigating and diagnosing suspected bacterial meningitis with cerebrospinal fluid parameters | What is the accuracy and effectiveness of cerebrospinal fluid investigations in diagnosing bacterial meningitis? | Diagnostic |
| [B4] Factors associated with brain herniation | What factors (individually or in combination) are associated with an increased risk of brain herniation following lumbar | Prognostic |

| Evidence review | Review question | Type of review |
|---|---|---|
| | puncture in people with suspected bacterial meningitis? | |
| [B5] Role of neuroimaging prior to lumbar puncture | What is the role of neuroimaging prior to lumbar puncture? | Intervention |
| [C1] Timing of antibiotics for bacterial meningitis | What is the optimal timing of antibiotic administration for people with suspected bacterial meningitis? | Intervention |
| [C2] Timing of antibiotics for meningococcal disease | What is the optimal timing of antibiotic administration for people with suspected meningococcal disease? | Intervention |
| [D1] Antibiotics for bacterial meningitis before or in the absence of identifying causative infecting organism in younger infants | What antibiotic treatment regimens are effective in treating suspected bacterial meningitis in younger infants (excluding neonates) before identifying the causative infecting organism, or in the absence of identifying the causative infecting organism? | Intervention |
| [D2] Antibiotics for bacterial meningitis before or in the absence of identifying causative infecting organism in older infants and children | What antibiotic treatment regimens are effective in treating suspected bacterial meningitis in older infants and children before identifying the causative infecting organism, or in the absence of identifying the causative infecting organism? | Intervention |
| [D3] Antibiotics for bacterial meningitis before or in the absence of identifying causative infecting organism in adults | What antibiotic treatment regimens are effective in treating suspected bacterial meningitis in adults before identifying the causative infecting organism, or in the absence of identifying the causative infecting organism? | Intervention |
| [E1] Antibiotics for bacterial meningitis caused by Streptococcus pneumoniae | What antibiotic treatment regimens are effective in treating bacterial meningitis caused by Streptococcus pneumoniae? | Intervention |

| Evidence review | Review question | Type of review |
|---|---|---|
| [E2] Antibiotics for bacterial meningitis caused by Haemophilus influenzae | What antibiotic treatment regimens are effective in treating bacterial meningitis caused by Haemophilus influenzae? | Intervention |
| [E3] Antibiotics for bacterial meningitis caused by Group B streptococcus | What antibiotic treatment regimens are effective in treating bacterial meningitis caused by Group B streptococcus? | Intervention |
| [E4] Antibiotics for bacterial meningitis caused by Gram-negative bacilli | What antibiotic treatment regimens are effective in treating bacterial meningitis caused by Gram-negative bacilli? | Intervention |
| [E5] Antibiotics for bacterial meningitis caused by Listeria monocytogenes | What antibiotic treatment regimens are effective in treating bacterial meningitis caused by Listeria monocytogenes? | Intervention |
| [E6] Antibiotics for bacterial meningitis caused by Neisseria meningitidis | What antibiotic treatment regimens are effective in treating bacterial meningitis caused by Neisseria meningitidis? | Intervention |
| [F1] Antibiotics for meningococcal disease | What antibiotic treatment regimens are effective in treating suspected or confirmed meningococcal disease? | Intervention |
| [G1] Fluid restriction in bacterial meningitis | What is the effectiveness of fluid restriction in bacterial meningitis? | Intervention |
| [G2] Osmotic agents for bacterial meningitis | What is the effectiveness of osmotic agents in bacterial meningitis? | Intervention |
| [G3] Intracranial pressure monitoring in bacterial meningitis | What is the effectiveness of intracranial pressure monitoring in bacterial meningitis? | Intervention |

| Evidence review | Review question | Type of review |
|---|---|---|
| [G4] Corticosteroids for treatment of bacterial meningitis | What is the effectiveness of corticosteroid treatment in bacterial meningitis? | Intervention |
| [H] Corticosteroids in meningococcal disease | What is the effectiveness of corticosteroid treatment in meningococcal disease? | Intervention |
| [I1] Long-term complications and follow-up for bacterial meningitis | What is the risk of long-term complications in bacterial meningitis? | Prognostic |
| [I2] Long-term complications and follow-up for meningococcal disease | What is the risk of long-term complications in meningococcal disease? | Prognostic |
| [J1] Factors associated with an increased risk of recurrent bacterial meningitis | What factors (individually or in combination) are associated with an increased risk of recurrent bacterial meningitis? | Prognostic |
| [J2] Factors associated with an increased risk of recurrent meningococcal disease | What factors (individually or in combination) are associated with an increased risk of recurrent meningococcal disease? | Prognostic |
| [K1] What information is valued by patients and their families or carers, when concerns arise about the possibility of bacterial meningitis or meningococcal disease? | What information is valued by patients and their families or carers, when concerns arise about the possibility of bacterial meningitis or meningococcal disease? | Qualitative |
| [K2] Support for suspected bacterial meningitis or | What support is valued by patients and their families or carers, when concerns arise about the possibility of bacterial | Qualitative |

| Evidence review | Review question | Type of review |
|---|---|---|
| meningococcal disease | meningitis or meningococcal disease? | |
| [K3] Information for confirmed bacterial meningitis or meningococcal disease | What information is valued by patients with confirmed bacterial meningitis or meningococcal disease, and their families or carers? | Qualitative |
| [K4] Support for confirmed bacterial meningitis or meningococcal disease | What support is valued by patients with confirmed bacterial meningitis or meningococcal disease, and their families or carers? | Qualitative |

The COMET database was searched for core outcome sets relevant to this guideline. A core outcome set, including death and neurological sequelae as core clinical outcomes, was identified for bacterial meningitis, but no core outcome sets were identified for meningococcal disease. Additional outcomes were chosen based on committee discussions.

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

- Supplement 2 (Meningitis NICE technical team list).

# Searching for evidence

## Scoping search

During the scoping phase, searches were conducted for previous guidelines, economic evaluations, health technology assessments, systematic reviews, and randomised controlled trials.

## Systematic literature search

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

Databases were searched using subject headings, free-text terms and, where appropriate, study type filters. Where possible, searches were limited to retrieve studies published in English. Limits to exclude animal studies, letters, editorials, news and conferences were applied where possible. All the searches were conducted in the following databases: Medline, Medline-in-Process, Cochrane Central Register of Controlled Trials (CCTR), Cochrane Database of Systematic Reviews (CDSR), and Embase. For review questions related to information and support, Emcare and PsycINFO were also searched.

Searches were run once for all reviews during development. Searches for the following questions were updated in November 2022.

- [A1] Symptoms and signs associated with bacterial meningitis
- [A2] Risk factors associated with bacterial meningitis
- [A3] Symptoms and signs associated with meningococcal disease
- [A4] Risk factors associated with meningococcal disease
- [B3] Investigating and diagnosing suspected bacterial meningitis with cerebrospinal fluid parameters
- [B5] Role of neuroimaging prior to lumbar puncture
- [D1] Antibiotics for bacterial meningitis before or in the absence of identifying causative infecting organism in younger infants
- [D2] Antibiotics for bacterial meningitis before or in the absence of identifying causative infecting organism in older infants and children
- [D3] Antibiotics for bacterial meningitis before or in the absence of identifying causative infecting organism in adults
- [E1] Antibiotics for bacterial meningitis caused by Streptococcus pneumoniae
- [E2] Antibiotics for bacterial meningitis caused by Haemophilus influenzae
- [E3] Antibiotics for bacterial meningitis caused by Group B streptococcus
- [E4] Antibiotics for bacterial meningitis caused by Gram-negative bacilli
- [E5] Antibiotics for bacterial meningitis caused by Listeria monocytogenes
- [E6] Antibiotics for bacterial meningitis caused by Neisseria meningitidis
- [F1] Antibiotics for meningococcal disease
- [G1] Fluid restriction in bacterial meningitis
- [G2] Osmotic agents for bacterial meningitis
- [G4] Corticosteroids for Bacterial Meningitis
- [H] Corticosteroids in meningococcal disease

Details of the search strategies, including the study-design filters used and databases searched, are provided in Appendix B of each evidence review.

## Economic systematic literature search

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

A single search, using the population search terms used in the evidence reviews, was conducted to identify economic evidence in the NHS Economic Evaluation Database (NHS EED) and HTA. Another single search, using the population search terms used in the evidence reviews combined with an economic evaluations search filter, was conducted in Medline, Medline in Process and Embase. Where possible, searches were limited to studies published in English. Limits to exclude animal studies, letters, editorials, news were applied where possible.

As with the general literature searches, the economic literature searches were run once during development, and updated in November 2022.

Details of the search strategies, including the study-design filters used and databases searched, are provided in Appendix B of each evidence review.

### Quality assurance

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

## Reviewing research evidence

### Systematic review process

The evidence was reviewed in accordance with the following approach.

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

Review questions selected as high priorities for economic analysis (and those selected as medium priorities and where economic analysis could influence recommendations) and complex review questions were subject to dual screening and study selection through a 5% random sample of articles. Any discrepancies were resolved by discussion between the first and second reviewers or by reference to a third (senior) reviewer. For the remaining review questions, quality assurance processes included consideration of the outcomes of screening, study selection and data extraction and the committee reviewed the results of study selection and data extraction. The review protocol for each question specifies whether dual screening and study selection was undertaken for that particular question. Drafts of all evidence reviews were quality assured by a senior reviewer.

### Type of studies and inclusion/exclusion criteria

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

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

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

For diagnostic or prediction rule reviews, test-and-treat RCTs were prioritised for inclusion. In the absence of such studies, test accuracy studies were considered for inclusion. Single-gate studies were prioritised.

For prognostic reviews, prospective and retrospective cohort studies were considered for inclusion. Studies that included multivariate analysis were prioritised.

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

The committee was consulted about any uncertainty regarding inclusion or exclusion of studies. A list of excluded studies for each review question, including reasons for exclusion is presented in Appendix J of the corresponding evidence review.

Narrative reviews, posters, letters, editorials, comment articles, unpublished studies and studies published in languages other than English were excluded. Conference abstracts were not considered for inclusion because conference abstracts typically do not have sufficient information to allow for full critical appraisal.

# Methods of combining evidence

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

## Data synthesis for intervention studies

### *Pairwise meta-analysis*

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

For dichotomous outcomes, such as mortality, the Mantel–Haenszel method with a fixed effect model was used to calculate risk ratios (RRs). When there was only data from 1 study and the event rate was less than 1% in 1 arm and more than 1% in 1 arm, Peto odds ratio (POR) was used if the combined event rate was less than 1% (when more than 1 study, the decision to use POR was based on whether the event rate was less than 1% in most arms across studies). The POR method performs well when events are rare (Bradburn 2007).

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

outcomes, such as quality of life, were meta-analysed using an inverse-variance method for pooling weighted mean differences (WMDs). Where SDs were not reported, these were calculated from other reported statistics where possible (standard errors or 95% confidence intervals; CIs) and then meta-analysis was conducted as described above.

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

For some reviews, evidence was either stratified from the outset or separated into subgroups when heterogeneity was encountered. The stratifications and potential subgroups were pre-defined at the protocol stage (see the protocols for each review for further detail). Where evidence was stratified or subgrouped the committee considered on a case by case basis if separate recommendations should be made for distinct groups. Separate recommendations may be made where there is evidence of a differential effect of interventions in distinct groups. If there is a lack of evidence in 1 group, the committee considered, based on their experience, whether it was reasonable to extrapolate and assume the interventions will have similar effects in that group compared with others

When meta-analysis was undertaken, the results were presented visually using forest plots generated with RevMan5 (see Appendix E of relevant evidence reviews).

## Data synthesis for diagnostic test accuracy reviews

When diagnostic test accuracy was measured dichotomously, sensitivity and specificity were used as outcomes. When diagnostic test accuracy was measured continuously, the area under the receiver-operating characteristic (ROC) curve (AUC) was used. These diagnostic test accuracy parameters were obtained directly from results reported in the source articles or calculated by the technical team using data reported in the articles.

Meta-analysis of diagnostic test accuracy parameters was not conducted for this guideline due to insufficient evidence after stratifications (for example, after stratifying for age, index test threshold, bacterial pathogen and reference standard used), or where there was sufficient evidence a high level of heterogeneity remained between studies in terms of study design, population and prevalence of bacterial meningitis.

## Data synthesis for prognostic reviews

ORs or RRs with 95% CIs reported in published studies were extracted or calculated by the technical team to examine relationships between risk factors and outcomes of interest. Adjusted estimates from multivariate analyses were prioritised where available.

Where multiple studies reported on the same factor and the definitions used and approach to analysis in the primary papers was sufficiently consistent, meta-analyses were conducted and the results were presented visually using forest plots generated with RevMan5 (see Appendix E of relevant evidence reviews).

### Data synthesis for qualitative reviews

Where possible, a meta-synthesis was conducted to combine evidence from qualitative studies. Whenever studies identified a qualitative theme relevant to the protocol, this was extracted, and the main characteristics were summarised. When all themes had been extracted from studies, common concepts were categorised and tabulated. This included information on how many studies had contributed to each theme identified by the technical team.

Themes from individual studies were integrated into a wider context and, when possible, overarching categories of themes with sub-themes were identified. Themes were derived from data presented in individual studies and theme names were assigned by the technical team.

Emerging themes were placed into a thematic map representing the relationship between themes and overarching categories. The purpose of such a map is to show relationships between overarching categories and associated themes.

# Appraising the quality of evidence

### Intervention studies

#### *Pairwise meta-analysis*

##### GRADE methodology for intervention reviews

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

When GRADE was applied, software developed by the GRADE working group (GRADEpro) was used to assess the quality of each outcome, taking account of individual study quality factors and any meta-analysis results. Results were presented in GRADE profiles (GRADE tables).

The selection of outcomes for each review question was agreed during development of the associated review protocol in discussion with the committee. The evidence for each outcome was examined separately for the quality elements summarised in Table 2. Criteria considered in the rating of these elements are discussed below. Each element was graded using the quality ratings summarised in Table 3. Footnotes to GRADE tables were used to record reasons for grading a particular quality element as having a ‘serious’ or ‘very serious’ quality issue. The ratings for each component were combined to obtain an overall assessment of quality for each outcome as described in Table 4.

The initial quality rating was based on the study design: RCTs and NRS assessed by ROBINS-I start as ‘high’ quality evidence, other non-randomised studies start as ‘low’ quality evidence. The rating was then modified according to the assessment of each quality element (Table 2). Each quality element considered to have a ‘serious’ or ‘very serious’ quality issue was downgraded by 1 or 2 levels, respectively (for example, evidence starting as ‘high’ quality was downgraded to ‘moderate’ or ‘low’

quality). In addition, there was a possibility to upgrade evidence from non-randomised studies (provided the evidence for that outcome had not previously been downgraded) if there was a large magnitude of effect, a dose–response gradient, or if all plausible confounding would reduce a demonstrated effect or suggest a spurious effect when results showed no effect.

**Table 2: Summary of quality elements in GRADE for intervention reviews**

| Quality element | Description |
|---|---|
| Risk of bias ('Study limitations') | This refers to limitations in study design or implementation that reduce the internal validity of the evidence |
| Inconsistency | This refers to unexplained heterogeneity in the results |
| Indirectness | This refers to differences in study populations, interventions, comparators or outcomes between the available evidence and inclusion criteria specified in the review protocol |
| Imprecision | This occurs when a study has few participants or few events of interest, resulting in wide confidence intervals that cross minimally important thresholds |
| Publication bias | This refers to systematic under- or over-estimation of the underlying benefit or harm resulting from selective publication of study results |

**Table 3: GRADE quality ratings (by quality element)**

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

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

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

*Assessing risk of bias in intervention reviews*

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

Risk of bias in RCTs was assessed using the Cochrane risk of bias 2.0 tool (see Appendix H in Developing NICE guidelines: the manual).

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

- randomisation process
- deviations from the intended interventions
- missing outcome data
- measurement of the outcome
- selection of the reported result.

A study with a poor methodological design does not automatically imply high risk of bias; the bias is considered individually for each outcome and it is assessed whether the chosen design and methodology will impact on the estimation of the intervention effect.

More details about the Cochrane risk of bias 2.0 tool can be found in Section 8 of the Cochrane Handbook for Systematic Reviews of Interventions (Higgins 2020).

For systematic reviews of RCTs the AMSTAR checklist was used and for systematic reviews of other study types the ROBIS checklist was used (see Appendix H in Developing NICE guidelines: the manual).

For non-randomised studies the ROBINS-I checklist was used (see Appendix H in Developing NICE guidelines: the manual).

*Assessing inconsistency in intervention reviews*

Inconsistency refers to unexplained heterogeneity in results of meta-analysis. When estimates of treatment effect vary widely across studies (that is, there is heterogeneity or variability in results), this suggests true differences in underlying effects. Inconsistency is, thus, only truly applicable when statistical meta-analysis is conducted (that is, results from different studies are pooled). When outcomes were derived from a single study the rating ‘no serious inconsistency’ was used when assessing this domain, as per GRADE methodology (Santesso 2016).

Inconsistency was assessed visually by inspecting forest plots and observing whether there was considerable heterogeneity in the results of the meta-analysis (for example if the point estimates of the individual studies consistently showed benefits or harms). This was supported by calculating the I-squared statistic for the meta-analysis with an I-squared value of more than 50% indicating serious heterogeneity, and more than 80% indicating very serious heterogeneity.

When serious or very serious heterogeneity was observed, subgroup analyses were performed as pre-specified in the review protocol where possible. If heterogeneity was serious and could not be accounted for by sub-group analyses the meta-analysis was re-run using the Der-Simonian and Laird method with a random effects model. If heterogeneity was very serious and could not be accounted for by sub-group analyses the data was not pooled.

### *Assessing indirectness in intervention reviews*

Directness refers to the extent to which populations, interventions, comparisons and outcomes reported in the evidence are similar to those defined in the inclusion criteria for the review and was assessed by comparing the PICO elements in the studies to the PICO defined in the review protocol. Indirectness is important when such differences are expected to contribute to a difference in effect size or may affect the balance of benefits and harms considered for an intervention.

### *Assessing imprecision and importance in intervention reviews*

Imprecision in GRADE methodology refers to uncertainty around the effect estimate and whether or not there is an important difference between interventions (that is, whether the evidence clearly supports a particular recommendation or appears to be consistent with several candidate recommendations). Therefore, imprecision differs from other aspects of evidence quality because it is not concerned with whether the point estimate is accurate or correct (has internal or external validity). Instead, it is concerned with uncertainty about what the point estimate represents. This uncertainty is reflected in the width of the CI.

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

Imprecision was assessed in the guideline evidence reviews by considering whether the width of the 95% CI of the effect estimate was relevant to decision making, considering each outcome independently. This is illustrated in Figure 1, which considers a positive outcome for the comparison of two treatments. Three decision-making zones can be differentiated, bounded by the thresholds for minimal importance (minimally important differences; MIDs) for benefit and harm.

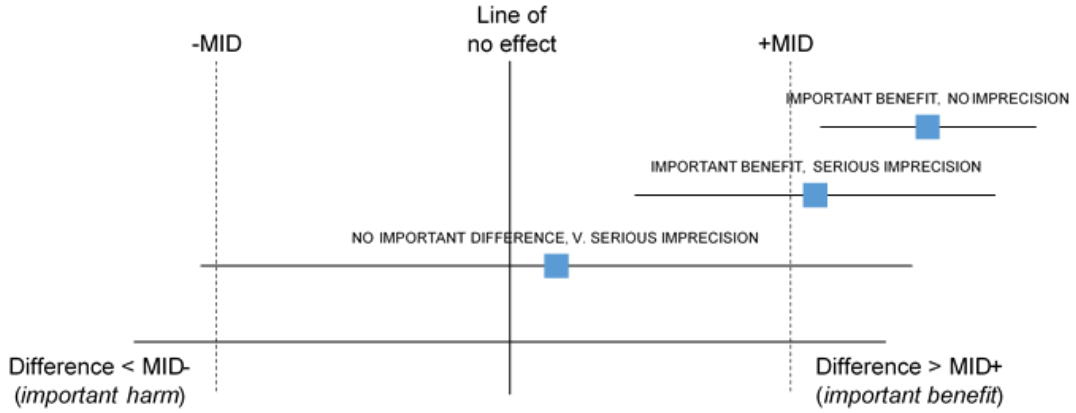
When the CI of the effect estimate is wholly contained in 1 of the 3 zones there is no uncertainty about the size and direction of effect, therefore, the effect estimate is considered precise; that is, there is no imprecision.

When the CI crosses 2 zones, it is uncertain in which zone the true value of the effect estimate lies and therefore there is uncertainty over which decision to make. The CI is consistent with 2 possible decisions, therefore, the effect estimate is considered to be imprecise in the GRADE analysis and the evidence is downgraded by 1 level (‘serious imprecision’).

When the CI crosses all 3 zones, the effect estimate is considered to be very imprecise because the CI is consistent with 3 possible decisions and there is therefore a considerable lack of confidence in the results. The evidence is therefore downgraded by 2 levels in the GRADE analysis (‘very serious imprecision’).

Implicitly, assessing whether a CI is in, or partially in, an important zone, requires the guideline committee to estimate an MID or to say whether they would make different decisions for the 2 confidence limits.

**Figure 1: Assessment of imprecision and importance in intervention reviews using GRADE**




*MID, minimally important difference*

### *Defining minimally important differences for intervention reviews*

The committee was asked whether there were any recognised or acceptable MIDs in the published literature and community relevant to the review questions under consideration. The committee agreed that there were a number of outcomes, namely all-cause mortality, brain herniation and serious intervention-related adverse effects leading to death, disability or prolonged hospitalisation, that were sufficiently serious that any statistically significant difference would be considered important. Further, they agreed that 1 day and 5mmHg could be considered accepted MIDs for length of hospitalisation and change in intracranial pressure, respectively.

For the remaining outcomes, in the absence of published or accepted MIDs, the committee agreed to use the GRADE default MIDs to assess imprecision. For dichotomous outcomes, minimally important thresholds for a RR of 0.8 and 1.25 respectively were used as default MIDs in the guideline. The committee also chose to use 0.8 and 1.25 as the MIDs for ORs & HRs in the absence of published or accepted MIDs. ORs were predominantly used in the guideline when Peto OR were indicated due to low event rates, at low event rates OR are mathematically similar to RR making the extrapolation appropriate. While no default MIDs exist for HR, the committee agreed for consistency to continue to use 0.8 and 1.25 for these outcomes.

If risk difference was used for meta-analysis, for example if the majority of studies had zero events in either arm, imprecision was assessed based on sample size using 200 and 400 as cut-offs for very serious and serious imprecision respectively. The committee used these numbers based on commonly used optimal information size thresholds.

For continuous outcomes default MIDs were used of half the standard deviation (SD) of the control groups at baseline (or at follow-up if the SD is not available at baseline).

MIDs, the line of no effect, and both 95% and 90% confidence intervals (CIs) were used to assess whether there were important differences in outcomes between

groups. Outcomes were considered to have an important benefit/harm, possible important benefit/harm, no evidence of an important difference, or no important difference using the following approach:

- Where the point estimate (PE) was greater than the upper MID and the 95% CI did not cross line of no effect, an intervention was described as having an important benefit
- Where the PE was greater than the upper MID and the 95% CI crossed the line of no effect, but the 90% CI did not, an intervention was described as having a possible important benefit
- Where the PE was greater than the upper MID **or** lower than the lower MID, and the 90% CI crossed the line of no effect, the result was described as no evidence of an important difference
- Where the PE was between two MIDs, the result was described as no important difference
- Where the PE was lower than the lower MID and the 95% CI crossed the line of no effect, but the 90% CI did not, an intervention was described as having a possible important harm
- Where the PE was lower than the lower MID and the 95% CI did not cross line of no effect, an intervention was described as having an important harm.

This approach was used for all evidence reviews which informed decision making on the guideline. Please note that the above descriptions were based on positive outcomes (where high values indicate better outcomes or events are positive). If the outcomes were negative (where high values indicate worse outcomes or events are negative) then whether an intervention is considered to have an important benefit or important harm would be switched (for example, where the PE is greater than the upper MID and the 95% CI do not cross the line of no effect, an intervention would be described as having an important harm; where the PE is lower than the lower MID and the 95% CI do not cross line of no effect, an intervention would be described as having an important benefit).

90% CIs are reported in the summary of the evidence section of the evidence reviews only when they were used to determine a possible important difference (that is, when interventions had a possible important benefit/harm).

### *Assessing publication bias in intervention reviews*

Where 10 or more studies were included as part of a single meta-analysis, a funnel plot was produced to graphically assess the potential for publication bias. Where fewer than 10 studies were included for an outcome, the committee subjectively assessed the likelihood of publication bias based on factors such as the proportion of trials funded by industry and the propensity for publication bias in the topic area.

## Prognostic studies

### *Adapted GRADE methodology for prognostic reviews*

For prognostic reviews with evidence from comparative studies an adapted GRADE approach was used. As noted above, GRADE methodology is designed for intervention reviews but the quality assessment elements were adapted for prognostic reviews.

The evidence for each outcome in the prognostic reviews was examined separately for the quality elements listed and defined in Table 5. The criteria considered in the rating of these elements are discussed below. Each element was graded using the quality levels summarised in Table 3. Footnotes to GRADE tables were used to record reasons for grading a particular quality element as having 'serious' or 'very serious' quality issues. The ratings for each component were combined to obtain an overall assessment of quality for each outcome as described in Table 4.

**Table 5: Adaptation of GRADE quality elements for prognostic reviews**

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

*RR, relative risk; OR, odds ratio*

### *Assessing risk of bias in prognostic reviews*

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

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

### *Assessing inconsistency in prognostic reviews*

Where multiple results were deemed appropriate to meta-analyse (that is, there was sufficient similarity between risk factor and outcome under investigation) inconsistency was assessed by visually inspecting forest plots and observing

whether there was considerable heterogeneity in the results of the meta-analysis. This was assessed by calculating the I-squared statistic for the meta-analysis with an I-squared value of more than 50% indicating serious heterogeneity, and more than 80% indicating very serious heterogeneity. When serious or very serious heterogeneity was observed, subgroup analyses were performed as pre-specified in the review protocol where possible.

When no plausible explanation for the heterogeneity could be found, data were not pooled.

*Assessing indirectness in prognostic reviews*

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

*Assessing imprecision and importance in prognostic reviews*

Prognostic studies may have a variety of purposes, for example, establishing typical prognosis in a broad population, establishing the effect of patient characteristics on prognosis, and developing a prognostic model. While by convention MIDs relate to intervention effects, the committee agreed to use GRADE default MIDs for risk ratios as a starting point from which to assess whether the size of an outcome effect in a prognostic study would be large enough to be meaningful in practice. Specifically, the committee agreed that these values would correspond to a moderate association between the prognostic factor and the outcome, with any statistically significant association being considered a small association, and risk ratios <0.5 and >2.00 being considered a strong association between the factor and the outcome. The latter threshold was selected for consistency with estimated effect sizes where it is possible to consider upgrading non-RCT evidence in GRADE.

## Diagnostic studies

### *Adapted GRADE methodology for diagnostic reviews and prediction models*

For diagnostic reviews and prediction models, an adapted GRADE approach was used. GRADE methodology is designed for intervention reviews but the quality assessment elements and outcome presentation were adapted by the guideline developers for diagnostic test accuracy reviews and prediction models. For example, GRADE tables were modified to include diagnostic test accuracy measures (sensitivity, specificity, and AUC values).

The evidence for each outcome in the diagnostic reviews and prediction models was examined separately for the quality elements listed and defined in Table 6. The criteria considered in the rating of these elements are discussed below. Each element was graded using the quality levels summarised in Table 3. Footnotes to GRADE tables were used to record reasons for grading a particular quality element as having a ‘serious’ or ‘very serious’ quality issue. The ratings for each component were combined to obtain an overall assessment of quality for each outcome as described in Table 4.

The initial quality rating was based on the study design: cross-sectional or cohort studies start as ‘high’ quality and case–control studies start as ‘low’ quality.

**Table 6: Adaptation of GRADE quality elements for diagnostic reviews**

| Quality element | Description |
|---|---|
| Risk of bias ('Study limitations') | Limitations in study design and implementation may bias estimates of diagnostic accuracy. High risk of bias for the majority of the evidence reduces confidence in the estimated effect. Diagnostic accuracy studies are not usually randomised and therefore would not be downgraded for study design from the outset (they start as high quality) |
| Inconsistency | This refers to unexplained heterogeneity in test accuracy measures (such as sensitivity and specificity) between studies |
| Indirectness | This refers to differences in study populations, index tests, reference standards or outcomes between the available evidence and inclusion criteria specified in the review protocol |
| Imprecision | This occurs when a study has relatively few participants and the probability of a correct diagnosis is low. Accuracy measures would therefore have wide confidence intervals around the estimated effect |

*Assessing risk of bias in diagnostic reviews and prediction models*

Risk of bias in diagnostic reviews and prediction models was assessed using the Quality Assessment of Diagnostic Accuracy Studies version 2 (QUADAS-2) checklist (see Appendix H in Developing NICE guidelines: the manual).

Risk of bias in primary diagnostic accuracy reviews or prediction models in QUADAS-2 consists of 4 domains:

- participant selection
- index test
- reference standard
- flow and timing.

More details about the QUADAS-2 tool can be found on the developer's website.

*Assessing inconsistency in diagnostic reviews and prediction models*

Inconsistency refers to the unexplained heterogeneity of the results in meta-analysis. When estimates of diagnostic accuracy and prediction model parameters vary widely across studies (that is, there is heterogeneity or variability in results), this suggests true differences in underlying effects. Inconsistency is, thus, only truly applicable when statistical meta-analysis is conducted (that is, results from different studies are pooled).

Inconsistency for diagnostic reviews and prediction models was assessed based on visual inspection of the point estimates and confidence intervals of the included studies. If these varied widely (for example, point estimates for some studies lying outside the CIs of other studies) the evidence was downgraded for inconsistency.

*Assessing indirectness in diagnostic reviews and prediction models*

Indirectness in diagnostic reviews and prediction models was assessed using the QUADAS-2 checklist by assessing the applicability of the studies in relation to the review question in the following domains:

- participant selection
- index test
- reference standard.

More details about the QUADAS-2 tool can be found on the developer's website.

*Assessing imprecision and importance in diagnostic reviews and prediction models*

The judgement of precision for diagnostic and prediction model evidence was based on the CIs of sensitivity and specificity. The committee defined 3 decision thresholds for each measure, a value below which the test would be considered of no use, a value above which the test could be considered moderately useful, and a value above which the test would be considered very useful. These thresholds were based on the committee's experience and consensus.

The thresholds were:

- sensitivity: not a useful test <50%, moderately useful test ≥50% and <90%, very useful ≥90%
- specificity: not a useful test <50%, moderately useful test >50% and <90%, very useful ≥90%

The following cut-offs were used when summarising the performance of diagnostic tests or prediction models in terms of AUC:

- very useful >0.80
- moderately useful test >0.70 and ≤0.80
- not a useful test ≤0.70.

## Qualitative studies

### *GRADE-CERQual methodology for qualitative reviews*

For qualitative reviews an adapted GRADE Confidence in the Evidence from Reviews of Qualitative research (GRADE-CERQual) approach (Lewin 2015) was used. In this approach the quality of evidence is considered according to themes in the evidence. The themes may have been identified in the primary studies or they may have been identified by considering the reports of a number of studies. Quality elements assessed using GRADE-CERQual are listed and defined in Table 7. Each element was graded using the levels of concern summarised in Table 8. The ratings for each component were combined (as with other types of evidence) to obtain an overall assessment of quality for each theme as described in Table 9.

**Table 7: Adaptation of GRADE quality elements for qualitative reviews**

| Quality element | Description |
|---|---|
| Risk of bias ('Methodological limitations') | Limitations in study design and implementation may bias interpretation of qualitative themes identified. High risk of bias for the majority of the evidence reduces confidence in review findings. Qualitative studies are not usually randomised and therefore would not be downgraded for study design from the outset (they start as high quality) |

| Quality element | Description |
|---|---|
| Relevance (or applicability) of evidence | This refers to the extent to which the evidence supporting the review findings is applicable to the context specified in the review question |
| Coherence of findings | This refers to the extent to which review findings are well grounded in data from the contributing primary studies and provide a credible explanation for patterns identified in the evidence |
| Adequacy of data (theme saturation or sufficiency) | This corresponds to a similar concept in primary qualitative research, that is, whether a theoretical point of theme saturation was achieved, at which point no further citations or observations would provide more insight or suggest a different interpretation of the particular theme. Individual studies that may have contributed to a theme or sub-theme may have been conducted in a manner that by design would have not reached theoretical saturation at an individual study level |

**Table 8: CERQual levels of concern (by quality element)**

| Level of concern | Definition |
|---|---|
| None or very minor concerns | Unlikely to reduce confidence in the review finding |
| Minor concerns | May reduce confidence in the review finding |
| Moderate concerns | Will probably reduce confidence in the review finding |
| Serious concerns | Very likely to reduce confidence in the review finding |

**Table 9: Overall confidence in the evidence in CERQual (by review finding)**

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

*Assessing methodological limitations in qualitative reviews*

Methodological limitations in qualitative studies were assessed using the Critical Appraisal Skills Programme (CASP) checklist for qualitative studies (see appendix H in Developing NICE guidelines: the manual). Overall methodological limitations were derived by assessing the methodological limitations across the 6 domains summarised in Table 10.

**Table 10: Methodological limitations in qualitative studies**

| | |
|---|---|
| Aim and appropriateness of qualitative evidence | This domain assesses whether the aims and relevance of the study were described clearly and whether qualitative research methods were appropriate for investigating the research question |
| Rigour in study design or validity of theoretical approach | This domain assesses whether the study approach was documented clearly and whether it was based on a theoretical framework (such as ethnography or grounded theory). This does not necessarily mean that the framework has to be stated explicitly, but a detailed description ensuring transparency and reproducibility should be provided |
| Sample selection | This domain assesses the background, the procedure and reasons for the method of selecting participants. The assessment should include consideration of any relationship between the researcher and the participants, and how this might have influenced the findings |
| Data collection | This domain assesses the documentation of the method of data collection (in-depth interviews, semi-structured interviews, focus groups or observations). It also assesses who conducted any interviews, how long they lasted and where they took place |
| Data analysis | This domain assesses whether sufficient detail was documented for the analytical process and whether it was in accordance with the theoretical approach. For example, if a thematic analysis was used, the assessment would focus on the description of the approach used to generate themes. Consideration of data saturation would also form part of this assessment (it could be reported directly or it might be inferred from the citations documented that more themes could be found) |
| Results | This domain assesses any reasoning accompanying reporting of results (for example, whether a theoretical proposal or framework is provided) |

*Assessing relevance of evidence in qualitative reviews*

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

### *Assessing coherence of findings in qualitative reviews*

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

### *Assessing adequacy of data in qualitative reviews*

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

### *Assessing importance in qualitative reviews*

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

## Reviewing economic evidence

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

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

| **Inclusion criteria** |
|---|
| Intervention or comparators in accordance with the guideline scope |
| Study population in accordance with the guideline scope |
| Full economic evaluations (cost-utility, cost effectiveness, cost-benefit or cost-consequence analyses) assessing both costs and outcomes associated with interventions of interest |
| **Exclusion criteria** |
| Abstracts containing insufficient methodological details |
| Cost-of-illness type studies |

Once the screening of titles and abstracts was completed, full-text copies of potentially relevant articles were requested for detailed assessment. Inclusion and exclusion criteria were applied to articles obtained as full-text copies.

Details of economic evidence study selection, lists of excluded studies, economic evidence tables, and the results of quality assessment of economic evidence are in Appendix G of the evidence reports.

### Appraising the quality of economic evidence

The quality of economic evidence was assessed using the economic evaluations checklist specified in Developing NICE guidelines: the manual.

## Economic modelling

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

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

- [B2] Investigating and diagnosing suspected meningococcal disease with blood and urine investigations
- [B5] Role of neuroimaging prior to lumbar puncture

However, no modelling was ultimately undertaken for either review. The committee were not persuaded that the clinical evidence was sufficiently strong to make a recommendation for procalcitonin which would have represented a change in current NHS practice for the investigation and diagnosis of suspected meningococcal disease. Furthermore, whilst the evidence review included many studies it was not possible to synthesise the data because of the heterogeneity between them. Finally, it was thought that the data would be lacking to map diagnostic test accuracy to “hard” health outcomes and QALYs. Assumptions could have been made to address this but given other model uncertainties it was thought that any output from such a model would be difficult to draw substantive conclusions from.

It was also decided, following the presentation of the clinical evidence, that health economic modelling would not aid the committee decision making on the role of neuroimaging prior to lumbar puncture. The evidence for effectiveness was not derived from randomised controlled trial data and was generally low quality. Furthermore, neuroimaging prior to lumbar puncture is not currently recommended and the effectiveness data, such as it was, did not show evidence of benefit.

As new economic analysis was not undertaken in this guideline, the committee made a qualitative judgement regarding cost effectiveness by considering expected differences in resource and cost use between options, alongside clinical effectiveness evidence identified from the clinical evidence review.

### Cost effectiveness criteria

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

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

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

## Developing recommendations

### Guideline recommendations

Recommendations were drafted on the basis of the committee's interpretation of the available evidence, taking account of the balance of benefits, harms and costs between different courses of action. When effectiveness and economic evidence was of poor quality, conflicting or absent, the committee drafted recommendations based on their expert opinion. The considerations for making consensus-based recommendations include the balance between potential benefits and harms, the economic costs or implications compared with the economic benefits, current practices, recommendations made in other relevant guidelines, person's preferences and equality issues.

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

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

### Research recommendations

When areas were identified for which evidence was lacking, the committee considered making recommendations for future research. For further details refer to Developing NICE guidelines: the manual and NICE's Research recommendations process and methods guide.

## Validation process

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

# Updating the guideline

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

# References

**Bradburn 2007**

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

**Dixon-Woods 2005**

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

**Hayden 2013**

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

**Higgins 2011**

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors) (2020) Cochrane Handbook for Systematic Reviews of Interventions Version 6.1 [updated September 2020] The Cochrane Collaboration. Available from www.training.cochrane.org/handbook (accessed 07 June 2023)

**Lewin 2015**

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

**McGowan 2016**

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

**NICE 2018**

National Institute for Health and Care Excellence (NICE) (2014) NICE Policy on conflicts of interest (updated 2017). Available from https://www.nice.org.uk/Media/Default/About/Who-we-are/Policies-and-procedures/declaration-of-interests-policy.pdf (accessed 07 June 2023)

**Santesso 2016**

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