National Institute for Health and Care Excellence

Final

Lyme disease: diagnosis and management

NICE guideline: methods

NICE guideline 95

Methods

April 2018

Final

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

Update information

Minor changes since publication

May 2018: The definition for the Jarisch-Herxheimer reaction was amended to clarify when the reaction can start.

ISBN: 978-1-4731-2919-1

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 professionals.

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

NICE received the remit for this guideline from NHS England. NICE commissioned the NGC to produce the guideline.

The remit for this guideline is:

to develop a clinical guideline on the diagnosis and management of Lyme disease.

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 Saul Faust in accordance with guidance from NICE.

The group met approximately every 4–6 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

This guideline includes adults, young people and children with suspected or confirmed Lyme disease. A particular focus lies on the assessment, diagnosis and management of Lyme disease, as well as information needs of people with suspected or confirmed Lyme disease. The incidence of Lyme disease in the UK and the transmission of the disease are also covered. 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 the diagnosis and management of other tick-borne infections or the prevention of Lyme disease.

1.3.3 Relationships between the guideline and other NICE guidance

Related NICE guidelines:

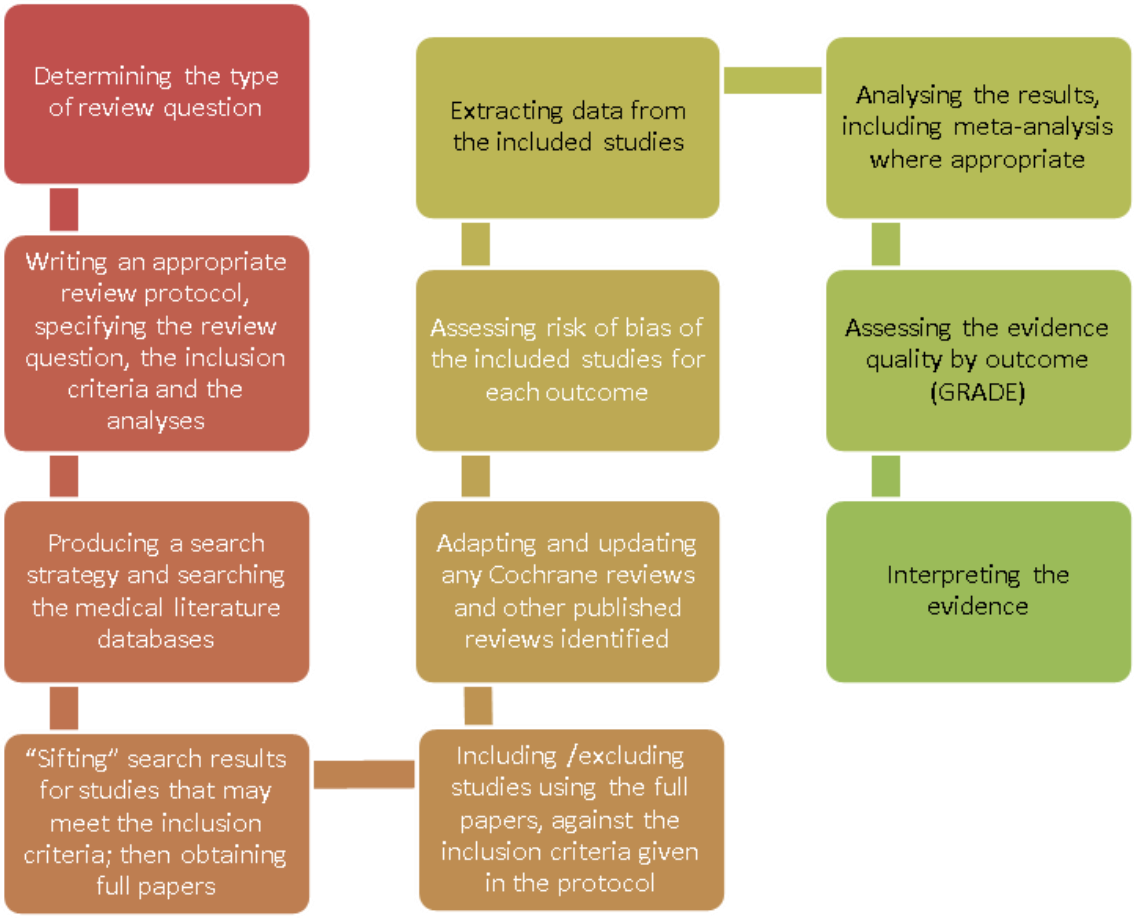
- Patient experience in adult NHS services: improving the experience of care for people using adult NHS services. NICE guideline CG138 (2012).
- Common mental health problems: identification and pathways to care. NICE guideline CG123 (2011).
- Medicines adherence: involving patients in decisions about prescribed medicines and supporting adherence. NICE guideline CG76 (2009).
- Chronic fatigue syndrome/myalgic encephalomyelitis (or encephalopathy): diagnosis and management. NICE guideline CG53 (2007).

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.[4]

Sections 2.1 to 2.3 describe the process used to identify and review clinical evidence (summarised in Figure 1), sections 2.2 and 2.4 describe the process used to identify and review the health economic evidence, and section 2.5 describes the process used to develop recommendations.

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; and using a framework of population, setting and context for qualitative reviews.

The review questions on the transmission and incidence of Lyme disease were developed using a framework of population, target condition and measures of probability of occurrence, that is, legitimate incidence, prevalence or transmission risk estimates.

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. The questions were based on the key clinical areas identified in the scope.

A total of 15 review questions were identified.

Full literature searches, critical appraisals and evidence reviews were completed for all the specified review questions except for the question on the awareness of Lyme disease. The recommendations for raising awareness of Lyme disease were based on discussions, consensus and expert opinion of the committee and were also informed by other review questions. The committee agreed that there was no published evidence that could inform these recommendations.

The evidence included in the review chapter on the management of persistent symptoms was identified through the review on the management of non-specific symptoms. The committee agreed that the evidence on persistent symptoms associated with Lyme disease should be separated out because the study populations represented a different patient group seen in clinical practice. No separate review was undertaken for the management of persistent symptoms associated with Lyme disease as the committee agreed that treatment failure and duration of symptoms should be considered as part of each management review.

Table 1: Review questions

Evidence report	Type of review	Review questions	Outcomes
A	No formal review was undertaken Recommendations were based on discussions, consensus and expert opinion and informed by the review on the incidence of Lyme disease	In whom should Lyme disease be suspected?	Identify people who may have Lyme disease and should undergo further investigation
A	Epidemiological	What is the incidence of Lyme disease in the UK?	• Incidence of Lyme disease • Prevalence of Lyme disease
B	Diagnostic	In people with suspected (or under investigation for) Lyme disease, how accurate are physiological signs and symptoms to identify whether Lyme disease is present?	Detecting Lyme disease Critical outcome: • Sensitivity Important outcomes: • Specificity • Receiver Operating Characteristic (ROC) curve or area under curve
C	Diagnostic	In people with suspected (or under investigation for) Lyme disease, what is the most accurate initial test to identify whether Lyme disease is present?	Detecting Lyme disease Critical outcome: • Sensitivity Important outcomes:

Evidence report	Type of review	Review questions	Outcomes
			• Specificity • Receiver Operating Characteristic (ROC) curve or area under curve
C	Diagnostic	In people with a positive test for Lyme disease, what is the most accurate test to confirm or rule out Lyme disease?	Detecting Lyme disease Critical outcome: • Specificity Important outcomes: • Sensitivity • Receiver Operating Characteristic (ROC) curve or area under curve
C	Diagnostic	In people with suspected (or under investigation for) Lyme disease, what is the most accurate combination of tests to diagnose or rule out Lyme disease?	Detecting Lyme disease Critical outcome: • Sensitivity Important outcomes: • Specificity • Receiver Operating Characteristic (ROC) curve or area under curve
D	Intervention	What is the most clinically and cost-effective treatment for people with an erythema chronicum migrans?	Critical outcomes: • Quality of life • Cure (resolution of symptoms) • Reduction of clinical symptoms • Symptom relapse Important outcomes: • Adverse events
E	Intervention	What is the most clinically and cost-effective treatment for seropositive people, who have non-specific symptoms that may be related to Lyme disease?	Critical outcomes: • Quality of life • Cure (resolution of symptoms) • Reduction of clinical symptoms • Symptom relapse Important outcomes: • Adverse events
F	Intervention	What is the most clinically and cost-effective treatment for people with symptoms consistent with neuroborreliosis?	Critical outcomes: • Quality of life • Cure (resolution of symptoms) • Reduction of clinical symptoms

Evidence report	Type of review	Review questions	Outcomes
			• Symptom relapse Important outcomes: • Adverse events
G	Intervention	What is the most clinically and cost-effective treatment for people with arthritis related to Lyme disease?	Critical outcomes: • Quality of life • Cure (resolution of symptoms) • Reduction of clinical symptoms • Symptom relapse Important outcomes: • Adverse events
H	Intervention	What is the most clinically and cost-effective treatment for people with acrodermatitis chronica atrophicans related to Lyme disease?	Critical outcomes: • Quality of life • Cure (resolution of symptoms) • Reduction of clinical symptoms • Symptom relapse Important outcomes: • Adverse events
I	Intervention	What is the most clinically and cost-effective treatment for people with carditis related to Lyme disease?	Critical outcomes: • Quality of life • Cure (resolution of symptoms) • Reduction of clinical symptoms • Symptom relapse Important outcomes: • Adverse events
J	Intervention	What is the most clinically and cost-effective treatment for people with lymphocytoma related to Lyme disease?	Critical outcomes: • Quality of life • Cure (resolution of symptoms) • Reduction of clinical symptoms • Symptom relapse Important outcomes: • Adverse events
K	Intervention	What is the most clinically and cost-effective treatment for people with non-neurological ocular manifestations related to Lyme disease (for example, keratitis, uveitis, iritis, scleritis)?	Critical outcomes: • Quality of life • Cure (resolution of symptoms) • Reduction of clinical symptoms

Evidence report	Type of review	Review questions	Outcomes
			• Symptom relapse Important outcomes: • Adverse events
L	No formal review was undertaken The included evidence was identified through the review on the management of non-specific symptoms related to Lyme disease	What is the most clinically and cost-effective treatment for seropositive people, who have non-specific symptoms that may be related to Lyme disease?	Critical outcomes: • Quality of life • Cure (resolution of symptoms) • Reduction of clinical symptoms • Symptom relapse Important outcomes: • Adverse events
M	Epidemiological	What are the patterns of person-to-person transmission of Lyme disease?	Transmission risk of Lyme disease
N	Qualitative	What information do people with suspected, confirmed or treated Lyme disease need?	Any type of information described by studies. • Content of information required and how this information is delivered • Information for carers and family members as well as information for patients • Timing of information

2.2 Searching for evidence

2.2.1 Clinical and health economic 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 economic evidence relevant to the review questions. Searches were undertaken according to the parameters stipulated within the NICE guidelines manual 2014.[4] Databases were searched using relevant medical subject headings, free-text terms and study-type filters where appropriate. An exclusion filter can be applied to remove certain study designs and publication types by using the Boolean operator 'NOT'. Studies published in languages other than English were not reviewed, where possible searches were restricted to English language. All searches were updated on 3 July 2017. Papers published or added to databases after this date were not considered. If new evidence falls outside of the timeframe for the guideline searches, for example, 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. Medline search strategies were checked by a second information specialist before being run. Searches were crosschecked with reference lists of highly relevant papers, searches in other systematic reviews analysed, and committee members requested to highlight any additional studies they were aware of.

The titles and abstracts of records retrieved by the searches were sifted for relevance, with potentially significant publications obtained in full text. These were assessed against the inclusion criteria in the protocols.

During the scoping stage, a search was conducted for guidelines and reports on the websites listed below from organisations relevant to the topic.

- 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)
- NHS 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 the committee considered for pharmaceutical interventions may be different from that the MHRA and European Medicines Agency considered 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.[4] Qualitative studies were critically appraised using the GRADE CERQual approach for rating confidence in the body of evidence as a whole and using an NGC checklist for the methodological limitations section of the quality assessment. Epidemiological studies were critically appraised using an adapted version of The Joanna Briggs Institute critical appraisal checklist for studies reporting incidence and prevalence data.[3]
- 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.
    - Diagnostic data studies: coupled sensitivity and specificity values were summarised in forest plots. No meta-analyses were undertaken for the 3 review questions on

diagnostic tests. This was due to heterogeneity in terms of different types of diagnostic tests and their manufacturers, differences in how the tests were performed and their results were analysed, and differences in the study populations analysed. Where meta-analysis was performed for the review question on signs and symptoms, coupled sensitivity and specificity values were also presented on summary Receiver Operating Characteristic (sROC) plots along with the results of the meta-analysis (the summary sensitivity and specificity point and 95% confidence region) and the summary curve. Where evidence was not meta-analysed, because studies differed in population or outcome, then no alternative pooling strategies were carried out on the basis that such pooling would have little meaning. Results from single studies were presented.

  - Qualitative data were synthesised across studies and presented as summary statements with accompanying GRADE CERQual ratings for each review finding.
  - Epidemiological data were presented as individual values or as a range of values. No meta-analyses were undertaken because the majority of studies based their incidence calculations on samples tested at reference laboratories in England and Scotland. As such, meta-analysing the individual results would mean that samples could be counted multiple times.
- 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.

The key population inclusion criterion was:

- People of all ages with any clinical presentation of Lyme disease.

The key population exclusion criterion was:

- People with other tick-borne infections.

Literature reviews, posters, letters, editorials, comment articles, unpublished studies, conference abstracts 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. In the qualitative review for this guideline, however, due to a general lack of evidence all identified studies were included.

2.3.2 Type of studies

Randomised trials, non-randomised intervention studies, and other observational studies (including diagnostic and epidemiological 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 there was limited evidence from RCTs, well-conducted non-randomised studies were included. Please refer to the review protocols in each evidence report for full details on the study design of studies selected for each review question.

For diagnostic accuracy review questions, cross-sectional studies and retrospective studies were considered the most appropriate study design. Case–control studies were also included due to a general lack of evidence from cross-sectional and retrospective studies.

For epidemiological review questions, any studies reporting an incidence or prevalence estimate or a transmission risk estimate for Lyme disease were included.

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)[10] software to combine the data given in all studies for each of the outcomes of interest for the review question.

All analyses were stratified for age (under 18 years and 18 years or over), which meant that different studies with predominant age-groups in different age strata were not combined and analysed together. For some questions, additional stratification was used, and this is documented in the individual review question protocols in each evidence report. When additional strata were used, this led to substrata (for example, using 2 stratification criteria leads to 4 substrata, using 3 stratification criteria leads to 9 substrata) which were analysed separately.

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:

- Quality of life (if dichotomised)
- cure or resolution of specific symptoms
- reduction of specific clinical symptoms (if dichotomised)
- the relapse of specific symptoms
- adverse events.

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. If there were zero events in both arms in an individual study or in at least one study in a meta-analysis, the risk difference was calculated.

Continuous outcomes

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

- Symptom scales (such as visual analogue scale)
- function and activities of daily living.

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[10] 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.[10] 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 substantial heterogeneity) as well as the distribution of effects. Where significant heterogeneity was present, predefined subgrouping of studies was carried out for either:

- Pregnant women
- People who are immunocompromised
- People in whom a previous course of treatment had failed
- People who have been partially treated
- People with ehrlichiosis
- For the review on the management of erythema migrans, different presentations of the erythema migrans rash (single versus multiple erythema migrans).

If the subgroup analysis resolved heterogeneity within any 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.

If any 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 diagnostic test accuracy reviews

For diagnostic test accuracy studies, a positive result on the index test was found if the person had values of the measured quantity above or below a threshold value, and different thresholds could be used. The threshold of a diagnostic test is defined as the value at which the test can best differentiate between those with and without the target condition. The committee could not prespecify any thresholds because different manufactures or test kits used different thresholds. The committee also agreed that the majority of studies would not provide such a level of detail. Instead, studies would only list positive, negative and sometimes equivocal test results. When the full-text of identified studies was assessed, no information to allow for thresholds to be set could be identified.

Diagnostic test accuracy measures used in the analysis were: sensitivity and specificity, and (if appropriate) area under the receiver operating characteristics (ROC) curve (AUC). Positive and negative predictive values were not used because they are not intrinsic to the test and depend on the prevalence of Lyme disease. If a test has a high sensitivity, then very few people with the condition will be missed (few false negatives). For example, a test with a sensitivity of 97% will only miss 3% of people with the condition. Conversely, if a test has a high specificity then few people without the condition would be incorrectly diagnosed (few false positives). For example, a test with a specificity of 97% will only incorrectly diagnose 3% of people who do not have the condition as positive. For the review questions on signs and symptoms, initial tests and test combinations, sensitivity was considered more important than specificity due to the consequences of a missed Lyme disease diagnosis (false negative result). Initial tests with a high sensitivity are often followed by a confirmatory test with a high specificity to eliminate as many false positive test results as possible. Specificity was therefore considered more important for the review question on confirmatory tests. Coupled forest plots of sensitivity and specificity with their 95% CIs across studies were produced for each test, using RevMan5.[10] In order to do this, 2×2 tables (the number of true positives, false positives, true negatives and false negatives) were directly taken from the study if given, or else were derived from raw data or calculated from the set of test accuracy statistics.

Diagnostic meta-analysis was conducted where appropriate, that is, when 3 or more studies were available per test, sign or symptom, and when populations did not differ considerably between studies in terms of their clinical presentations. The accuracy of a test, sign or symptom was pooled using the bivariate method for the direct estimation of summary sensitivity and specificity using a random-effects approach in WinBUGS software.[13] The advantage of this approach is that it produces summary estimates of sensitivity and specificity that account for the correlation between the 2 statistics. Other advantages of this method have been described elsewhere.[9,11,12] The bivariate method uses logistic regression on the true positives, true negatives, false positives and false negatives reported in the studies. Overall sensitivity and specificity and confidence regions were plotted (using methods outlined by Novielli 2010.[7]) Pooled sensitivity and specificity and their 95% CIs were reported in the clinical evidence summary tables. If values could not be pooled, then the individual sensitivity values and their coupled specificity were presented in the clinical evidence summary.

Heterogeneity or inconsistency amongst studies is usually visually inspected in the forest plots and pooled diagnostic meta-analysis plots. The study populations in the included studies in all the guideline diagnostic test reviews differed from each other in terms of clinical presentation and disease duration, and in many studies, there was a general lack of detail on patient characteristics. There were also considerable differences in the type of test and their manufacturers, the way the tests were performed, the analysis of samples, and the interpretation of the test results. Sensitivity and specificity data were grouped by age, clinical

presentation and type of test and graphically presented in forest plots. No sensitivity and specificity data in the diagnostic test accuracy reviews were meta-analysed, however, as a result of the underlying heterogeneity explained above. Furthermore, the committee could not pre-specifiy a sensitivity or specificity threshold above which they would consider recommending the test. The committee knew that none of the commercially available tests for Lyme disease were 100% accurate but was not aware of the general accuracy of these tests. In the absence of a decision-making threshold and given the large number of coupled sensitivity and specificity data points included, inconsistency was not assessed.

The committee could not decide on any minimum sensitivity thresholds needed for recommending a diagnosis of Lyme disease based in the presence of specific signs or symptoms either. This was because of the dichotomous nature of signs and symptoms in the included studies. In the absence of a decision-making threshold, inconsistency could not be assessed in the signs and symptoms review.

2.3.3.3 Data synthesis for qualitative study reviews

The main findings for each included paper were identified and the thematic analysis method was used to synthesise this information. Broad and overarching patterns and themes were identified in the data. Through an iterative process these were further synthesised and refined into the main findings, each summarised with a concise statement of review finding. The evidence was presented in the form of a narrative summary detailing the data 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 based on the checklist were listed. A summary evidence table with the succinct summary statements for each review finding was produced including the associated quality assessment per outcome (review finding).

2.3.3.4 Data synthesis for epidemiological study reviews

Estimates of incidence, prevalence and transmission risk were presented in summary evidence tables as reported in the included studies and, where appropriate, narratively summarised. No meta-analyses or other statistical analyses were undertaken.

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

Quality element	Description
	lack of blinding of the patient, healthcare professional or assessor) and attrition 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 professionals)	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 • performance in outcome measures

Limitation	Explanation
	• 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.

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.

2.3.4.1.2 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.

2.3.4.1.3 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 by subgroup analyses, 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.

2.3.4.1.4 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 'GRADE default' method, as follows:

- For categorical outcomes the MIDs were taken to be RRs or Peto ORs of 0.75 and 1.25. For 'positive' outcomes such as 'patient satisfaction', the RR or Peto OR of 0.75 is taken as the line denoting the boundary between no clinically important effect and a clinically important harm, while the RR or Peto OR of 1.25 is taken as the line denoting the boundary between no clinically important effect and a clinically important benefit. For 'negative' outcomes such as 'bleeding', the opposite occurs, so the RR or Peto OR of 0.75 is taken as the line denoting the boundary between no clinically important effect and a clinically important benefit, while the RR or Peto OR of 1.25 is taken as the line denoting the boundary between no clinically important effect and a clinically important 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 important 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 important 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.

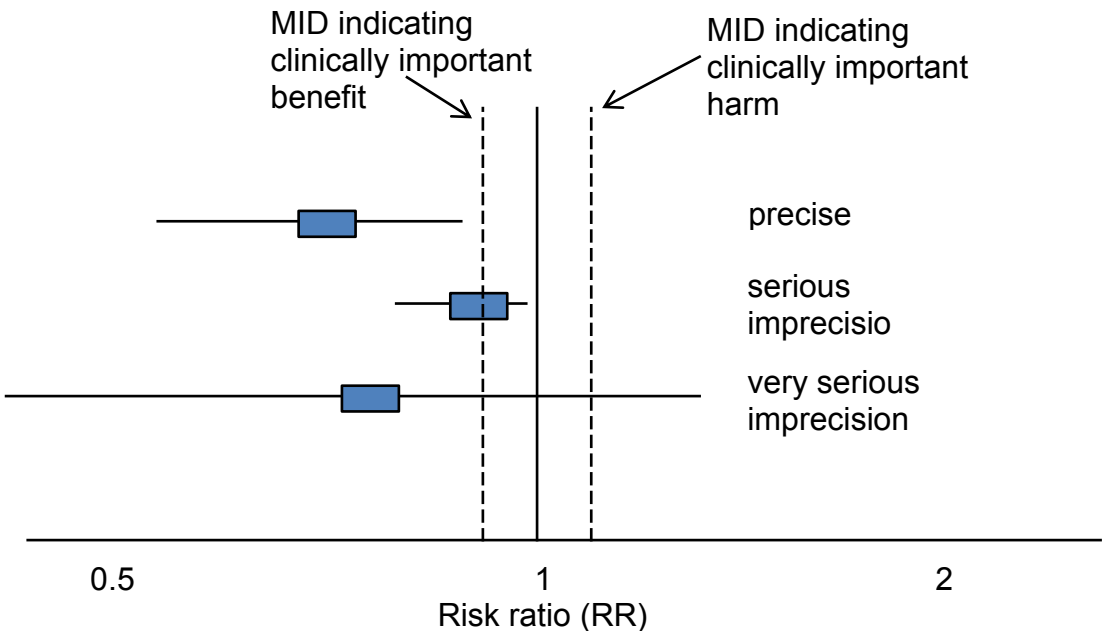
- If standardised mean differences have been used, then the MID will be set at the absolute value of +0.5. This follows because standardised mean differences are mean differences normalised to the pooled standard deviation of the 2 groups, and are thus effectively expressed in units of 'numbers of standard 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.

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, MIDs were found in the literature for the continuous health-related quality of life outcome SF-36 which were used to assess imprecision and clinical importance (see section 2.3.5 below).[2] No other appropriate MIDs were found in the literature, and so the default method was adopted for all other continuous and dichotomous outcomes.

Risk difference was used when individual studies or some studies in the meta-analysis had a zero event rate in both arms. The absolute cut-offs for assessing importance were used as MIDs for imprecision in this case.

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)



2.3.4.1.5 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 4. 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 4: 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 Diagnostic studies

Risk of bias and indirectness of evidence for diagnostic data were evaluated by study using the Quality Assessment of Diagnostic Accuracy Studies version 2 (QUADAS-2) checklists (see appendix H in the NICE guidelines manual 2014[4]). Risk of bias and applicability in primary diagnostic accuracy studies in QUADAS-2 consists of 4 domains (see Figure 3):

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

Figure 3: Summary of QUADAS-2 with list of signalling, risk of bias and applicability questions.

Domain	Patient selection	Index test	Reference standard	Flow and timing
Description	Describe methods of patient selection. Describe included patients (prior testing, presentation, intended use of index test and setting)	Describe the index test and how it was conducted and interpreted	Describe the reference standard and how it was conducted and interpreted	Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2×2 table (refer to flow diagram). Describe the time interval and any interventions between index test(s) and reference standard
Signalling questions (yes/no/ unclear)	Was a consecutive or random sample of patients enrolled?	Were the index test results interpreted without knowledge of the results of the reference standard?	Is the reference standard likely to correctly classify the target condition?	Was there an appropriate interval between index test(s) and reference standard?
	Was a case–control design	If a threshold was used, was it pre-	Were the reference	Did all patients receive a reference standard?

Domain	Patient selection	Index test	Reference standard	Flow and timing
	avoided?	specified?	standard results interpreted without knowledge of the results of the index test?	
	Did the study avoid inappropriate exclusions?			Did all patients receive the same reference standard?
				Were all patients included in the analysis?
Risk of bias; (high/low/ unclear)	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Concerns regarding applicability (high/low/ unclear)	Are there concerns that the included patients do not match the review question?	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?	

2.3.4.2.1 Inconsistency

Inconsistency refers to an unexplained heterogeneity of results for an outcome across different studies. Inconsistency is usually assessed by inspection of the sensitivity OR specificity value (based on the primary measure) using the point estimates and 95% CIs of the individual studies on the forest plots. Particular attention is placed on values above or below 50% (diagnosis based on chance alone) and the threshold set by the committee (the threshold above which it would be acceptable to recommend a test). For example, the committee might have set a threshold of 90% as an acceptable level to recommend a test.

As mentioned above, the committee could not set any decision-making thresholds. This was because the committee knew that the available tests for Lyme disease were not 100% accurate, but the committee was not aware of the general accuracy of these tests. To determine a decision making threshold, the committee would have had to either set a very low threshold for sensitivity or specificity to be able to consider available tests or a very high threshold and risk not being able to recommend any test.

Similarly, thresholds could not be defined for signs and symptoms because of the dichotomous nature of these in the included studies. A sign or symptom was either present or absent. While in practice, signs and symptoms can vary by degree, the included studies did not provide any such level of detail.

In the absence of decision-making thresholds for the diagnostic accuracy reviews in this guideline and the large number of coupled sensitivity and specificity data points included, the pragmatic approach not to assess for inconsistency was chosen.

2.3.4.2.2 Imprecision

The judgement of precision was based on visual inspection of the confidence region around the summary sensitivity and specificity point from the diagnostic meta-analysis, if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted, imprecision was assessed according to the range of point estimates or, if only 1 study contributed to the evidence, the 95% CI around the single study. As a general rule

(after discussion with the committee), a variation of 0–20% was considered precise, 20–40% serious imprecision, and >40% very serious imprecision. Imprecision was assessed on the primary outcome measure for decision-making.

2.3.4.2.3 Overall grading

Quality rating started at High for cross-sectional studies, and each major limitation (risk of bias, indirectness, inconsistency and imprecision) brought the rating down by 1 increment to a minimum grade of Very Low, as explained for intervention reviews.

The quality rating for case-control studies started at Low without the option to upgrade due to the inherent issues around the selection of patients and the potential overestimate of the diagnostic test accuracy. This is because populations in case-control studies tend to differ from 'true populations' found in clinical practice as cases tend to be more severely ill than the average patient population in clinical practice in order to fit inclusion criteria of studies. Controls, on the other hand, are usually drawn from a healthy population or include known specific cross-reactivity controls.

2.3.4.3 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 5.

Table 5: 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 NGC 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.

2.3.4.3.1 Methodological limitations

Each review finding had its methodological limitations assessed within each study first using an NGC 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.

2.3.4.3.2 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.

2.3.4.3.3 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.

2.3.4.3.4 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 reasonably 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.

2.3.4.3.5 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 6. 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 6: 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.

2.3.4.4 Epidemiological reviews

In the absence of any established study limitations checklists for epidemiological reviews, risk of bias and indirectness of evidence were assessed using an adapted version of a checklist for incidence and prevalence studies published by The Joanna Briggs Institute[3]. The published checklist was adapted for the purpose of this guideline because the sections on an adequate sample size and response rate were not applicable for this guideline, since the reviews on the incidence and transmission of Lyme disease used clinical data, such as serological samples tested, to establish a risk estimate. Response rates to surveys, such as a general census, or drop-out rates were therefore not applicable.

Table 7: Description of quality elements for incidence and prevalence studies (as adapted from a checklist developed by The Joanna Briggs Institute[3] for the purpose of this guideline)

Quality element	Description
Was the sample frame appropriate to address the target population?	This question relies upon knowledge of the broader characteristics of the population of interest and the geographical area. Consideration to specific population characteristics in the study, including age range, gender, morbidities, medications, and other potentially influential factors should be given. A sample frame may be appropriate when it includes almost all the members of the target population (i.e. a census, or a complete list of participants or complete registry data), but may be inappropriate if only a certain group has been used (such as one profession).
Were the study participants sampled in an appropriate way?	Random probabilistic sampling from a defined subset of the population (sample frame) should be employed in most cases, however, random probabilistic sampling is not needed when everyone in the sampling frame will be included or analysed. For example, reporting on all the data from a good census is appropriate as a good census will identify everybody. When using cluster sampling, such as a random sample of villages within a region, the methods need to be clearly stated as the precision of the final prevalence estimate incorporates the clustering effect. Convenience samples, such as a street survey or interviewing many people at a public gatherings are not considered to provide a representative sample of the base population.
Were the study subjects and setting described in detail?	Certain diseases or conditions vary in prevalence across different geographic regions and populations. The study sample should be described in sufficient detail so that other researchers can determine if it is comparable to the population of interest to them.
Was the data analysis conducted with sufficient coverage of the identified sample?	Coverage bias can occur when not all subgroups of the identified sample respond at the same rate. For instance, the overall response rate for the study may be very high, but the response rate for a certain subgroup (for example, older adults) may be quite low.

Quality element	Description
Were valid methods used for the identification of the condition?	Many health problems are not easily diagnosed or defined and some measures may not be capable of including or excluding appropriate levels or stages of the health problem. If the outcomes were assessed based on existing definitions or diagnostic criteria, then the answer to this question is likely to be yes. If the outcomes were assessed using observer reported, or self-reported scales, the risk of over- or under-reporting is increased, and objectivity is compromised. Importantly, it should be determined if the measurement tools used were validated instruments as this has a significant impact on outcome assessment validity.
Was the condition measured in a standard, reliable way for all patients?	Considerable judgment is required to determine the presence of some health outcomes. Were those involved in collecting data trained or educated in the use of the instrument(s)? If there was more than one data collector, were they similar in terms of level of education, clinical or research experience, or level of responsibility in the piece of research being appraised? When there was more than one observer or collector, was there comparison of results from across the observers? Was the condition measured in the same way for all participants?
Was there appropriate statistical analysis?	Importantly, the numerator and denominator should be clearly reported, and percentages should be given with confidence intervals. The methods section should be detailed enough for reviewers to identify the analytical technique used and how specific variables were measured. Additionally, it is also important to assess the appropriateness of the analytical strategy in terms of the assumptions associated with the approach as differing methods of analysis are based on differing assumptions about the data and how it will respond.
Other limitations	Were there any other issues that could reduce the confidence in the result?

Each question has to be answered with ‘yes’, ‘no’, ‘unclear’ or ‘not applicable’. Based on the quality elements described in Table 7, studies were given a Low, Moderate, or High risk of bias.

Inconsistency was not assessed as no meta-analyses or other pooling strategies were performed. Imprecision could not be assessed because the majority of included studies did not provide any confidence intervals for the incidence and transmission estimates.

2.3.5 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 dichotomised outcomes in the intervention reviews that if at least 100 more participants per 1,000 (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 adverse events, 50 events or more per 1,000 (5%) represented clinical harm. For continuous outcomes if the mean difference was greater than the minimally important difference (MID), then this represented a clinical benefit or harm.

Established MIDs were found in the literature for the outcome SF-36 and the values used for imprecision and clinical importance are provided in Table 8. For all other outcome, the default approach was used.

Table 8: MIDs for assessing between group differences

Outcome	MID for imprecision	MID for clinical importance	Source
SF-36	Physical component summary: 2 Mental component summary: 3 Physical functioning: 3 Role-physical: 3 Bodily pain: 3 General health: 2 Vitality: 2 Social functioning: 3 Role-emotional: 4 Mental health: 3		User's manual for the SF-36v2 Health Survey, Third Edition[2]

This assessment was carried out by the committee for each critical outcome, and an evidence summary table was produced to compile the committee's assessments of clinical importance per outcome, alongside the evidence quality and the uncertainty in the effect estimate (imprecision).

2.3.6 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 and the number of participants 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).

2.4 Identifying and analysing evidence of cost effectiveness

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.[4]

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 health economic exploratory analysis in priority areas.

2.4.1 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.[4]
- 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.

2.4.1.1 Inclusion and exclusion criteria

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 2001 and studies from non-OECD countries or the US 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. Where exclusions occurred on this basis, it is noted in the relevant evidence report. 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[4]) 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.

2.4.1.2 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.[4] 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.[8]

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[4]

2.4.2 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. The committee agreed on the priority areas for new analysis after formation of the review questions and consideration of the existing health economic evidence.

The committee identified diagnosis as the highest priority area for original health economic modelling.

The committee identified diagnosis as a high priority because it affects the largest number of people in the guideline (that is, all those tested), there are a number of uncertainties over the most appropriate approach to testing, and there are no includable health economic analyses to aid consideration of cost-effectiveness.

Current practice in the NHS is a 2-tier testing strategy; an 'initial' test (an ELISA) followed by a 'confirmatory' test (an immunoblot) for those with a positive or equivocal initial test result. The committee were interested to establish if the current 2-tier testing was cost effective compared to a single test. They also highlighted uncertainty as to whether other tests, not currently being used in the NHS, may be of value. Based on the review of the clinical evidence identified, the committee agreed to make recommendations that reflected current practice (as described above) with some exceptions, which are discussed in more detail in section 4.4 in chapter C. A full cost–utility analysis to establish whether or not the current 2-tier testing approach is cost effective compared to initial testing only was considered inappropriate as there is too much uncertainty around model inputs and too many tenuous assumptions would be required. As a result, a simple exploratory analysis was conducted to justify the additional cost of 2-tier testing (ELISA including C6 IgM and IgG followed by confirmatory immunoblot if ELISA is positive) over initial testing only (ELISA including C6 IgM and IgG) in people with suspected Lyme disease. More detail on the rationale for this approach and the methodology and results are available in appendix H of chapter C.

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

- Methods were consistent with the NICE reference case for interventions with health outcomes in NHS settings,[4,6] although this analysis was restricted to costs only and so QALYs were not used. Furthermore, a PSA was not deemed useful for this exploratory analysis. Further detail is provided in the full write up.
- The committee was involved in the selection of inputs and interpretation of the results.
- Inputs were based on the systematic review of the clinical literature supplemented with other published data sources where possible.
- Inputs and assumptions were reported fully and transparently.
- The results were subject to sensitivity analysis and limitations were discussed.
- The analysis was peer-reviewed by another health economist at the NGC.

Full methods and results of the exploratory analysis comparing 2-tier testing to single testing for Lyme disease are described in chapter C.

2.4.3 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.[5] 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'.[5]

When QALYs or life-years-gained are not used in the analysis, results are difficult to interpret unless 1 strategy dominates the others with respect to every relevant health outcome and cost.

2.4.4 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.

2.5 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 A-N).
- 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 and summary ROC curves (in appendices to the relevant evidence reports).
- A description of the methods and results of the cost-effectiveness analysis undertaken for the guideline (in a separate economic analysis report).

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 1 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 consensus recommendations were agreed through discussions in the committee. 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 2.5.1 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[4]).

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

2.5.1 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.

2.5.2 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.

2.5.3 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.

2.5.4 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.

2.5.5 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
ACA	Acrodermatitis chrnoica atrophicans
AUC	Area under curve
BIA	British Infection Association
BNF	British National Formulary
Borrelia burgdorferi s.l.	A complex diverse group of 18 globally distributed bacteria of the *Borrelia burgdorferi* species 7 of which are known to infect humans.
CDC	Center for Disease Control and Prevention
CE	*Conformité Européene* or European conformity
CEA	Cost-effectiveness analysis
CI	Confidence interval
CLR	Clarithromycin
CNS	Central nervous system
CRP	C-reactive protein
CSF	Cerebrospinal fluid
CTX	Ceftriaxone
CUA	Cost–utility analysis
DOX	Doxycycline
EAN	European Academy of Neurology
ECM	Erythema chronicum migrans (see EM)
EFNS	European Federation of Neurological Societies
EIA	Enzyme immunoassay
ELISA	Enzyme-linked immunosorbent assay
EM	Erythema migrans
ESCMID	European Society of Clinical Microbiology and Infectious Diseases
ESGBOR	ESCMID Study Group for Lyme Borreliosis
ESR	Erythrocyte sedimentation rate
FCE	Finished consultant (hospital) episode
FN	False negative
FP	False positive
FSS	Fatigue severity score
GC	Guideline Committee
GRADE	Grading of recommendations assessment, development and evaluation
GV	Genzyme Virotech
HCQ	Hydroxychloroquine
HES	Hospital episode statistics
HSCIC	Health and Social Care Information Centre
IB	Immunoblotting
ICER	Incremental cost-effectiveness ratio
IF	Immunofluorescence
IU	International unit
IV	Intravenous

Acronym or abbreviation	Description
LA	Lyme arthritis
LB	Lyme borreliosis (see LD)
LC	Lyme carditis
LD	Lyme disease
LNB	Lyme neuroborreliosis
LP	Lumbar puncture
NB	Neuroborreliosis
NeBoP	Neuroborreliosis prediction
NGC	National Guideline Centre
NICE	National Institute of Health and Care Excellence
NPV	Negative predictive value
OECD	Organisation for Economic Co-operation and Development
PCR	Polymerase chain reaction
Pen V	Phenoxymethylpenicillin
PHE	Public Health England
PO	Per os (by mouth, orally)
PPV	Positive predictive value
QALY	Quality-adjusted life year
RCT	Randomised control trial
RDEH	Royal Devon and Exeter Hospital
RIPL	Rare and Imported pathogens laboratory
ROC	Receiver operating characteristic
RR	Relative risk
SLE	Systemic lupus erythematosus
TN	True negative
TP	True positive
VIsE	Surface lipoprotein E
WB	Western blot
WBC	White blood cell
WC	Whole cell
WCS	Whole cell sonicate

4 Glossary

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

4.1 Guideline-specific terms

Term	Definition
Acrodermatitis chronica atrophicans (ACA)	A late cutaneous manifestation of Lyme borreliosis occurring months to years after inoculation.
Acute disseminated Lyme disease	Symptoms usually occurring in the first 3 months following inoculation which may include but are not limited to multiple erythema migrans lesions, heart block, facial palsy, radiculitis or acute large joint arthritis.
Arrhythmia	Disturbance of the heart's usual rhythm. It's also known as cardiac dysrhythmia.
Assay	An investigative (analytic) procedure in laboratory medicine and molecular biology for qualitatively assessing or quantitatively measuring the presence, amount, or functional activity of a target entity (the analyte) such as an antibody or antigen.
Borrelia burgdorferi (*B. burgdorferi*)	A bacterial species of the spirochete class of the genus *Borrelia*. *B. burgdorferi* and is the predominant causative agent of Lyme disease in the United Kingdom and United States.
Borrelia burgdorferi sensu lato (*Borrelia burgdorferi s.l.*)	A complex diverse group of 18 globally distributed bacteria of the *Borrelia burgdorferi* species 7 of which are known to infect humans.
Borrelia burgdorferi sensu stricto	See *Borrelia burgdorferi.*
C6	The synthetic peptide used in commercial assays is the C6 peptide and the term is in common usage. The domain in the VlsE protein antibodies, which some of the recommended diagnostic tests in this guideline are designed to detect, is the IR6 domain. See IR6.
CE marking	A mandatory conformity marking for certain products sold within the European Economic Area (EEA) since 1985.
Cure	Resolution of symptoms.
CXCL13	A small cytokine belonging to the CXC chemokine family. As its name suggests, this chemokine is selectively chemotactic for B cells belonging to both the B-1 and B-2 subsets, and elicits its effects by interacting with chemokine receptor CXCR5.
Early Lyme disease	Early localised symptoms can begin 1 to 2 weeks after the tick bite. One of the earliest signs is a 'bull's-eye', erythema migrans rash, which is a sign that bacteria are multiplying at the inoculation site. If present, the rash occurs at the site of the tick bite as an area of spreading erythema, which may have central clearing. It may be warm to the touch, but it isn't painful and doesn't itch. This rash will disappear after 4 weeks. Additionally, in early disease, individuals may have an atypical rash, no rash or be asymptomatic.
Enzyme-linked immunosorbent assay (ELISA)	A test that detects and measures antibodies in your blood. This test can be used to determine if you have antibodies related to certain infectious conditions.
Erythema migrans (EM)	A rash often, but not always, seen in the early stage of Lyme disease. It can appear anywhere from one day to one month after a tick bite. This rash does not represent an allergic reaction to the bite, but rather an actual skin infection with the Lyme bacteria, *Borrelia burgdorferi sensu lato*. A target lesion (bull's-eye rash) is characteristic of a *Borrelia burgdorferi s.l.* infection. The rash is characterised by a spreading erythematous lesion with central clearing, is not itchy or painful, can increase in size and can appear for several weeks.

Term	Definition
Facial palsy	Weakness of the facial muscles, resulting from temporary or permanent damage to the facial nerve.
Haemodynamically comprised	A compromised circulatory system, which may result in poor blood flow to the essential organs. Causes include low blood pressure (hypotension), poor cardiac function due to an arrthymia and sepsis.
Heart block	See arrhythmia.
IR6	The domain in the VlsE protein antibodies, which some of the recommended diagnostic tests in this guideline are designed to detect, is the IR6 domain. This is often referred to as the C6 peptide.
Jarisch–Herxheimer reaction	A systemic reaction, thought to be caused by the release of cytokines when large numbers of bacteria are killed by antibiotics. Symptoms include a worsening of fever, chills, muscle pains, and headache. The reaction can start between 1 and 12 hours after antibiotics are started but can also occur later and can last for a few hours or up 1 or 2 days. The reaction is self-limiting and usually resolves within 24 to 48 hours. It was originally reported in the treatment of syphilis but has been documented in tick-borne diseases including Lyme disease, leptospirosis and relapsing fever.
Lyme arthritis	Lyme arthritis is a feature of late-stage infection with the tick-borne spirochete, *Borrelia burgdorferi s.l.*, usually beginning months after the initial tick bite. Sometimes, the earlier phases of the infection are asymptomatic and arthritis is the presenting manifestation of the disease. Patients with Lyme arthritis have intermittent or persistent attacks of joint swelling and pain in 1 or a few large joints, especially the knee, usually over a period of several years, without prominent systemic manifestations.
Lyme carditis	A type of Lyme disease that affects the conducting system of the heart and can result in arrhythmias. Symptoms may include light-headedness, fainting, shortness of breath, heart palpitations or chest pain. People with Lyme carditis may also experience other symptoms such as fever and body aches or erythema migrans rash.
Lyme disease	An infection caused by bacterium carried by deer ticks. If caught early, it is easily treated with antibiotics. If left untreated, the infection can spread to the joints, heart, and nervous system, causing a complex debilitating disorder that is more difficult to treat. It is used in this guideline to refer to both the disease and to tests for an antibody response. This reflects the terminology used in clinical practice. See *Borrelia burgdorferi sensu lato*.
Lyme meningitis	When Lyme disease affects the nervous system, it may produce symptoms of meningitis. Meningitis is characterized by headaches that fluctuate in intensity from mild to severe with or without associated nausea, vomiting, light sensitivity, neck stiffness, or pain on eye motion.
Lyme neuroborreliosis	See neuroborreliosis.
Lymphocytoma	Lymphocytoma cutis is a rare, chronic, benign cutaneous B-cell lymphoproliferative condition caused by *Borrelia burgdorferi s.l.* usually presenting with red-purple papules, nodules or plaques predominantly on the head and neck.
miniVIDAS	See VIDAS.
NeBoP score	A clinical prediction test for evaluation of children with Lyme Neuroborreliosis in Europe. This weighted score is derived from facial palsy, fever, fatigue, erythema chronicum migrans or lymphocytoma, and pleocytosis in cerebrospinal fluid.
Neuroborreliosis	A disorder of the central nervous system; a neurological manifestation

Term	Definition
	of Lyme disease that is caused by a systemic infection of spirochetes of the genus *Borrelia burgodorferi s.l.* such as meningitis.
Neurosyphilis	An infection of the brain or spinal cord caused by the spirochete Treponema pallidum. It usually occurs in people who have had chronic, untreated syphilis, usually about 10 to 20 years after first infection and develops in about 25%–40% of persons who are not treated.
Person-to-person transmission	People infected through human contact, for example, sexually or vertically. See vertical transmission.
Photosensitvity	An abnormal sensitivity to light.
PRISMA	An evidence-based minimum set of items for reporting in systematic reviews and meta-analyses. PRISMA focuses on the reporting of reviews evaluating randomized trials, but can also be used as a basis for reporting systematic reviews of other types of research, particularly evaluations of interventions.
Radiculopathy	The consequence of nerve root damage due to any cause, symptoms may include pain, numbness (paraesthesia) and weakness localised to the site of damage.
Red rash	Erythemous migrans rash. See Erythema chronicum migrans.
Relapse	Return of symptoms after cure established.
Seropositive	A positive result after testing a person's blood (serum) for a particularly antibody.
SF-36 component	Short Form (form 36) is a health survey of 36-items. It is a patient-reported survey of patient health.
Spirochaete	A spirochaete or spirochete is a member of the phylum Spirochaetes which contains distinctive diderm (double-membrane) bacteria, most of which have long, helically coiled (corkscrew-shaped or spiralled, hence the name) cells. *Borrelia* are a genus of the spircohaete family.
Systemically unwell	A condition involving the body as a whole, as opposed to limited conditions that affect particular parts of the body.
Tick-borne spirochaete	Tick-borne bacteria. See spirochaete.
Vertical transmission	Disease transmitted from a mother to her child.
VIDAS	An instrument that is multiparametric immunoassay system designed to help provide better care and the most accurate laboratory results.

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

Term	Definition
	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.
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 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.
Case–control study	A study to find out the cause(s) of a disease or condition. This is 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.
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.
Clinical efficacy	The extent to which an intervention is active when studied under controlled research conditions.

Term	Definition
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).
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 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

Term	Definition
	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.
Deterministic analysis	In economic evaluation, this is an analysis that uses a point estimate for each input. In contrast, see Probabilistic analysis
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).

Term	Definition
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 (clinical study)	Criteria that define who is not eligible to participate in a clinical study.
Exclusion criteria (literature review)	Explicit standards used to decide which studies should be excluded from consideration as potential sources of evidence.
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.

Term	Definition
Incidence	Proportion of new cases within a specified time period in the population initially at risk.
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 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.
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).
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’).
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.
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).

Term	Definition
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 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

Term	Definition
	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. 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.
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.
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)
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).
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.
Pre-test probability	In diagnostic tests: The proportion of people with the target disorder in the population at risk at a specific time point or time interval. Prevalence may depend on how a disorder is diagnosed.

Term	Definition
Prevalence	See Pre-test probability.
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 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.

Term	Definition
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 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

Term	Definition
	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. 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 that 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:

2. Maruish ME. User's Manual for the SF-36v2 Health Survey. Quality Metric Incorporated. 2011. Available from: https://books.google.co.uk/books?id=a0vYnQEACAAJ

3. Munn Z, Moola S, Lisy K, Riitano D, Tufanaru C. Methodological guidance for systematic reviews of observational epidemiological studies reporting prevalence and cumulative incidence data. International Journal of Evidence-Based Healthcare. 2015; 13(3):147-153

4. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual. London. National Institute for Health and Care Excellence, 2014. Available from: http://www.nice.org.uk/article/PMG20/chapter/1%20Introduction%20and%20overview

5. 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

6. National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal 2013. London. National Institute for Health and Clinical Excellence, 2013. Available from: http://publications.nice.org.uk/pmg9

7. Novielli N, Cooper NJ, Abrams KR, Sutton AJ. How is evidence on test performance synthesized for economic decision models of diagnostic tests? A systematic appraisal of Health Technology Assessments in the UK since 1997. Value in Health. 2010; 13(8):952-957

8. Organisation for Economic Co-operation and Development (OECD). Purchasing power parities (PPP). 2012. Available from: http://www.oecd.org/std/ppp Last accessed: 08/08/2017.

9. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. Journal of Clinical Epidemiology. 2005; 58(10):982-990

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

11. Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Statistics in Medicine. 2002; 21(4):589-624

12. Van Houwelingen HC, Zwinderman KH, Stijnen T. A bivariate approach to meta-analysis. Statistics in Medicine. 1993; 12(24):2273-2284

13. WinBUGS [Computer programme] version 1.4. Cambridge. MRC Biostatistics Unit University of Cambridge, 2015. Available from: http://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/