National Institute for Health and Care Excellence

FINAL

Epilepsies in children, young people and adults: diagnosis and management

NICE guideline: methods

NICE guideline NG217

Methods

April 2022

FINAL

Developed by the National Guideline Centre

Disclaimer

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

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

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 All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

ISBN: 978-1-4731-4513-9

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 patients 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:

An update of the diagnosis and management of epilepsies in children, young people and adults.

ISBN: 978-1-4731-4513-9

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 Stephen Ward in accordance with guidance from NICE.

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

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

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

1.3.1 What this guideline covers

Children, young people and adults with suspected or confirmed epilepsy.

Diagnosis and assessment of epilepsy

Information and support needs

Pharmacological management of epileptic seizures and epilepsy syndromes

Pharmacological management of epileptic seizures and epilepsy syndromes in girls and women who are able to get pregnant (including those who are pregnant or breastfeeding).

Non-pharmacological management of epileptic seizures

Ongoing monitoring, including referral to specialist services and drug withdrawal.

Psychological, neurodevelopmental, cognitive and behavioural comorbidities.

Reducing the risk of epilepsy-related mortality.

Service design and delivery

Transition from children's and young people's services to adults' services.

Pharmacological management of childhood-onset epileptic seizures and epilepsy syndromes

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

ISBN: 978-1-4731-4513-9

1.3.2 What this guideline does not cover

New-born babies (under 28 days) with acute symptomatic seizures.

Managing non-epileptic seizures

1.3.3 Relationships between the guideline and other NICE guidance

NICE technology appraisals to be incorporated in this guidance:

- Cannabidiol with Clobazam for treating seizures associated with Dravet syndrome TA614 (2019)

Cannabidiol with Clobazam for treating seizures associated with Lennox-Gastaut syndrome TA615 (2019)**Related NICE interventional procedures guidance:**

- Deep brain stimulation for refractory epilepsy in adults IPG678 (2020)
- Vagus nerve stimulation for refractory epilepsy in children IPG50 (2004)

Related NICE guidelines:

- Transient loss of consciousness (blackouts) in over 16s CG109 (2014)
- Cannabis-based medicinal products NG144 2021)

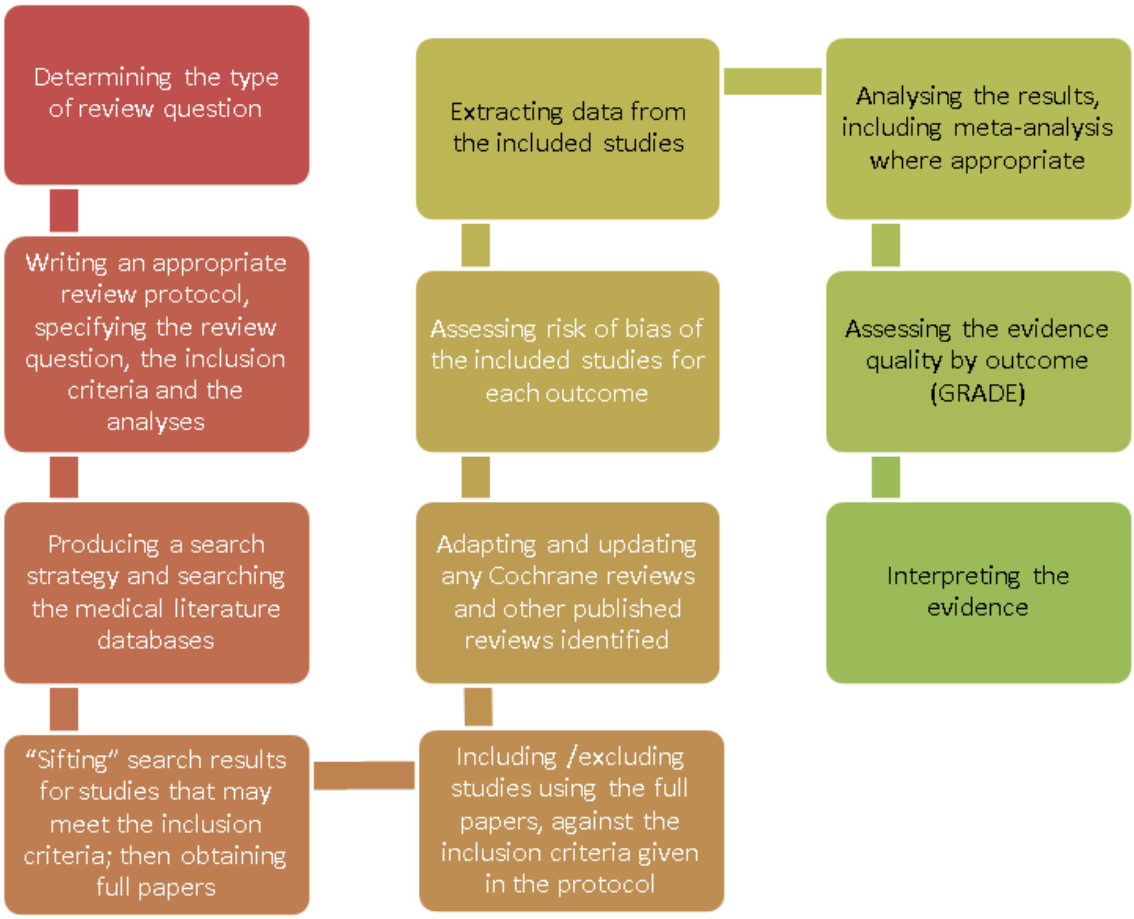
ISBN: 978-1-4731-4513-9

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, last updated October 2020.[3]

Sections 2.1 to 2.3 describe the process used to identify and review clinical evidence (summarised in Figure 1), sections 0 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; using population, presence or absence of factors under investigation (for example prognostic factors) and outcomes for prognostic reviews; and using a framework of population, setting and context for qualitative reviews.

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

ISBN: 978-1-4731-4513-9

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

A total of 23 review questions were identified.

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

Table 1: Review questions

Evidence report	Type of review	Review questions	Outcomes
01	Diagnostic	What are the most accurate tools for predicting a further seizure, in people who have had a single seizure?	• Discrimination: sensitivity, specificity, C statistic. These measures assess how accurately the tool can predict those who will and will not have a second seizure. • Calibration: tests how well the tool results predict the absolute risk of a second seizure. • Net classification Improvement: a sensitive method for evaluating the different levels of predictive accuracy accruing from a change in the prediction tool.
02	Prognostic	What are the modifiable risk factors for a further seizure after a first seizure, and what is the magnitude of risk of those factors?	• Vascular disease • Blood pressure • Activity/exercise levels • Alcohol/ recreational drugs • Psychological factors / stress • Psychosocial factors • Sleep deprivation • ASM use • Other drugs that reduce seizures threshold • Tumours • Drugs affecting sleep • Systemic illness
03	Diagnostic	What is the most accurate approach for 1) diagnosis of epilepsy, and 2) differentiation between types of epilepsy.	Any diagnostic strategies used in papers to detect • epilepsy • type of epilepsy. These may include (for example) symptoms/signs, imaging, EEG, ECG, serum measures, either singly or in combination.
04	Qualitative	2.2 What information and support is needed by people, parents or carers in relation to epilepsy, and when should this be provided?	Information and support do people and their families or carers need (for example, advice on lifestyle, driving, and their treatment). The synthesis of qualitative data will follow a thematic analysis approach. Information will be synthesised into main

ISBN: 978-1-4731-4513-9

Evidence report	Type of review	Review questions	Outcomes
			review findings.
05	Intervention	2.1 What is the clinical and cost effectiveness of digital health technologies (for example, night monitors, wearable devices and Apps) in people with epilepsy?	• Mortality including SUDEP at 12 months • Medicines adherence at 12 months • Healthcare resource impact (including changes in medication use, consultations and hospitalisations) at 12 months • Frequency of seizure-associated risks (such as falls and fractures) at 12 months • Quality of life (measured with a validated scale) at 12 months • Seizure frequency (50% or greater reduction in seizure frequency) at 12 months • Adverse events (total adverse events, anxiety (measured using a validated scale), and false alarms (each reported separately))
06	Intervention	What Anti-seizure Medications (ASMs)_ (individually or add-ons) are safe in the treatment of epilepsies in women and girls who are pregnant and already taking ASMs and in those women who are breastfeeding?	• Major congenital malformations such as neural tube defects (spina bifida), limb defects (club foot), cleft lip and palate, urogenital defects (hypospadias, absent kidneys, abnormal genitalia), cardiac-related (congenital heart disease, including ventricular or atrial septal defect) gastric related (oesophageal atresia and gastroschisis), lung-related (congenital lung cysts) • Minor (less major) congenital malformations such as a missing digit or additional digit, cavernous haemangioma of the skin, or minor versions of congenital heart disease, or spina bifida occulta. • Intellectual quotient (IQ) (Wechsler Intelligence Scale for Children, the Differential Ability Scales) • Development quotient (DQ): (Griffiths and the Bayley Scales) • Other cognitive outcomes: language, memory, attention and executive functioning (Clinical Evaluation of Language Fundamentals, Peabody picture naming. The Children's Memory Scale, Rivermead Memory Test, NEPSY: Neuropsychological Assessment) • Adaptive Behaviour (Vinelands Adaptive Behaviour Scale, the Adaptive Behaviour Assessment System (both have been used in this area) • Neurodevelopmental disorders such as

ISBN: 978-1-4731-4513-9

Evidence report	Type of review	Review questions	Outcomes
			autism, ADHD, dyspraxia
07	Intervention	Monitoring a) When should monitoring be carried out for people with epilepsy? b) How should monitoring be carried out for people with epilepsy, and who should do it?	• Mortality at a minimum of 1 year • Seizure recurrence at a minimum of 1 year • Seizure frequency at a minimum of 1 year • Seizure freedom at a minimum of 1 year • Drug adherence at a minimum of 1 year • Quality of life at a minimum of 1 year • Health care use at a minimum of 1 year • Unplanned hospital admission at a minimum of 1 year • Attendance at ed during a minimum of 1 year • Social functioning (measures of adaptive functioning or adaptive behaviour using a validated scale) at a minimum of 1 year • Cognitive outcomes (including neuropsychological measures of global cognitive functioning, executive functioning and memory using a validated scale) at a minimum of 1 year • In children and young people: neurodevelopmental outcomes (behavioural and emotional outcomes measured with a validated scale) at a minimum of 1 year • Educational outcome at a minimum of 1 year • Placement breakup (change in care location during a minimum of 1 year)

H2	Intervention	What is the appropriate serial monitoring of drug levels, including timing, in girls or women who are thinking about conceiving, are pregnant or in the post-partum period?	• Mortality of mother or baby at study follow-up
			• Seizure freedom during pregnancy and at 6 months post-partum
			• Reduction in seizure frequency (50% or greater reduction in seizure frequency)
			• Time to first seizure in pregnancy up to 6 weeks and time to subsequent seizure up to 1 year
			• Anti-seizure medication exposure (mean daily)
			• Quality of life (any validated measures) at study follow-up
			• Adverse events
			  ○ Anti-seizure medication-related (toxicity)
			  ○ Pregnancy complications in mother and baby (admission to HDU/ICU for mother, admission to NICU for baby)

ISBN: 978-1-4731-4513-9

Evidence report	Type of review	Review questions	Outcomes
			○ Seizures during labour ○ Attendance at ED ○ Congenital anomalies (neural tube defects (spina bifida), limb defects (club foot), cleft lip and palette etc) • Neurodevelopmental outcomes (Griffith Mental Development Scales and the Bayley Scales of Infant and Toddler Development scale)
09	Intervention	What antiepileptic drugs (monotherapy) are effective in the treatment of status epilepticus?	• Mortality (including SUDEP) • Time to seizure cessation, (5 min after drug administration, 10 min, 30 min, 60 min, 1 to 24 hours, up to 24 hours for convulsive, non-convulsive- up to 1 month • Time to event seizure cessation • Seizure recurrence < within less than 24 hours after administration of monotherapy • Time to seizure recurrence after administration of monotherapy • Quality of life (QOLIE-31, QOLIE-AD-48) • Length of ICU stay • Length of hospital stay • Mean Glasgow outcome scale (% difference in the means between the two groups) • Adverse events ○ Respiratory depression ○ Hypotension ○ Frequency of endotracheal intubation ○ ICU admission ○ Neuropsychological events such as confusion, anxiety, challenging behaviour, mood disturbance • Healthcare resource use
09	Intervention	What antiepileptic drugs (add-on therapy) are effective in the treatment of status epilepticus?	• Mortality (including SUDEP) • Time to seizure cessation, (5 min after drug administration, 10 min, 30 min, 60 min, less than or equal to 24 hours (convulsive), up to 1 month (non-convulsive)) • Time to event seizure cessation • Seizure recurrence greater than or less than 24 hours after administration of monotherapy • Time to seizure recurrence after administration of monotherapy • Quality of life (QOLIE-31, QOLIE-AD-48)

ISBN: 978-1-4731-4513-9

Evidence report	Type of review	Review questions	Outcomes

			• Length of ICU stay
			• Length of hospital stay
			• Mean Glasgow outcome scale (% difference in the means between the two groups)
			• Adverse events
			  ○ Respiratory depression
			  ○ Hypotension
			  ○ Frequency of endotracheal intubation
			  ○ ICU admission
			  ○ Neuropsychological events such as confusion, anxiety, challenging behaviour, mood disturbance
			• Healthcare resource use

10	Intervention	What ASMs (monotherapy) are effective in the treatment of repeated seizures or clusters of seizures?	• Mortality (including SUDEP)
			• Time to seizure cessation, within 24 hours after drug administration, 24 to 72 hours, greater than 72 hours 1 week
			• Time to event seizure cessation
			• Quality of life (QOLIE-31, QOLIE-AD-48)
			• Length of hospital stay
			• Adverse events
			  ○ Respiratory depression
			  ○ Hypotension
			  ○ Frequency of endotracheal intubation
			  ○ ICU admission
			  ○ Neuropsychological events such as confusion, anxiety, challenging behaviour, mood disturbance
			• Healthcare resource use

10	Intervention	What ASMs (add-on therapy) are effective in the treatment of repeated seizures or clusters of seizures?	• Mortality (including SUDEP)
			• Time to seizure cessation, within 24 hours after drug administration, 24 to 72 hours, greater than 72 hours 1 week
			• Time to event seizure cessation
			• Quality of life (QOLIE-31, QOLIE-AD-48)
			• Length of hospital stay
			• Adverse events
			  ○ Respiratory depression
			  ○ Hypotension
			  ○ Frequency of endotracheal intubation
			  ○ ICU admission
			  ○ Neuropsychological events such as confusion, anxiety, challenging behaviour, mood disturbance
			• Healthcare resource use

ISBN: 978-1-4731-4513-9

Evidence report	Type of review	Review questions	Outcomes

K2	Intervention	What ASMs (monotherapy) are effective in the treatment of prolonged seizures?*	• Mortality (including SUDEP)
			• Time to seizure cessation, within 24 hours after drug administration, 24 to 72 hours, greater than 72 hours 1 week
			• Time to event seizure cessation
			• Quality of life (QOLIE-31, QOLIE-AD-48)
			• Length of hospital stay
			• Adverse events
			  ○ Respiratory depression
			  ○ Hypotension
			  ○ Frequency of endotracheal intubation
			  ○ ICU admission
			  ○ Neuropsychological events such as confusion, anxiety, challenging behaviour, mood disturbance
			• Healthcare resource use

12	Intervention	What is the effectiveness of ketogenic diets in drug-resistant epilepsy?	• Seizure freedom (100% reduction in seizure frequency at study endpoint • Seizure frequency (50% or greater reduction in seizure frequency at study endpoint • Quality of life (as measured by validated scales) • Adverse events (all e.g., Diarrhoea / constipation / vomiting / renal stones (all GI heading)) at study endpoint • Attrition rate
13	Intervention	What is the clinical and cost effectiveness of different criteria for referral to epilepsy surgical services?	• Appropriateness of referral decisions
13	Intervention	What is the effectiveness of surgical intervention in epilepsy?	• Mortality at short-term follow-up of 12-24 months and longer-term follow-up of >24-60 months • Seizure freedom at short-term follow-up of 12 to 24 months and longer-term follow-up of >24-60 months • Due to anticipated heterogeneity in reporting of seizure freedom, data will be extracted as presented within included studies. Where a study reports multiple variants, then all data will be extracted. For decision making priority will be given to data based on hazards (of first seizure) rather than risks or odds. • Seizure frequency (50% or greater reduction in seizure frequency) at short-term follow-up of 12 to 24 months and longer-term follow-up of >24-60 month

ISBN: 978-1-4731-4513-9

Evidence report	Type of review	Review questions	Outcomes
			• Quality of life (measured with a validated scale) at short-term follow-up of 12 to 24 months and longer-term follow-up of >24-60 months • Healthcare resource use • Social functioning (measures of adaptive functioning or adaptive behaviour using a validated scale) short-term follow-up of 12 to 24 months and longer-term follow-up of >24-60 months • Cognitive outcomes (including neuropsychological measures of global cognitive functioning, executive functioning and memory using a validated scale) short-term follow-up of 12 to 24 months and longer-term follow-up of >24-60 months • In children and young people: neurodevelopmental outcomes (behavioural and emotional outcomes measured with a validated scale) short-term follow-up of 12 to 24 months and longer-term follow-up of >24-60 months • Serious adverse events (such as infection, stroke, severe bleeding)
14	Intervention	What is the effectiveness of vagus nerve stimulation in epilepsy?	• Mortality at short-term follow-up of 12 months and longer-term follow-up of up to 60 months • Seizure freedom (100% reduction in seizure frequency) at short-term follow-up of 12 months and longer-term follow-up of up to 60 months • Seizure frequency (50% or greater reduction in seizure frequency) at short-term follow-up of 12 months and longer-term follow-up of up to 60 months • Quality of life (measured with a validated scale) at short-term follow-up of 12 months and longer-term follow-up of 60 months • Healthcare resource use • Social functioning (measures of adaptive functioning or adaptive behaviour using a validated scale) short-term follow-up of 12 months and longer-term follow-up of up to 60 months • Cognitive outcomes (including neuropsychological measures of global cognitive functioning, executive functioning and memory using a validated scale) short-term follow-up of 12 months and longer-term follow-up of up to 60 months

ISBN: 978-1-4731-4513-9

Evidence report	Type of review	Review questions	Outcomes
			• In children and young people: neurodevelopmental outcomes (behavioural and emotional outcomes measured with a validated scale) short-term follow-up of 12 months and longer-term follow-up of up to 60 months • Adverse events (analysed separately): ○ Lead fracture ○ Infection ○ Hoarse voice ○ Cardiac difficulties ○ Device removal
15	Prevalence	Prevalence of depression, anxiety, learning disability and behavioural/cognitive difficulties in people with epilepsies	In people with Epilepsy, the prevalence of: • Depression • Anxiety • Learning disabilities • Cognitive difficulties • Dementia • psychosis
16	Intervention	What is the effectiveness of psychological treatments on HRQoL for people with epilepsy	Validated HRQoL outcomes
17	Risk prognostic test accuracy	What are the most accurate tools to predicting death, including SUDEP, in people with epilepsy?	• Discrimination: sensitivity, specificity, C statistic. These measures assess how accurately the tool can predict those who will and will not get SUDEP/die from any cause • Calibration: tests how well the tool results predict the absolute risk of getting SUDEP/dying from any cause • Net classification Improvement: a sensitive method for evaluating the different levels of predictive accuracy accruing from a change in the prediction tool.
18	Prognostic	What are the modifiable risk factors for epilepsy-related mortality, including SUDEP, and what is the magnitude of risk of the factors?	• Death, related to epilepsy • SUDEP
19	Intervention	What interventions are effective in reducing the risk of seizure-related mortality, including Sudden Unexpected Death in Epilepsy (SUDEP), in people with epilepsy?	• SUDEP at longest study follow-up • Total non-SUDEP seizure-related mortality (including other seizure-related causes such as accident-related mortality, status epilepticus-related mortality, and unexplained mortality) at longest study follow-up • Adverse events (total) at longest study follow-up

ISBN: 978-1-4731-4513-9

Evidence report	Type of review	Review questions	Outcomes
20	Qualitative	How should the transition from children's and young people's services to adults' epilepsy services be managed?	Themes will be derived from the evidence identified for this review and may include driving, teratogenicity of certain anti-epileptic medications, the interaction of anti-epileptic medications with contraception, the effect of alcohol/ recreational drugs on seizures, psychosocial aspects of epilepsy as people move out of the parental home e.g., to work, to University, independence. Memory, stigma, mental health.

** adjunct therapy for prolonged seizures was dropped from the protocol following discussion with the committee because the ILAE changes to the definition of prolonged seizures being between 2 and 5 minutes would mean the population would automatically fall into the category being covered by the review for the status epilepticus population.*

2.2 Searching for evidence

2.2.1 Clinical and health economics literature searches

The full 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.

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.[3] Databases were searched using relevant medical subject headings, free-text terms and study-type filters where appropriate. Studies published in languages other than English were not reviewed, and where possible, searches were restricted to English language. All searches were updated on 13 May 2021. If new evidence falls outside of the timeframe for the guideline searches, e.g., from stakeholder comments, the impact on the guideline will be considered, and any further action agreed between the developer and NICE staff with a quality assurance role.

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

During the scoping stage, a search was conducted for guidelines and reports on the websites, including:

- Guidelines International Network database (www.g-i-n.net)
- North American Guidelines (https://www.ahrq.gov/gam/index.html)
- NHS Evidence Search (www.evidence.nhs.uk).
- TRIP (www.tripdatabase.com)

Searching for unpublished literature was not undertaken.

ISBN: 978-1-4731-4513-9

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

2.3.1 Inclusion and exclusion criteria

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

The key population inclusion criterion was:

Children, young people and adults with suspected or confirmed epilepsy.

ISBN: 978-1-4731-4513-9

Specific consideration will be given to:

- children and young people
- girls and women who are able to get pregnant (including those who are pregnant or breastfeeding)
- older people
- people with learning disabilities

The key population exclusion criterion was:

- New-born babies (under 28 days) with acute symptomatic seizures.

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

2.3.1.1 Saturation of qualitative studies

Data extraction in qualitative reviews is a thorough process and may require more time compared to intervention reviews. It is common practice to stop extracting data once saturation has been reached. This is the point when no new information emerges from studies that match the review protocol. The remaining identified studies are, however, not directly excluded from the review as they nevertheless fit the criteria defined in the review protocol. Any studies for which data were not extracted due to saturation having been reached but that fit the inclusion criteria of the protocol were listed in the table for studies 'identified but not included due to saturation' in an appendix to the qualitative evidence review.

2.3.1.2 Prevalence review inclusion criteria

The pragmatic decision to only include systematic reviews for this review was made in anticipation of the search, otherwise generating an impractically large number of results to work with. Limiting to systematic reviews would enable the incorporation of synthesised prevalence evidence in a more time-sensitive manner.

2.3.2 Type of studies

Randomised trials, non-randomised intervention studies, and other observational studies (including diagnostic or prognostic studies) were included in the evidence reviews as appropriate.

For most intervention reviews in this guideline, parallel randomised controlled trials (RCTs) were included because they are considered the most robust type of study design that can produce an unbiased estimate of the intervention effects. If non-randomised intervention studies were considered appropriate for inclusion (for example, where no randomised evidence was available for critical outcomes), the committee stated a priori in the protocol that either certain identified variables must be equivalent at baseline or else the analysis had to adjust for any baseline differences. If the study did not fulfil either criterion, it was excluded. Please refer to the review protocols in each evidence report for full details on the study design of studies selected for each review question.

For diagnostic review questions, diagnostic RCTs, cross-sectional studies and retrospective studies were included. For prognostic review questions, prospective and retrospective cohort studies were included. Case-control studies were not included.

ISBN: 978-1-4731-4513-9

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 (not limited to):

- Mortality including SUDEP
- Medicine adherence
- Seizure frequency
- 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.

Where sufficient information was provided, hazard ratios were calculated in preference for outcomes such as mortality. where the time to the event occurring was important for decision-making.

Continuous outcomes

Continuous outcomes were analysed using an inverse variance method for pooling weighted mean differences. These outcomes included (not limited to):

- heath related quality of life (HRQoL)
- length of hospital stay

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

ISBN: 978-1-4731-4513-9

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 significant heterogeneity) as well as the distribution of effects. Where significant heterogeneity was present, predefined subgrouping of studies was carried as defined in each individual review question protocol.

If the subgroup analysis resolved heterogeneity within all of the derived subgroups, then each of the derived subgroups were adopted as separate outcomes (providing at least 1 study remained in each subgroup. For example, instead of the single outcome of '*missed diagnosis'*, this was separated into 2 outcomes '*missed diagnosis in people aged under 65'* and '*missed diagnosis in people aged 65 and over'*. Assessments of potential differences in effect between subgroups were based on the chi-squared tests for heterogeneity statistics between subgroups. Any subgroup differences were interpreted with caution, as separating the groups, breaks the study randomisation; and as such, is subject to uncontrolled confounding.

For some questions, additional subgrouping was applied, and this is documented in the individual review question protocols. These additional subgrouping strategies were applied independently, so subunits of subgroups were not created, unlike the situation with strata. Other subgrouping strategies were only used if the age category subgroup was unable to explain heterogeneity, then these further subgrouping strategies were applied in order of priority. Again, once a subgrouping strategy was found to explain heterogeneity from all derived subgroups, further subgrouping strategies were not used.

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

2.3.3.1.4 Complex analysis

Network meta-analysis was considered for the comparison of interventional treatments but was not pursued because of insufficient data available for the relevant outcomes.

ISBN: 978-1-4731-4513-9

Where studies had used a crossover design, paired continuous data were extracted where possible, and forest plots were generated in RevMan5[10] with the generic inverse variance function. When a crossover study had categorical data, and the number of subjects with an event in both interventions was known, the standard error (of the log of the risk ratio) was calculated using the simplified Mantel–Haenszel method for paired outcomes. Forest plots were also generated in RevMan5[10] with the generic inverse variance function. If paired continuous or categorical data were not available from the crossover studies, the separate group data were analysed in the same way as data from parallel groups, on the basis that this approach would overestimate the confidence intervals and thus artificially reduce study weighting resulting in a conservative effect. Where a meta-analysis included a mixture of studies using both paired and parallel-group approaches, all data were entered into RevMan5[10] using the generic inverse variance function.

2.3.3.2 Data synthesis for diagnostic test accuracy reviews

Two separate review protocols were produced to reflect the 2 different diagnostic study designs.

2.3.3.2.1 Diagnostic RCTs

Diagnostic RCTs (sometimes referred to as *test and treat trials*) are a randomised comparison of 2 diagnostic tests, with study outcomes being clinically important consequences of the diagnosis (patient-related outcome measures similar to those in intervention trials, such as mortality). Patients are randomised to receive test A or test B, followed by identical therapeutic interventions based on the results of the test [so everyone with a positive result would all receive the same treatment as each other (treatment X) regardless of whether they were diagnosed by test A or test B, and everyone with a negative result would also receive the same treatment as each other (treatment Y, which is usually no treatment) regardless of whether they were diagnosed by test A or test B]. Downstream patient outcomes are then compared between the 2 groups. As the selection of available treatments for both positive and negative tests is the same in both arms of the trial, any differences in patient outcomes will reflect the appropriateness of treatment choice (rather than the treatments themselves), which is, of course, a result of the accuracy of the tests in correctly establishing who does and does not have the condition. Diagnostic RCTs are, therefore a way of measuring the efficacy of diagnostic tests through their capacity to lead to appropriate management choices through accurate diagnosis. Data were synthesised using the same methods for intervention reviews (see section 2.3.3.1.1 above).

2.3.3.2.2 Diagnostic accuracy studies

Diagnostic accuracy studies measure how well a test can detect those people who truly have the condition and also how well the test can detect those people who truly do not have the condition. The true existence of the condition is determined by a gold standard test, which is regarded as infallible. A two by two table (Figure 1) contains all the information required to calculate diagnostic accuracy, with the data being counts of people and all cells being mutually exclusive and exhaustive. The two columns carry information about the gold standard results, and the two rows contain information about the test under investigation (the index test).

	Gold standard positive = truly have the condition	**Gold standard negative = truly do NOT have the condition**
Index test positive	98	22
Index test negative	2	178
Total	100	200

ISBN: 978-1-4731-4513-9

Figure 2: A two by two table for diagnostic accuracy

In the example above, there are 100 people defined by the gold standard as truly having the condition. Of these, 98 are correctly identified as having the condition by the index test (positive index test), so the sensitivity of the index test is 98/100 = 98%. There are also 200 people defined by the gold standard as truly NOT having the condition. Of these, 178 are correctly identified as not having the condition by the index test (negative index test), so the specificity of the index test is 178/200 = 89%.

In many diagnostic tests, the index test is based on a continuous, or ordinal measurement, and the test is designated positive if the test result is beyond a specific threshold on that continuous scale. The position of this threshold can be varied, and as the threshold changes, there is a trade-off between sensitivity and specificity. Assuming that higher values of the measurement are associated with the condition, a low threshold will tend to lead to more people testing positive because detection is triggered by all values *above* that threshold. A low threshold will thus have greater sensitivity, but because it may also tend to pick up people who don't have the condition, it will also lead to a lower specificity. In contrast, a high threshold may miss people who truly have the condition because it won't detect people with the condition who have a value below that high threshold. A higher threshold will therefore have lower sensitivity but will tend to pick out those who don't have the condition and so will have a high specificity. Plotting the sensitivities and specificities across these different thresholds yields the receiver operated characteristics (ROC) curve if specificity is plotted as 1-specificity, and the area under this curve provides an overall measure of accuracy over all thresholds. For this guideline, where the diagnostic accuracy study concerned the detection of epilepsy, the thresholds tended to be fixed, and multiple thresholds were not used. Therefore, only sensitivity and specificity at the fixed threshold were used, rather than ROC curves. If a test did use different thresholds, these were treated as separate tests.

For this guideline, where the diagnostic accuracy review concerned the detection of epilepsy, sensitivity and specificity were given equal priority, and minimum standards for recommendation were set at 0.9 for both. This was due to the consequences of failing to detect epilepsy (in someone who truly has it) being considered as serious as misdiagnosing someone as having epilepsy (when in reality they don't).

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. Test accuracy for the studies 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.

For scores with fewer than 3 studies, median sensitivity and the paired specificity were reported where possible. If an even number of studies were reported, the results of the study with the lower sensitivity value of the 2 middle studies was reported.

ISBN: 978-1-4731-4513-9

2.3.3.3 Data synthesis for prognostic factor reviews

To investigate the effects of *individual* baseline risk factors upon later outcomes such as second seizure or epilepsy-related death, odds ratios (ORs), risk ratios (RRs), or hazard ratios (HRs) [with their 95% CIs] for the effect of the prespecified prognostic risk factors upon the outcome were extracted from the studies. Studies were only included if the confounders prespecified by the committee were either matched at baseline or were adjusted for in the multivariate analysis. This ensured that the effects for the individual risk factors upon the outcome were *independent* – that is, that they could be assumed to be the causal effects of the risk factor upon the outcome, without being confounded by correlative effects from intervening variables. This information was of particular use for recommendations concerning prevention of the outcome – if a risk factor is known to cause the outcome, then the outcome may be prevented through the elimination of that risk factor. Individual risk factors may not be ideal for the prediction of outcomes such as a second seizure or SUDEP; however, as such outcomes are usually multifactorial, and so prognostic risk tools that utilise a combination of prognostic variables are more useful for that purpose (see next section).

Studies of lower risk of bias were preferred, taking into account the analysis and the study design. In particular, prospective cohort studies were preferred if they reported multivariable analyses that adjusted for key confounders identified by the committee at the protocol stage for that outcome.

Data were not combined in meta-analyses for prognostic studies.

2.3.3.4 Data synthesis for risk prognostic test accuracy reviews (also called 'risk prediction tools' or 'risk prediction rules').

2.3.3.4.1 Prognostic accuracy test studies

A prognostic accuracy test aims to accurately determine who will and who will not attain a particular prognostic outcome (for example, second seizure or epilepsy-related death) in the future. These tests usually utilise a combination of symptoms, signs, measures or characteristics that together can help to differentiate between people at risk and not at risk of the future event. This is analogous to a diagnostic test, which aims to accurately determine who has, and who does not have, a particular disease. The difference between a prognostic test and a diagnostic test is that whilst the diagnostic test measures the accuracy of detecting a current condition, the prognostic test measures the accuracy of predicting a later event (determining who actually gets the outcome or not). Therefore, while the gold standard for diagnostic tests is the best available method of diagnosis, the gold standard for prognostic tests is always the later measurement of the outcome. In the review for detection of second seizure, the later outcome was second seizure. For the review for detection of epilepsy-related death, the later outcome was SUDEP or other epilepsy-related death.

The key difference between a prognostic risk factor review (see the previous section) and a prognostic accuracy test (or risk prediction) review is that whilst the former attempts to identify individual risk factors for the outcome, so that those risk factors can be managed to prevent the outcome, prognostic accuracy tests aim to predict the risk of the outcome in individuals. Thus, the two kinds of review have different functions that support each other: the prognostic accuracy review identifies who is at risk and the risk tool review helps to determine what risk factors can be managed in those individuals who have been identified as at risk.

<u>C statistics</u>

In this guideline, the accuracy of different prediction tools was analysed at a variety of test thresholds within each study, and so areas under the ROC curve (AUC or 'C statistic') were useful measures of overall accuracy (see section 2.3.3.2.2). The AUC describes the overall

ISBN: 978-1-4731-4513-9

diagnostic accuracy across the full range of thresholds. The following criteria were used for evaluating AUCs:

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

C statistics across different studies were meta-analysed using the generic inverse variance option (for continuous variables) on RevMan. The derived forest plots were amended using the 'paint' program so that the null line was removed. Unlike the measures of effect in most meta-analyses, C statistics are not measures representing the differences or ratios between two groups and are instead a single group value (although the ultimate frame of reference is the gold standard). A null line indicating that there is 'no difference between groups' therefore has little meaning in this context.

Sensitivity and specificity

Sensitivity and specificity data were also collected for specific thresholds where available in the papers. This was necessary as prediction tools will be used clinically with specific thresholds, and so knowledge of accuracy at these specific thresholds is vital.

Sensitivity and specificity data for the prognostic reviews were not meta-analysed because these data were only available for some tools.

Calibration

Measures of calibration assess the ability of a risk prediction model to predict accurately the absolute level of risk that is subsequently observed. Calibration concerns how well the predicted risks compare to observed risks. A model is well-calibrated if, for every 100 patients given a prediction of p%, the observed number of events is close to p. Calibration is evaluated either by calculating the Hosmer-Lemeshow test statistic or preferably by plotting predicted risks against observed risks (calibration plot). This involves predicted outcome probabilities (on the x-axis) plotted against observed outcome frequencies (on the y-axis). A well-calibrated model shows predictions lying on or around the 45° line of the calibration plot; perfect calibration shows a slope of 1 and an intercept of 0, although some caveats have recently been identified. Other informative measures of model performance include the R^2 and the Brier score. R^2 characterizes the degree of variation in risk explained by the model. The adjusted R^2 has been proposed as a better measure, as it accounts for the number of predictors and helps to prevent overfitting. Brier scores are a similar measure of performance, which are used when the outcome of interest is categorical instead of continuous.

Calibration measures the accuracy of absolute risk prediction better than discrimination methods (such as C statistics or sensitivity/specificity). The absolute level of bleeding risk is what will be used clinically to allow the clinician and patient to make a shared decision on risk reduction through attention to modifiable risk factors for bleeding. Therefore, calibration was regarded as a particularly important measure of effect for the prediction of bleeding risk.

Calibration data were mostly synthesised using narrative methods because data were often presented graphically. However, where appropriate, data were meta-analysed.

2.3.3.5 Data synthesis for qualitative study reviews

The main findings for each included paper were identified, and thematic analysis methods were used to synthesise this information into broad overarching themes, which were

ISBN: 978-1-4731-4513-9

summarised into the main review findings. The evidence was presented in the form of a narrative summary detailing the evidence from the relevant papers and how this informed the overall review finding, plus a statement on the level of confidence for that review finding. Considerable limitations and issues around relevance were listed. A summary evidence table with the succinct summary statements for each review finding was produced, including the associated quality assessment.

2.3.3.6 Data synthesis for prevalence reviews

Prevalence statistics were reported as seen in the systematic reviews. Since the majority of the evidence was reported as prevalence percentages, a meta-analysis could not be carried out, and the evidence was presented narratively in a summary table with the risk of bias rating derived through the risk of bias ratings.

2.3.4 Appraising the quality of evidence by outcomes

2.3.4.1 Intervention reviews

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

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

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

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

ISBN: 978-1-4731-4513-9

Quality element	Description
	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 a risk of bias in just 1 domain, the risk of bias was given a 'serious' rating of −1, but if there was a 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: Principal 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 • 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.

ISBN: 978-1-4731-4513-9

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, 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 on 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

ISBN: 978-1-4731-4513-9

MID (no clinically important effect, clinical benefit and clinical harm). This is illustrated in Figure 3. As for inconsistency, since the imprecision score was based on the meta-analysis results, the score represented the whole outcome, and so weighted averaging across studies was not necessary.

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

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

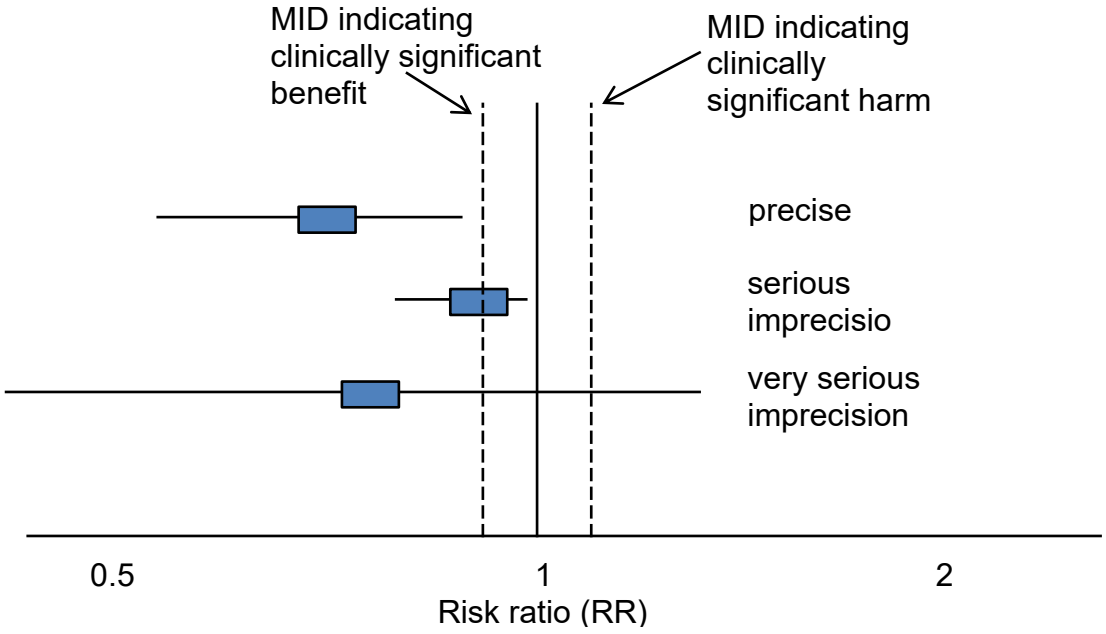
- For categorical outcomes, the MIDs were taken to be RRs of 0.8 and 1.25. For 'positive' outcomes such as 'patient satisfaction', the RR of 0.8 is taken as the line denoting the boundary between no clinically important effect and a clinically significant harm, whilst the RR of 1.25 is taken as the line denoting the boundary between no clinically important effect and a clinically significant benefit. For 'negative' outcomes such as 'bleeding', the opposite occurs, so the RR of 0.8 is taken as the line denoting the boundary between no clinically important effect and a clinically significant benefit, whilst the RR of 1.25 is taken as the line denoting the boundary between no clinically important effect and a clinically significant harm.
- For mortality any change was considered to be clinically important, and the imprecision was assessed on the basis of the whether the confidence intervals crossed the line of no effect, that is whether the result was consistent with both benefit and harm.
- For continuous outcome variables, the MID was taken as half the median baseline standard deviation of that variable, across all studies in the meta-analysis. Hence the MID denoting the minimum clinically significant benefit was positive for a 'positive' outcome (for example, a quality-of-life measure where a higher score denotes better health), and negative for a 'negative' outcome (for example, a visual analogue scale [VAS] pain score). Clinically significant harms will be the converse of these. If baseline values are unavailable, then half the median comparator group standard deviation of that variable will be taken as the MID.
- If standardised mean differences have been used, then the MID will be set at the absolute value of +0.5. This follows because standardised mean differences are mean differences normalised to the pooled standard deviation of the 2 groups, and are thus effectively expressed in units of, numbers of standard deviations. The 0.5 MID value in this context therefore indicates half a standard deviation, the same definition of MID as used for non-standardised mean differences.

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, no appropriate MIDs for continuous or dichotomous outcomes were found in the literature, and so the default method was adopted.

ISBN: 978-1-4731-4513-9

Figure 3: 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

ISBN: 978-1-4731-4513-9

2.3.4.2 Prognostic risk factor studies

2.3.4.2.1 Risk of bias

QUIPS was used to assess risk of bias in prognostic risk factor studies

Table 5: Description of quality elements for prospective studies using QUIPS

Quality element	Description of cases with high risk of bias	Description of cases with moderate risk of bias	Description of cases with low risk of bias
Study participation	The relationship between the PF and outcome is very likely to be different for participants and eligible nonparticipant	The relationship between the PF and outcome may be different for participants and eligible nonparticipants	The relationship between the PF and outcome is unlikely to be different for participants and eligible nonparticipants
Patient attrition	The relationship between the PF and outcome is very likely to be different for completing and non-completing participants	The relationship between the PF and outcome may be different for completing and non-completing participants	The relationship between the PF and outcome is unlikely to be different for completing and non-completing participants
Prognostic factor measurement	The measurement of the PF is very likely to be different for different levels of the outcome of interest	The measurement of the PF may be different for different levels of the outcome of interest	The measurement of the PF is unlikely to be different for different levels of the outcome of interest
Outcome measurement	The measurement of the outcome is very likely to be different related to the baseline level of the PF	The measurement of the outcome may be different related to the baseline level of the PF	The measurement of the outcome is unlikely to be different related to the baseline level of the PF
Study confounding	The observed effect of the PF on the outcome is very likely to be distorted by another factor related to PF and outcome	The observed effect of the PF on outcome may be distorted by another factor related to PF and outcome	The observed effect of the PF on outcome is unlikely to be distorted by another factor related to PF and outcome
Statistical analysis and reporting	The reported results are very likely to be spurious or biased related to analysis or reporting	The reported results may be spurious or biased related to analysis or reporting	The reported results are unlikely to be spurious or biased related to analysis or reporting

2.3.4.2.2 Inconsistency

This was not applicable as meta-analysis was not carried out.

2.3.4.2.3 Imprecision

The position of the 95% CIs in relation to the null line determined the existence of imprecision. If the 95% CI did not cross the null line, then no serious imprecision was recorded. If the 95% CI crossed the null line, then serious imprecision was recorded.

ISBN: 978-1-4731-4513-9

2.3.4.2.4 Overall grading

Quality rating started at High for prospective studies, and each major limitation brought the rating down by 1 increment to a minimum grade of Very Low, as explained for interventional reviews. For prognostic reviews, prospective cohort studies with a multivariate analysis are regarded as the gold standard because RCTs are usually inappropriate for these types of review for ethical or pragmatic reasons. Furthermore, if the study is looking at more than 1 risk factor of interest then randomisation would be inappropriate as it can only be applied to 1 of the risk factors.

2.3.4.3 Prognostic accuracy studies

2.3.4.3.1 Risk of bias

Risk of bias was initially assessed per study using the PROBAST tool.

PROBAST criteria were as follows:

- Appropriateness of data sources?
- Appropriateness of inclusion and exclusion criteria?
- Appropriate similarity of health across participants?
- Were predictors defined or assessed in the same way for all?
- Predictor assessments made without knowledge of outcome data?
- Predictors all available at time model meant to be used?
- All relevant predictors analysed?
- Pre-specified outcome used?
- Predictors excluded from outcome definition?
- Outcome defined in same way for all?
- Outcome determined without knowledge of predictor information?
- Reasonable number of outcome events? (100)
- Time interval between baseline and outcome appropriate? (5 years)
- All enrolled included in analysis?
- Missing data handled appropriately?
- Non-binary predictors handled appropriately?
- Complexities in data accounted for?
- Relevant performance measures?
- Model recalibrated or likely that calibration not needed?

Possible responses were not applicable, unclear, yes or no.

For each study risk of bias was downgraded by 1 (serious risk of bias) if blinding of assessors was not reported, and/or attrition bias (>10% loss) was suspected. Risk of bias was downgraded by 2 (very serious risk of bias) if the studies with the aforementioned limitations also had insufficient numbers of events (<100) and/or inappropriately short follow up times (<5 years) to be able to accurately predict risk.

An overall risk of bias rating was then pooled across studies covering the same outcome, using the meta-analysis weighting.

2.3.4.3.2 Indirectness

Indirectness was assessed by the extent to which the population, index test or outcome differed from the protocol definition. Indirectness was planned to be downgraded by 1 (serious risk of indirectness) if there was one departure from protocol, or by 2 (very serious risk of indirectness) if there were two or more departures from protocol. However, no studies were downgraded for indirectness.

ISBN: 978-1-4731-4513-9

2.3.4.3.3 Inconsistency

Where data were pooled, an I^2 of 50-74% was deemed serious inconsistency and an I^2 of 75% or above was deemed very serious inconsistency. If no pooling were possible, inconsistency was assessed by inspection of the degree of overlap of confidence intervals between studies: if one of more Cis did not overlap then a rating of serious inconsistency was given.

2.3.4.3.4 Imprecision

The judgement of precision was based on the spread of confidence intervals. For C statistic data, two clinical thresholds were used: AUCs of 0.5 and 0.7. The threshold of 0.5 marked the boundary between no predictive value better than chance and a predictive value better than chance. The threshold of 0.7 marked the boundary above which the committee might consider recommendations. If the 95% Cis crossed one of these thresholds, a rating of serious imprecision was given, and if they crossed both thresholds a rating of very serious imprecision was given.

For the NRI data if the lower 95% CI passed across 0 then this was graded as seriously imprecise.

For R^2 calibration data, if the upper 95% CIs were 20-40% greater than the point estimate then they were graded as seriously imprecise, and if the upper 95% CIs were >40% greater than the point estimate then they were graded as very seriously imprecise

2.3.4.3.5 Overall rating

Quality rating started at High for prospective and retrospective 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.

2.3.5 Prevalence review

2.3.5.1 Risk of bias

The risk of bias in systematic reviews (ROBIS) tool was uses to determine risk of bias ratings for the evidence included in the prevalence report. The ROBIS tool comprises of 3 phases: (1) assess relevance (optional), (2) identify concerns with the review process and (3) judge risk of bias in the review. Phases 2 and 3 have a subset of signalling questions to enable an accurate rating (see below). The rating from each phase is taken into consideration when determining the overall risk of bias rating.

Figure 4: Phase 1: Assessing relevance (optional)

Intervention reviews:

Category	Target question (e.g. overview or guideline)	Review being assessed
Patients/Population(s):		
Intervention(s):		
Comparator(s):		
Outcome(s):		

ISBN: 978-1-4731-4513-9

For aetiology reviews:

Category	Target question (e.g. overview or guideline)	Review being assessed
Patients/Population(s):		
Exposure(s) and comparator(s):		
Outcome(s):		

For DTA reviews:

Category	Target question (e.g. overview or guideline)	Review being assessed
Patients):		
Index test(s):		
Reference standard:		
Target condition:		

For prognostic reviews:

Category	Target question (e.g. overview or guideline)	Review being assessed
Patients:		
Outcome to be predicted:		
Intended use of model:		
Intended moment in time:		

Source: Tables taken from http://www.bristol.ac.uk/population-health-sciences/projects/robis/

Figure 5: Phase 2: identifying concerns with review process

DOMAIN 1: STUDY ELIGIBILITY CRITERIA	
Describe the study eligibility criteria, any restrictions on eligibility and whether there was evidence that objectives and eligibility criteria were pre-specified:	
1.1 Did the review adhere to pre-defined objectives and eligibility criteria?	Y/PY/PN/N/NI
1.2 Were the eligibility criteria appropriate for the review question?	Y/PY/PN/N/NI
1.3 Were eligibility criteria unambiguous?	Y/PY/PN/N/NI
1.4 Were any restrictions in eligibility criteria based on study characteristics appropriate (e.g. date, sample size, study quality, outcomes measured)?	Y/PY/PN/N/NI
1.5 Were any restrictions in eligibility criteria based on sources of information appropriate (e.g. publication status or format, language, availability of data)?	Y/PY/PN/N/NI
Concerns regarding specification of study eligibility criteria	LOW/HIGH/UNCLEAR
Rationale for concern:	

ISBN: 978-1-4731-4513-9

DOMAIN 2: IDENTIFICATION AND SELECTION OF STUDIES	
Describe methods of study identification and selection (e.g. number of reviewers involved):	
2.1 Did the search include an appropriate range of databases/electronic sources for published and unpublished reports?	Y/PY/PN/N/NI
2.2 Were methods additional to database searching used to identify relevant reports?	Y/PY/PN/N/NI
2.3 Were the terms and structure of the search strategy likely to retrieve as many eligible studies as possible?	Y/PY/PN/N/NI
2.4 Were restrictions based on date, publication format, or language appropriate?	Y/PY/PN/N/NI
2.5 Were efforts made to minimise error in selection of studies?	Y/PY/PN/N/NI
Concerns regarding methods used to identify and/or select studies Rationale for concern:	LOW/HIGH/UNCLEAR

DOMAIN 3: DATA COLLECTION AND STUDY APPRAISAL	
Describe methods of data collection, what data were extracted from studies or collected through other means, how risk of bias was assessed (e.g. number of reviewers involved) and the tool used to assess risk of bias:	
3.1 Were efforts made to minimise error in data collection?	Y/PY/PN/N/NI
3.2 Were sufficient study characteristics available for both review authors and readers to be able to interpret the results?	Y/PY/PN/N/NI
3.3 Were all relevant study results collected for use in the synthesis?	Y/PY/PN/N/NI
3.4 Was risk of bias (or methodological quality) formally assessed using appropriate criteria?	Y/PY/PN/N/NI
3.5 Were efforts made to minimise error in risk of bias assessment?	Y/PY/PN/N/NI
Concerns regarding methods used to collect data and appraise studies Rationale for concern:	LOW/HIGH/UNCLEAR

DOMAIN 4: SYNTHESIS AND FINDINGS	
Describe synthesis methods:	
4.1 Did the synthesis include all studies that it should?	Y/PY/PN/N/NI
4.2 Were all pre-defined analyses reported or departures explained?	Y/PY/PN/N/NI
4.3 Was the synthesis appropriate given the nature and similarity in the research questions, study designs and outcomes across included studies?	Y/PY/PN/N/NI
4.4 Was between-study variation (heterogeneity) minimal or addressed in the synthesis?	Y/PY/PN/N/NI
4.5 Were the findings robust, e.g. as demonstrated through funnel plot or sensitivity analyses?	Y/PY/PN/N/NI
4.6 Were biases in primary studies minimal or addressed in the synthesis?	Y/PY/PN/N/NI
Concerns regarding the synthesis and findings Rationale for concern:	LOW/HIGH/UNCLEAR

Source: Tables taken from http://www.bristol.ac.uk/population-health-sciences/projects/robis/

ISBN: 978-1-4731-4513-9

Figure 6: Phase 3: Judging risk of bias

RISK OF BIAS IN THE REVIEW	
Describe whether conclusions were supported by the evidence:	
A. Did the interpretation of findings address all of the concerns identified in Domains 1 to 4?	Y/PY/PN/N/NI
B. Was the relevance of identified studies to the review's research question appropriately considered?	Y/PY/PN/N/NI
C. Did the reviewers avoid emphasizing results on the basis of their statistical significance?	Y/PY/PN/N/NI
Risk of bias in the review	RISK: LOW/HIGH/UNCLEAR
Rationale for risk:	

Source: Table taken from http://www.bristol.ac.uk/population-health-sciences/projects/robis/

2.3.6 Assessing clinical importance

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

The assessment of clinical benefit, harm, or no benefit or harm was based on the point estimate of absolute effect for intervention studies, which was standardised across the reviews. The committee considered for most of the outcomes in the intervention reviews that if at least 100 more participants per 1000 (10%) achieved the outcome of interest in the intervention group compared to the comparison group for a positive outcome then this intervention was considered beneficial. The same point estimate but in the opposite direction applied for a negative outcome. For the critical outcome of mortality any reduction represented a clinical benefit. For adverse events 50 events or more per 1000 (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. For outcomes such as mortality any reduction or increase was considered to be clinically important.

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).

The quality of evidence for prognostic studies was evaluated according to the criteria given in Table 5. If data were meta-analysed, the quality for pooled studies was presented. If the data were not pooled, then a quality rating was presented for each study.

2.3.6.1 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.[3] Risk of bias and applicability in primary diagnostic accuracy studies in QUADAS-2 consists of 4 domains (see Figure 7):

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

ISBN: 978-1-4731-4513-9

Figure 7: 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 avoided?	If a threshold was used, was it pre-specified?	Were the reference standard results interpreted without knowledge of the results of the index test?	Did all patients receive a reference standard?
	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?	

2.3.6.1.1 Inconsistency

Inconsistency refers to an unexplained heterogeneity of results for an outcome across different studies. Inconsistency was assessed by visual inspection of the sensitivity and specificity plots. If there were any studies with 95% CIs that did not overlap with any other, then a rating of serious inconsistency was given. For tools with only single studies no inconsistency rating was given.

ISBN: 978-1-4731-4513-9

2.3.6.1.2 Imprecision

The judgement of precision was based on the position of the 95% confidence intervals for sensitivity and specificity relative to two clinical thresholds at 0.60 and 0.90. The 0.60 threshold represented the threshold accuracy below which the tool would not be clinically useful, and the 0.90 threshold represented the threshold above which the tool might be recommended. Serious imprecision was recorded if the 95% CIs crossed one of these clinical thresholds, and very serious imprecision was recorded if the 95% CIs crossed both clinical thresholds.

If a meta-analysis was undertaken the 95% CIs of the summary sensitivity/specificity was used. If only 2 studies were available, then the 95% CIs of the median sensitivity value and paired specificity value were used. If only 1 study was available, the 95% CI of the single sensitivity and specificity values were used.

2.3.6.1.3 Overall grading

Quality rating started at High for prospective and retrospective 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.

2.3.6.2 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 6.

Table 6: 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.

2.3.6.2.1 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:

ISBN: 978-1-4731-4513-9

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

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

2.3.6.2.2 Coherence

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

2.3.6.2.3 Relevance

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

2.3.6.2.4 Adequacy

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

2.3.6.2.5 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

ISBN: 978-1-4731-4513-9

significance of these overall ratings is explained in Table 7. 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 7: 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.

2.3.7 Assessing clinical importance

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

The assessment of clinical benefit, harm, or no benefit or harm was based on the point estimate of absolute effect for intervention studies, which was standardised across the reviews. The committee considered for most of the outcomes in the intervention reviews that if at least 100 more participants per 1000 (10%) achieved the outcome of interest in the intervention group compared to the comparison group for a positive outcome then this intervention was considered beneficial. The same point estimate but in the opposite direction applied for a negative outcome. For the critical outcome of mortality any reduction represented a clinical benefit. For adverse events, 50 events or more per 1000 (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. For outcomes such as mortality any reduction or increase was considered to be clinically important.

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).

2.4 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

ISBN: 978-1-4731-4513-9

uncertainties must be offset by a compelling argument in favour of the recommendation. The cost impact or savings potential of a recommendation should not be the sole reason for the committee's decision.[3]

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

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

2.4.1 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.[3]
- Extracted key information about the studies' methods and results into health economic evidence tables (which can be found in appendices to the relevant evidence reports).
- Generated summaries of the evidence in NICE health economic evidence profile tables (included in the relevant evidence report for each review question) – see below for details.

2.4.1.1 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 2004 and studies from non-OECD countries or the USA were also excluded, on the basis that the applicability of such studies to the present UK NHS context is likely to be too low for them to be helpful for decision-making.

The 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, this 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 8 below and the economic evaluation checklist (appendix H of the NICE guidelines manual[3]) and the health economics review protocol, which can be found in each of the evidence reports.

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

ISBN: 978-1-4731-4513-9

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.[3] It also shows the incremental costs, incremental effects (for example, quality-adjusted life years [QALYs]) and incremental cost-effectiveness ratio (ICER) for the base case analysis in the study, as well as information about the assessment of uncertainty in the analysis. See Table 8 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 8: Content of NICE health economic evidence profile

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

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

2.4.2 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

ISBN: 978-1-4731-4513-9

selected areas. Priority areas for new analysis were agreed by the committee after the formation of the review questions and consideration of the existing health economic evidence.

The committee noted the significant clinical benefits surgery could have for people with drug refractory epilepsy, however they noted that there are also significant resource costs. The committee were also aware there is little cost effectiveness evidence for resective epilepsy surgery, therefore it identified this as the highest priority for modelling.

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

- Methods were consistent with the NICE reference case for interventions with health outcomes in NHS settings.[3, 6]
- The committee was involved in the design of the model, selection of inputs and interpretation of the results.
- Model inputs were based on the systematic review of the clinical literature supplemented with other published data sources where possible.
- When published data were not available committee expert opinion was used to populate the model.
- Model inputs and assumptions were reported fully and transparently.
- The results were subject to sensitivity analysis and limitations were discussed.
- The model was peer-reviewed by another health economist at the NGC.

Full methods and results of the cost-effectiveness analysis for resective epilepsy surgery are presented in the economic analysis report.

2.4.3 Cost-effectiveness criteria

NICE sets out the principles that committees should consider when judging whether an intervention offers good value for money.[3-5] 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 factors set out in NICE methods manuals.[3]

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

2.4.4 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.

ISBN: 978-1-4731-4513-9

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–B]).
- 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(ses) undertaken for the guideline (in a separate economic analysis report).

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

When clinical and health economic evidence was of poor quality, conflicting or absent, the committee drafted recommendations based on its expert opinion. The considerations for making consensus-based recommendations include the balance between potential harms and benefits, the economic costs compared to the economic benefits, current practices, recommendations made in other relevant guidelines, patient preferences and equality issues. The consensus recommendations were agreed through discussions in the committee, or methods of formal consensus [insert method of consensus] were applied. 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 effects 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.

ISBN: 978-1-4731-4513-9

- The strength of the recommendation (for example the word 'offer' was used for strong recommendations and 'consider' for weaker recommendations).
- The involvement of patients (and their carers if needed) in decisions on treatment and care.
- Consistency with NICE's standard advice on recommendations about drugs, waiting times and ineffective interventions (see section 9.2 in the NICE guidelines manual[3]).

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

2.5.1 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[3], 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.

ISBN: 978-1-4731-4513-9

3 Additional information

3.1 MHRA review

A decision was made to incorporate into the guideline the guidance published by the Medicines and Healthcare products Regulatory Agency (MHRA) (2021) Public Assessment Report: Antiepileptic drugs: review to investigate the safety of ASMs in women and girls with epilepsy (see chapter 06). Therefore, a separate evidence review was not conducted by the National Guideline centre. The report reviewed available safety data relating to the use of the main ASMs in pregnancy for the risk of major congenital malformations, neurodevelopmental disorders and delay, and other effects on the baby. However, the report does not address the safety of ASMs in women and girls with epilepsy who are breastfeeding which was also a population of interest highlighted by the guideline committee. The committee therefore made consensus recommendations for this population.

The MHRA did not provide methodology data for the report, in the absence of which we were unable to derive quality ratings for the evidence included. The ROBIS tool for assessing risk of bias in systematic reviews (see 2.3.5.1) was applied to the report, and the ratings considered when discussing the evidence with the guideline committee.

3.2 Cochrane collaboration

The National Guideline Centre collaborated with Cochrane for 2 areas of this guideline:

- *What is the effectiveness of ketogenic diets in drug-resistant epilepsy?*
- *What is the effectiveness of psychological treatments on HRQoL for people with epilepsy?*

An overlap was identified between the Cochrane systematic reviews and the review questions within the draft NICE Epilepsies guideline scope. Commissioning briefs were developed with NICE for the Cochrane review group to update the systematic reviews for both areas. The NGC technical team and the Epilepsies guideline committee worked with the Cochrane group to finalise the review protocols and the systematic reviews were updated by the respective Cochrane teams and incorporated within the guideline. For summaries of the evidence see chapter 12 and chapter 16.

3.3 Investigating Prevalence

Review question: Prevalence of depression, anxiety, learning disability and behavioural/cognitive difficulties in people with epilepsies.

The scope for this guideline set out to investigate the prevalence of psychological disorders, neurodevelopmental and cognitive disorders, and behavioural disorders in people with epilepsy. The technical team, with support from the guideline committee, recognised the difficulty in addressing such a vast area in a limited timeframe. It was therefore agreed with NICE to refine the review question to investigating the prevalence of depression, anxiety, learning disability and behavioural/cognitive difficulties in people with epilepsies, as these were the comorbidities the committee recognised as priority for clinical awareness (see chapter 15).

The pragmatic decision to only include systematic reviews for this review was made. The committee agreed that despite narrowing down the review question, the evidence search would still generate an impractically large number of results to work with. Limiting to systematic reviews would enable incorporation of already synthesised evidence.

ISBN: 978-1-4731-4513-9

4 Updating the guideline

Following publication, and in accordance with the Developing 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.

ISBN: 978-1-4731-4513-9

5 Terms used in the guideline

5.1 Acronyms and abbreviations

Acronym or abbreviation	Description
ACTH	Adrenocorticotropic Hormone
ADNFLE	Autosomal Dominant Nocturnal Frontal Lobe Epilepsy
APE2 score	Antibody Prevalence in Epilepsy and Encephalopathy 2 score
ARS	Acute Repetitive Seizures
ASM	Anti-seizure medication
BECTS	Benign epilepsy with centrotemporal spikes former term of reference for what is now referred to as (~SLECTS) Self Limited Epilepsy with Centrotemporal Spikes
BNF	British National Formulary
BYI_II	Beck Anxiety Inventory for Youth
CAE	Childhood Absence Epilepsy
CESS	Children's Epilepsy Surgery Service
CFM	Cerebral Function Monitor
CGI	Clinical Global Impressions Scale
CHF	Congestive heart failure
CNS	Central Nervous System
CSE	Convulsive Status Epilepticus
CSWS	Continuous spike and wave during slow sleep, former term of reference now referred to as Developmental and Epileptic Encephalopathy with spike activation in sleep
CT/MRI	Computed Tomography/Magnetic Resonance Imaging
CVD	Cardiovascular disease
DNAR	Do Not Attempt Resuscitation
DSM III	Diagnostic and Statistical Manual of Mental Disorders
EEG	Electroencephalogram
EMG	Electromyography
EMSE	Epidemiology-based Mortality Score in Status Epilepticus
FCSE	Focal convulsive status epilepticus
FDG-PET	Fluorodeoxyglucose-Positron Emission Tomography
FIRES	Febrile infection-related epilepsy syndrome
FLE	Frontal lobe epilepsy
FLEP	Frontal Lobe Epilepsy and Parasomnias scale
GCS	Glascow Comma Score
GCSE	Generalised convulsive status epilepticus
GFAP	Glial Fibrillary Acidic Protein
GTCS	Generalized Tonic–Clonic Seizure
HMPAO PSECT	Hexamethylpropylene Amine Oxime Single-Photon Emission Computed Tomography
ICTRP	International Clinical Trials Registry Platform
IGE	Immunoglobulin E
IGE	Idiopathic generalised epilepsy
JAE	Juvenille Absence Epilepsy

ISBN: 978-1-4731-4513-9

Acronym or abbreviation	Description
JME	Juvenile Myoclonic epilepsy
KD	Ketogenic Diet
LGIT	Low Glycaemic Index Treatment
LKS	Landau–Kleffner syndrome: CSWS (Developmental and Epileptic Encephalopathy with Spike Activation in Sleep)
LOC	Loss of Consciousness
LTM	Long-Term Monitoring
MAD	Modified Atkins Diet
MASC 2	Multidimensional Anxiety Scale for Children
MCT	Medium-Chain Triglyceride
MEG	Magnetoencephalography
MEMS	Medication Event Monitoring System
MESS	Multicentre trial for Early Epilepsy and Single Seizures
MGLS	Morisky Green and Levine Scale
MPR	Medication Possession Ratio
MTLE	Mesial Temporal Lobe Epilepsy
NCSE	Nonconvulsive Status Epilepticus
NDDI-E	Neurological Disorders Depression Inventory for Epilepsy
NEAD	
NFLE	Nocturnal frontal lobe epilepsy
NORSE	New Onset Refractory Status Epileptics
OPCS	Oligodendrocyte Precursor Cells
PET	Positron Emission Tomography
PNES	Psychogenic Nonepileptic Seizures
PoSERS	Post Stroke Epilepsy Risk Scale
PSG	Polysomnography
PSSi	Post-Ischemic Stroke Seizure
RAVLT	Rey Auditory Verbal Learning Test
RCMAS-2	Revised Children's Manifest Anxiety Scale-Second Edition
RSE	Refractory status epilepticus
RTLE	Right-Sided Temporal Lobe Epilepsy
SCARED	Self-Report for Childhood Anxiety Related Emotional Disorders
SCAS	Spence Children's Anxiety Scale
SE	Status epilepticus
SGPT	Serum Glutamic Pyruvic Transaminase
SLECTS	Self-Limited Epilepsy with Centrotemporal Spikes formerly referred to as Benign epilepsy with centrotemporal spikes (BECTS)
SPECT	Single-Photon Emission Computerized Tomography
STAIC	Self-Trait Anxiety Inventory for Children
SUDEP	Sudden Unexpected Death in Epilepsy
TCI	Transitory cognitive impairment
TCS	Tonic-Clonic Seizure
TDM	Therapeutic Drug Monitoring
TIA	Transient Ischaemic Attack
TLE	Temporal Lobe Epilepsy

ISBN: 978-1-4731-4513-9

Acronym or abbreviation	Description
TMS	Transcranial Magnetic Stimulation
t-VNS	Transcutaneous Vagus Nerve Stimulation
VEEG	Video electroencephalography
VNS	Vagus Nerve Stimulation
WMH CIDI	World Mental Health Composite International Diagnostic Interview

5.2 Glossary

5.2.1 Guideline-specific terms

Term	Definition
Absence seizure	An incident where an individual loses awareness of their surroundings for a short time. Usually only lasting up to 15 seconds. They can happen several times per day. They mainly affect children, but can happen at any age.
Acidosis	When body fluids contain too much acid; this condition occurs when the kidneys and lungs can't maintain the body's pH balance.
Acute otitis media	The presence of inflammation in the middle ear, resulting in the rapid onset of an ear infection.
Anti-seizure medication	Medication taken daily to prevent the recurrence of epileptic seizures. Refer to the BNF or BNFC concerning the choice of drug, side effects and suitability to syndrome. Formerly referred to as anti-epileptic drugs.
Atonic seizure	A type of seizure that causes loss of muscle strength. These seizures are sudden and can cause individuals to collapse, but not lose consciousness.
Benign epilepsy with centrotemporal spikes (BECTS)	An epilepsy syndrome of childhood (5–14 years) characterised by focal motor and/or secondarily generalised seizures, the majority from sleep, in an otherwise normal individual, with centrotemporal spikes seen on EEG. Now referred to as SLECTS:
Bronchopneumonia	A type of pneumonia that causes inflammation in the lungs.
Callosotomy	A surgery used to treat 'drop' epileptic seizures when antiseizure medications don't work. The procedure involves cutting a band of fibres (the corpus callosum) in the brain, resulting in the nerves being unable to send seizure signals between the brain's two halves.
Cerebrovascular accident	Also known as a stroke, this occurs when blood flow to a part of the brain is stopped either by a blockage or the rupture of a blood vessel.
Childhood absence epilepsy	An epilepsy syndrome with an age of onset of 4–9 years, characterised by frequent absence seizures associated with 3 Hz spike wave activity on EEG.
Childhood occipital visual epilepsy syndrome (formerly Late-onset childhood occipital epilepsy (Gastaut type))	Epilepsy with an age of onset in mid-childhood to adolescence with frequent brief seizures characterised by initial visual hallucinations, ictal blindness, vomiting and post-ictal headache. EEG typically shows interictal occipital spikes attenuated by eye opening.
Complex febrile convulsion	A febrile convulsion is defined as a seizure occurring with fever in a child with no history of afebrile seizures age 6months and 3 years in the absence of a meningitis or encephalitis. They are complex if they last longer than 15 minutes. Or only involve one part of their body (this is known as focal seizure). If prolonged, they can be associated with increased risk for epilepsy in later life. Initial treatment is to make sure child is safe, call ambulance if > 5 minutes etc.
Complex needs	A person is described as having complex needs if they have been

ISBN: 978-1-4731-4513-9

Term	Definition
	diagnosed with an illness, disability or sensory impairment and needs high levels of additional support daily; often relying on a range of health and social care services.
Continuous spike and wave during slow sleep (CSWS)	An older term used in Epilepsy literature. This is an epilepsy syndrome with childhood onset, characterised by a plateau and regression of cognitive abilities associated with dramatic increase in spike wave activity in slow wave sleep (> 85% of slow sleep). There may be few seizures at presentation. Now called Developmental and Epileptic Encephalopathy with Spike Activation in Sleep formerly.
Convulsive seizure	A bilateral tonic clonic seizure with loss of consciousness.
Convulsive status epilepticus	When a convulsive seizure continues for a prolonged period (longer than 5 minutes), or when convulsive seizures occur one after the other with no recovery between them. Convulsive status epilepticus is an emergency and requires immediate medical attention.
Disconnective surgery	A surgical treatment for epilepsy where the individual's nerve fibres between hemispheres within the brain are cut to interrupt the spread of seizures from one part of the brain to another.
Doose syndrome	Also called Myoclonic-Atonic Epilepsy (MAE); this is a childhood onset epilepsy syndrome onset 2-6 years characterised by myoclonic atonic seizures and generalised tonic clonic seizures.
Dravet syndrome	Previously known as severe myoclonic epilepsy of infancy. An epilepsy syndrome with onset in infancy, characterised at onset by initial prolonged, typically lateralised, febrile seizures, with subsequent development of multiple seizure types including myoclonic, absence, focal and generalised tonic–clonic seizures, and developmental plateau.
Drug resistant	Epilepsy in which seizures persist despite an adequate trial of anti-seizure medications. Drug resistant epilepsy is formally defined as 'failure of adequate trials of 2 tolerated and appropriately chosen and used antiseizure medication schedules (whether as monotherapy or in combination) to achieve sustained seizure freedom'. (International League Against Epilepsy,
Dysphagia	Swallowing difficulties.
Electrocardiogram (ECG)	A test that records the heart's electrical activity.
Electroencephalogram (EEG)	An investigation that involves recording the electrical activity of the brain through electrodes are attached to standardised points on the person's head.
Endotracheal intubation	A medical procedure in which a tube is placed into the windpipe (trachea) through the mouth or nose.
Extratemporal resective surgery	Type of focal resection to remove a small part of the brain not within the temporal lobes.
Focal secondary generalised	Focal to bilateral tonic clonic or 'secondarily generalized' seizures begin in one part of the brain and then spread to both sides of the brain. In other words, the person first has a focal seizure, followed by a generalised seizure.
Focal seizure	A seizure that originates within networks limited to one hemisphere, discretely localised or more widely distributed. Replaces the terms partial seizure and localisation-related seizure.
Generalised seizure	A seizure that originates in, and rapidly engages, bilaterally distributed networks. Such bilateral networks can include cortical and subcortical structures but do not necessarily include the entire cortex.
Generalised tonic-clonic (GTC) seizure	A seizure of sudden onset involving generalised stiffening and subsequent rhythmic jerking of the limbs, the result of rapid widespread engagement of bilateral cortical and subcortical networks in the brain.

ISBN: 978-1-4731-4513-9

Term	Definition
Hemispherectomy	One side of the brain is either partly or totally removed from the rest of the brain. It is a surgical procedure done for epilepsy where that side of the brain is abnormal and causing seizures not responsive to medications. It is typically done in children and occasionally in adults.
Hemispherotomy	An operation that disconnects one half of the brain. (or hemisphere) from the other without removing it.
Haemorrhagic	Accompanied by or produced by haemorrhage.
Hyperammonaemic encephalopathy	An unusual complication characterized by a decreasing level of consciousness, focal neurological deficits, cognitive slowing, vomiting, drowsiness, and lethargy associated with high blood ammonia. This can be seen in people taking sodium valproate and does not necessarily associate with high valproate levels or deranged liver function tests.
Hypercalcinaemia	A high calcium level in the blood.
Hypercalcuria	Excess calcium in the urine
Hypercholesterolaemia	A high cholesterol concentration in the blood.
Hypotension	Low blood pressure.
Hypothalamic hamartoma	A rare, benign (noncancerous) brain tumour or lesion of the hypothalamus (small region of the brain located at the base of the brain, near the pituitary gland).
Ictal phenomenology	Description or history of ictal events (seizures).
Idiopathic generalised epilepsy (IGE)	A well-defined group of disorders characterised by typical absences, myoclonic and generalised tonic–clonic seizures, alone or in varying combinations in otherwise normal individuals. The EEG is also characteristic, demonstrating a distinct pattern of generalised polyspike wave discharges and/or generalised spike wave. Presumed to have a genetic aetiology. This terminology specifically refers to the group of four epilepsy syndromes Childhood Absence Epilepsy, Juvenile Absence Epilepsy, Juvenile Myoclonic Epilepsy and Epilepsy with Generalised Tonic Clonic seizures only.
Infantile Spasms Syndrome	Previously known as, West Syndrome, this is a group of symptoms characterized by epileptic or infantile spasms, abnormal brain wave patterns called hypsarrhythmia and intellectual disability.
Juvenile absence epilepsy (JAE)	An epilepsy syndrome with an age of onset of 9–13 years characterised by absence seizures, associated with 3–4 Hz spike wave on EEG. Generalised tonic–clonic seizures may occur.
Juvenile myoclonic epilepsy (JME)	An epilepsy syndrome with an age of onset of 5–20+ years (peak 10–16 years) characterised by myoclonic seizures that most commonly occur soon after waking. Absence and generalised tonic–clonic seizures may occur in between 50 and 80% of individuals with JME. EEG demonstrates 3–6 Hz generalised polyspike and wave activity, with photosensitivity in more than 30% of individuals.
Ketogenic diet	A specific diet that is high in fat but low in carbohydrates and protein.
Landau–Kleffner syndrome (LKS)	Landau Kleffener syndrome formerly related to a very rare epilepsy syndrome with an age of onset of 3–6 years characterised by loss of language (after a period of normal language development) associated with an epilepsy of centrotemporal origin, more specifically bitemporal spikes on EEG with enhancement in sleep or continuous spike and wave during slow sleep. Also referred to as Developmental and Epileptic Encephalopathy with spike activation in sleep.
Lennox-Gastaut syndrome	A severe form of epilepsy that typically becomes apparent during infancy or early childhood. Affected children experience several different types of seizures most commonly atonic, tonic and atypical absence seizures. Affected individuals have cognitive dysfunction and EEG findings can be characteristic with slow spike and wave activity.

ISBN: 978-1-4731-4513-9

Term	Definition
Meningitis/encephalitis	Meningitis is an infection of the meninges, the membranes that surround the brain and spinal cord. Encephalitis is inflammation of the brain itself.
Mesial TLE (MTLE)	Mesial temporal lobe epilepsy (MTLE) involves the medial or internal structures of the temporal lobe. Seizures often begin in a structure of the brain called the hippocampus or surrounding area. MTLE accounts for almost 80% of all temporal lobe seizures.
Messial temporal sclerosis	Also commonly referred to as hippocampal sclerosis, is the most common association with intractable temporal lobe epilepsy (TLE) in adults
MRI protocols	An MRI scan produces sets of images of the brain, or 'sequences', each of which provides specific information. An epilepsy MRI protocol is made up of a group of sequences, put together to improve the sensitivity and specificity in demonstrating possible structural abnormalities of the brain which cause epilepsy. The use of a regionally agreed standardised protocols aims to maximise diagnostic quality and deliver consistency in scan quality.
Myoclonic atonic epilepsy	(Doose syndrome) Also called Myoclonic-Atonic Epilepsy (MAE); this is a childhood onset epilepsy syndrome onset 2-6 years characterised by myoclonic atonic seizures and generalised tonic clonic seizures.
Myoclonic seizures	Sudden brief (less than 100 ms) and almost shock-like involuntary single or multiple jerks due to abnormal excessive or synchronous neuronal activity. Can be associated with polyspikes on EEG.
Nephrocalcinosis	A disorder that occurs when too much calcium is deposited in the kidneys.
Neuro-behavioural comorbidities	The additional health challenges which result from reduced nerve and brain function. This could include: cognitive impairment, psychiatric disorders and social problems all of which directly impact a person's diagnosis, management and access to care.
Neurocysticercosis	A preventable parasitic infection caused by larval cysts (enclosed sacs containing the immature stage of a parasite) of the pork tapeworm (Taenia solium). The larval cysts can infect various parts of the body causing a condition known as cysticercosis. When in the brain it is neurocysticercosis
Neurodevelopmental comorbidities	Neurodevelopmental disorders are the additional health challenges which result from impairment of the growth and development of the brain and/or central nervous system. This could include: attention-deficit hyperactivity disorder, or autism spectrum disorder. These disorders may contribute to cognitive impairment, affecting a person's intellectual functioning, reading ability, social skills, memory, attention or focus skills, and have implications for choice of treatment, diagnosis, management and access to care.
Neurodevelopmental disorders	A specific language delay or disorder, a learning (intellectual) disability or global developmental delay, a developmental coordination disorder. We have cross referred to the Autism spectrum disorders in under 19s and autism spectrum disorders in adults where further detail is provided.
Non-convulsive status epilepticus	A change in mental status or behaviour from baseline, associated with continuous seizure activity on EEG, which is also seen to be a change from baseline.
Non convulsive prolonged seizure	A seizure characterized by persistent change in mental status, behaviour or consciousness from baseline lasting more than 5 minutes, generally with epileptiform activity seen on EEG monitoring. These seizures occur in the absence of convulsive activity, with no motor abnormalities or feature subtle convulsive activity.
Non-epileptic attack	A disorder characterised by episodes of change in behaviour or

ISBN: 978-1-4731-4513-9

Term	Definition
disorder (NEAD)	movement, not caused by a primary change in electrical activity of the brain. Movements are varied, and the attacks can be difficult to differentiate from epileptic seizures. Refer to appendix A of the full guideline for the differentiation of epileptic attacks from NEAD and its subgroups. Also referred to as Psychogenic Non-Epileptic Seizures (PNES) and dissociative seizures: defined as the same condition
Provocation techniques during EEG	Methods used to provoke seizures, such as hyperventilation, photic stimulation, sleep deprivation, withdrawal of medication.
Refractory status Epilepticus	Continued status epilepticus despite treatment with two anticonvulsants in appropriate doses. This can occur in both convulsive and non-convulsive status epilepticus.
Renal stones	Kidney stones
Resective surgery	Most common type of epilepsy surgery; this involves removing a small portion of the brain. The surgeon removes brain tissue from the area of the brain where seizures occur. The aim is to remove the whole of the epileptogenic zone while minimising risk to the person with epilepsy.
Respiratory depression	Also known as hypoventilation, is a disorder in respiration typified by slow and ineffective breathing.
Rey Auditory Verbal Learning Test (RAVLT)	A neuropsychological assessment designed to evaluate verbal memory in patients, 16 years of age and older
Seizure semiology	Semiology refers to the characteristics of a seizure including symptoms and clinical signs. It can help in localising the epileptogenic zone.
Self-limited epilepsy with autonomic seizures	Formerly known as Panayiotopoulos syndrome this is an epilepsy syndrome presenting in early childhood (mean 4-7 years) with rare seizures that are prolonged. Characterised by autonomic features including vomiting pallor and sweating followed by tonic eye deviation, impairment of consciousness with possible evolution into a secondarily generalised seizure. Prognosis is excellent and treatment often unnecessary. EEG tends to show occipital spike discharges
Simple febrile seizure	A simple febrile convulsion is defined as brief (<15-minute) generalized seizure that occurs once during a 24-hour period in a febrile child who does not have an intracranial infection, metabolic disturbance, or history of afebrile seizures. More information is available on: https://www.nhs.uk/conditions/febrile-seizures/
Status epilepticus	A single seizure lasting more than five minutes or two or more seizures within a five-minute period without the person returning to normal between them.
Suboptimal (MRI)	An MRI scan would be deemed suboptimal if: • an inappropriate or inadequate set of sequences has been acquired. • image quality is poor, for example, due to patient movement.
Sudden unexpected (or unexplained) death in epilepsy (SUDEP)	Sudden, unexplained, witnessed or unwitnessed, non-traumatic and non-drowning death in individuals with epilepsy, with or without evidence for a seizure, and excluding documented status epilepticus, in which post-mortem examination does not reveal a toxicological or anatomic cause for death.
Syncope	A brief lapse in consciousness caused by transient reduction in blood flow to the brain. May be caused by many different factors, including emotional stress, vagal stimulation, vascular pooling in the legs, diaphoresis or sudden change in environmental temperature or body position.
Tachycardia	An elevated heart rate of over 100 beats per minute.
Temporal lobectomy	Removal of temporal lobe structures with the aim of reducing the number of seizures an individual has, make experienced seizures less

ISBN: 978-1-4731-4513-9

Term	Definition
	severe, or even stop them from happening altogether.
Temporoparietal occipital disconnection	A temporo-parietal-occipital disconnection surgery removes the front portion of the temporal lobe and disconnects the occipital and parietal lobes. The removed and disconnected lobes are all contained to one side of the brain, leaving the frontal lobe unaffected.
Tertiary epilepsy service	A service provided by epilepsy specialists who are adult or paediatric neurologists who undertake continuing professional development in the investigation, diagnosis and management of complex epilepsy. Offers access to additional specialist assessments including: • Neuropsychology • Neuropsychiatry • Specialised neuroimaging including 3T MRI • Specialised neurophysiology including video EEG telemetry Offers specialised assessment and management of particular patient groups: • Those with learning disability • Pregnancy and maternity care • Transition • Older people Offers access to: • specialised non-surgical treatments e.g.,, cannabidiol, ketogenic dietary therapy. • genetic diagnosis and counselling. • specialised assessment for surgery. • VNS implantation • participation in relevant clinical trials and research studies.
Tertiary epilepsy service	A service provided by epilepsy specialists who are adult or paediatric neurologists who undertake continuing professional development in the investigation, diagnosis and management of complex epilepsy. Offers access to additional specialist assessments including: • Neuropsychology • Neuropsychiatry • Specialised neuroimaging including 3T MRI • Specialised neurophysiology including video EEG telemetry Offers specialised assessment and management of particular patient groups: • Those with learning disability • Pregnancy and maternity care • Transition • Epilepsy in the elderly Offers access to: • specialised non-surgical treatments e.g., cannabidiol, ketogenic dietary therapy. • genetic diagnosis and counselling. • specialised assessment for surgery. • VNS insertion when appropriate. participation in relevant clinical trials and research studies.
Tonic seizure	An epileptic seizure characterised by abrupt generalised muscle stiffening possibly causing a fall. The seizure usually lasts less than a minute and recovery is rapid.
Tonic-clonic seizure	An epileptic seizure characterised by initial generalised muscle stiffening, followed by rhythmical jerking of the limbs, usually lasting a

ISBN: 978-1-4731-4513-9

Term	Definition
	few minutes. The individual may bite their tongue and may be incontinent. They may feel confused or sleepy afterwards, and take a while to recover fully.
Tuberous sclerosis complex	A rare genetic condition that causes mainly non-cancerous (benign) tumours to develop in different parts of the body. The tumours most often affect the brain, skin, kidneys, heart, eyes, and lungs.
Unsuccessful treatment	Treatment should be deemed unsuccessful if it has not managed to reduce or stop seizures, or if side effects are intolerable for the person with epilepsy.
Vagus nerve stimulation	The use of a device to stimulate the vagus nerve (the tenth cranial nerve or CN X, that interfaces with the parasympathetic control of the heart, lungs, and digestive tract) with electrical impulses.

5.2.2 General terms

Term	Definition
Abstract	Summary of a study, which may be published alone or as an introduction to a full scientific paper.
Algorithm (in guidelines)	A flow chart of the clinical decision pathway described in the guideline, where decision points are represented with boxes, linked with arrows.
Allocation concealment	The process used to prevent advance knowledge of group assignment in an RCT. The allocation process should be impervious to any influence by the individual making the allocation, by being administered by someone who is not responsible for recruiting participants.
Applicability	How well the results of a study or NICE evidence review can answer a clinical question or be applied to the population being considered.
Arm (of a clinical study)	Subsection of individuals within a study who receive one particular intervention, for example placebo arm.
Association	Statistical relationship between 2 or more events, characteristics or other variables. The relationship may or may not be causal.
Base case analysis	In an economic evaluation, this is the main analysis based on the most plausible estimate of each input. In contrast, see Sensitivity analysis.
Baseline	The initial set of measurements at the beginning of a study (after run-in period where applicable), with which subsequent results are compared.
Bayesian analysis	A method of statistics, where a statistic is estimated by combining established information or belief (the 'prior') with new evidence (the 'likelihood') to give a revised estimate (the 'posterior').
Before-and-after study	A study that investigates the effects of an intervention by measuring particular characteristics of a population both before and after taking the intervention, and assessing any change that occurs.
Bias	Influences on a study that can make the results look better or worse than they really are. (Bias can even make it look as if a treatment works when it does not.) Bias can occur by chance, deliberately or as a result of systematic errors in the design and execution of a study. It can also occur at different stages in the research process, for example, during the collection, analysis, interpretation, publication or review of research data. For examples see selection bias, performance bias, information bias, confounding factor, and publication bias.
Blinding	A way to prevent researchers, doctors and patients in a clinical trial from knowing which study group each patient is in so they cannot

ISBN: 978-1-4731-4513-9

Term	Definition
	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 nor 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.
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. Cohort studies can be retrospective or prospective. Retrospective cohort studies involve the use of hospital notes or databases that were often compiled before the research question was formulated, and that the researcher therefore looks at retrospectively. This is the origin of the term 'retrospective', and it is important to realise that the data within the database will have been originally collected in real-time (prospectively), with baseline data collected first, followed by outcome data after a follow up period. This is an important distinction with case-control studies, where the data are more truly 'retrospective', as they are collected largely from recall. The disadvantage of retrospective cohort studies is that often data on important confounding variables relevant to the research question

ISBN: 978-1-4731-4513-9

Term	Definition
	may not be present in the clinical database. Prospective cohort studies are usually the creation of the researcher and so the data are collected by the researcher prospectively. Prospective studies have the advantage that they can collect any data that is deemed appropriate, such as important confounding variable data. See also observational study.
Comorbidity	A disease or condition that someone has in addition to the health problem being studied or treated.
Comparability	Similarity of the groups in characteristics likely to affect the study results (such as health status or age).
Concordance	This is a recent term whose meaning has changed. It was initially applied to the consultation process in which doctor and patient agree therapeutic decisions that incorporate their respective views, but now includes patient support in medicine taking as well as prescribing communication. Concordance reflects social values but does not address medicine-taking and may not lead to improved adherence.
Confidence interval (CI)	There is always some uncertainty in research. This is because a small group of patients is studied to predict the effects of a treatment on the wider population. The confidence interval is a way of expressing how certain we are about the findings from a study, using statistics. It gives a range of results that is likely to include the 'true' value for the population. The CI is usually stated as '95% CI', which means that the range of values has a 95 in a 100 chance of including the 'true' value. For example, a study may state that "based on our sample findings, we are 95% certain that the 'true' population blood pressure is not higher than 150 and not lower than 110". In such a case the 95% CI would be 110 to 150. A wide confidence interval indicates a lack of certainty about the true 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	If we evaluated the effects of a treatment on a single group, it would not be possible to know if any change in outcome was due to the treatment or some intervening variable, such as natural recovery or the placebo effect. By having a control group, who do not have the study treatment, but who are otherwise prone to the same intervening effects, it is possible to discern the true treatment effects. For example, if the treatment group improve quality of life by 7 points and the control group improve quality of life by 5 points, we can assume, all things being equal, that the difference in quality of life (mean difference=2 points) must equal the true treatment effect. This is because the intervening effects will cancel each other out, leaving just the treatment difference. For this to work, the control group has

ISBN: 978-1-4731-4513-9

Term	Definition
	to be as similar as possible to the treatment group so that very similar intervening effects are experienced to the same degree. This also prevents secondary confounding resulting from the groups having different prognostic characteristics. Similarity across groups can be achieved through randomisation. In cases where randomisation is not possible statistical adjustment or propensity matching are good strategies to minimise bias. Instead of the study treatment, the control group may receive the standard treatment (sometimes called 'usual care') or a dummy treatment (placebo).
Cost–benefit analysis (CBA)	Cost–benefit analysis is one of the tools used to carry out an economic evaluation. The costs and benefits are measured using the same monetary units (for example, pounds sterling) to see whether the benefits exceed the costs.
Cost–consequences analysis (CCA)	Cost–consequences analysis is one of the tools used to carry out an economic evaluation. This compares the costs (such as treatment and hospital care) and the consequences (such as health outcomes) of a test or treatment with a suitable alternative. Unlike cost–benefit analysis or cost-effectiveness analysis, it does not attempt to summarise outcomes in a single measure (like the quality-adjusted life year) or in financial terms. Instead, outcomes are shown in their natural units (some of which may be monetary) and it is left to decision-makers to determine whether, overall, the treatment is worth carrying out.
Cost-effectiveness analysis (CEA)	Cost-effectiveness analysis is one of the tools used to carry out an economic evaluation. The benefits are expressed in non-monetary terms related to health, such as symptom-free days, heart attacks avoided, deaths avoided or life years gained (that is, the number of years by which life is extended as a result of the intervention).
Cost-effectiveness model	An explicit mathematical framework, which is used to represent clinical decision problems and incorporate evidence from a variety of sources in order to estimate the costs and health outcomes.
Cost–utility analysis (CUA)	Cost–utility analysis is one of the tools used to carry out an economic evaluation. The benefits are assessed in terms of both quality and duration of life, and expressed as quality-adjusted life years (QALYs). See also utility.
Credible interval (CrI)	The Bayesian equivalent of a confidence interval.
Decision analysis	An explicit quantitative approach to decision-making under uncertainty, based on evidence from research. This evidence is translated into probabilities, and then into diagrams or decision trees which direct the clinician through a succession of possible scenarios, actions and outcomes.
Deterministic analysis	In economic evaluation, this is an analysis that uses a point estimate for each input. In contrast, see Probabilistic analysis
Diagnostic odds ratio	The diagnostic odds ratio is a measure of the effectiveness of a diagnostic test. It is defined as the ratio of the odds of the test being positive if the subject has a disease relative to the odds of the test being positive if the subject does not have the disease.
Discounting	Costs and perhaps benefits incurred today have a higher value than costs and benefits occurring in the future. Discounting health benefits reflects individual preference for benefits to be experienced in the present rather than the future. Discounting costs reflects individual preference for costs to be experienced in the future rather than the present.
Disutility	The loss of quality of life associated with having a disease or condition. See Utility
Dominance	A health economics term. When comparing tests or treatments, an

ISBN: 978-1-4731-4513-9

Term	Definition
	option that is both less effective and costs more is said to be 'dominated' by the alternative.
Drop-out	A participant who withdraws from a trial before the end.
Economic evaluation	An economic evaluation is used to assess the cost effectiveness of healthcare interventions (that is, to compare the costs and benefits of a healthcare intervention to assess whether it is worth doing). The aim of an economic evaluation is to maximise the level of benefits – health effects – relative to the resources available. It should be used to inform and support the decision-making process; it is not supposed to replace the judgement of healthcare professionals. There are several types of economic evaluation: cost–benefit analysis, cost–consequences analysis, cost-effectiveness analysis, cost-minimisation analysis and cost–utility analysis. They use similar methods to define and evaluate costs, but differ in the way they estimate the benefits of a particular drug, programme or intervention.
Effect (as in effect measure, treatment effect, estimate of effect, effect size)	A measure that shows the magnitude of the outcome in one group compared with that in a control group. For example, if the absolute risk reduction is shown to be 5% and it is the outcome of interest, the effect size is 5%. The effect size is usually tested, using statistics, to find out how likely it is that the effect is a result of the treatment and has not just happened by chance (that is, to see if it is statistically significant).
Effectiveness	How beneficial a test or treatment is under usual or everyday conditions, compared with doing nothing or opting for another type of care.
Efficacy	How beneficial a test, treatment or public health intervention is under ideal conditions (for example, in a laboratory), compared with doing nothing or opting for another type of care.
Epidemiological study	The study of a disease within a population, defining its incidence and prevalence and examining the roles of external influences (for example, infection, diet) and interventions.
EQ-5D (EuroQol 5 dimensions)	A standardised instrument used to measure health-related quality of life. It provides a single index value for health status.
Evidence	Information on which a decision or guidance is based. Evidence is obtained from a range of sources including randomised controlled trials, observational studies, expert opinion (of clinical professionals or patients).
Exclusion criteria (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

ISBN: 978-1-4731-4513-9

Term	Definition
	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 such wide confidence intervals around the estimate of effect that their interpretation can change. For example, the confidence intervals may extend from effects denoting a clinical benefit to those denoting a clinical harm.
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. Furthermore, people who drop out of a treatment are often the worst responders, and so failure to include such people in the analysis can seriously skew results.
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 healthier diet.

ISBN: 978-1-4731-4513-9

Term	Definition
Intraoperative	The period of time during a surgical procedure.
Kappa statistic	A statistical measure of inter-rater agreement that takes into account the agreement occurring by chance.
Length of stay	The total number of days a participant stays in hospital.
Licence	See 'Product licence'.
Life years gained	Mean average years of life gained per person as a result of the intervention compared with an alternative intervention.
Likelihood ratio	The likelihood ratio combines information about the sensitivity and specificity. It tells you how much a positive or negative result changes the likelihood that a patient would have the disease. The likelihood ratio of a positive test result (LR+) is sensitivity divided by (1 minus specificity).
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).
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	The average number of patients who need to be treated to get a

ISBN: 978-1-4731-4513-9

Term	Definition
(NNT)	positive outcome. For example, if the NNT is 4, then 4 patients would have to be treated to ensure 1 of them gets better. The closer the NNT is to 1, the better the treatment. For example, if you give a stroke prevention drug to 20 people before 1 stroke is prevented, the number needed to treat is 20. See also number needed to harm, absolute risk reduction.
Observational study	Individuals or groups are observed, or certain factors are measured. No attempt is made to affect the outcome. For example, an observational study of a disease or treatment would allow 'nature' or usual medical care to take its course. Changes or differences in one characteristic (for example, whether or not people received a specific treatment or intervention) are studied without intervening. There is a greater risk of selection bias than in experimental studies.
Odds ratio	Odds are a way to represent how likely it is that something will happen (the probability). An odds ratio compares the probability of something in one group with the probability of the same thing in another. An odds ratio of 1 between 2 groups would show that the probability of the event (for example a person developing a disease, or a treatment working) is the same for both. An odds ratio greater than 1 means the event is more likely in the first group. An odds ratio less than 1 means that the event is less likely in the first group. Sometimes probability can be compared across more than 2 groups – in this case, one of the groups is chosen as the 'reference category', and the odds ratio is calculated for each group compared with the reference category. For example, to compare the risk of dying from lung cancer for non-smokers, occasional smokers and regular smokers, non-smokers could be used as the reference category. Odds ratios would be worked out for occasional smokers compared with non-smokers and for regular smokers compared with non-smokers. See also confidence interval, risk ratio.
Opportunity cost	The loss of other healthcare programmes displaced by investment in or introduction of another intervention. This may be best measured by the health benefits that could have been achieved had the money been spent on the next best alternative healthcare intervention.
Outcome	The impact that a test, treatment, policy, programme or other intervention has on a person, group or population. Outcomes from interventions to improve the public's health could include changes in knowledge and behaviour related to health, societal changes (for example, a reduction in crime rates) and a change in people's health and wellbeing or health status. In clinical terms, outcomes could include the number of patients who fully recover from an illness or the number of hospital admissions, and an improvement or deterioration in someone's health, functional ability, symptoms or situation. Researchers should decide what outcomes to measure before a study begins.
P value	The p value is a statistical measure that indicates whether or not an effect is statistically significant. For example, if a study comparing 2 treatments found that one seems more effective than the other, the p value is the probability of obtaining these results by chance. By convention, if the p value is below 0.05 (that is, there is less than a 5% probability that the results occurred by chance) it is considered that there probably is a real difference between treatments. If the p value is 0.001 or less (less than a 1% probability that the results occurred by chance), the result is seen as highly significant. If the p value shows that there is likely to be a difference between treatments, the confidence interval describes how big the difference in effect might be.
Perioperative	The period from admission through surgery until discharge,

ISBN: 978-1-4731-4513-9

Term	Definition
	encompassing the preoperative and postoperative periods.
Placebo	A fake (or dummy) treatment given to participants in the control group of a clinical trial. It is indistinguishable from the actual treatment (which is given to participants in the experimental group). The aim is to determine what effect the experimental treatment has had – over and above any placebo effect caused because someone has received (or thinks they have received) care or attention.
Polypharmacy	The use or prescription of multiple medications.
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.
Post-test probability	In diagnostic tests: The proportion of patients with that particular test result who have the target disorder (post-test odds/[1 plus post-test odds]).
Power (statistical)	The ability to demonstrate an association when one exists. Power is related to sample size; the larger the sample size, the greater the power and the lower the risk that a possible association could be missed.
Preoperative	The period before surgery commences.
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.
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.
Prior distribution	In Bayesian statistics this is the probability distribution for a statistic based on previous evidence or belief.
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.

ISBN: 978-1-4731-4513-9

Term	Definition
Quality of life	See ‘Health-related quality of life’.
Quality-adjusted life year (QALY)	A measure of the state of health of a person or group in which the benefits, in terms of length of life, are adjusted to reflect the quality of life. One QALY is equal to 1 year of life in perfect health. QALYS are calculated by estimating the years of life remaining for a patient following a particular treatment or intervention and weighting each year with a quality-of-life score (on a scale of 0 to 1). It is often measured in terms of the person’s ability to perform the activities of daily life, freedom from pain and mental disturbance.
Randomisation	Assigning participants in a research study to different groups without taking any similarities or differences between them into account. For example, it could involve using a random numbers table or a computer-generated random sequence. It means that each individual (or each group in the case of cluster randomisation) has the same chance of receiving each intervention.
Randomised controlled trial (RCT)	A study in which a number of similar people are randomly assigned to 2 (or more) groups to test a specific drug or treatment. One group (the experimental group) receives the treatment being tested, the other (the comparison or control group) receives an alternative treatment, a dummy treatment (placebo) or no treatment at all. The groups are followed up to see how effective the experimental treatment was. Outcomes are measured at specific times and any difference in response between the groups is assessed statistically. This method is also used to reduce bias.
RCT	See ‘Randomised controlled trial’.
Receiver operated characteristic (ROC) curve	A graphical method of assessing the accuracy of a diagnostic test. Sensitivity is plotted against 1 minus specificity. A perfect test will have a positive, vertical linear slope starting at the origin. A good test will be somewhere close to this ideal.
Reference standard	The test that is considered to be the best available method to establish the presence or absence of the outcome – this may not be the one that is routinely used in practice.
Reporting bias	See ‘Publication bias’.
Resource implication	The likely impact in terms of finance, workforce or other NHS resources.
Retrospective cohort study	See cohort study and case control study.
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.

ISBN: 978-1-4731-4513-9

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

ISBN: 978-1-4731-4513-9

Term	Definition
	• organisations representing healthcare professionals.
State transition model	See Markov model
Systematic review	A review in which evidence from scientific studies has been identified, appraised and synthesised in a methodical way according to predetermined criteria. It may include a meta-analysis.
Time horizon	The time span over which costs and health outcomes are considered in a decision analysis or economic evaluation.
Transition probability	In a state transition model (Markov model), this is the probability of moving from one health state to another over a specific period of time.
Treatment allocation	Assigning a participant to a particular arm of a trial.
Univariate	Analysis which separately explores each variable in a data set.
Utility	In health economics, a 'utility' is the measure of the preference or value that an individual or society places upon a particular health state. It is generally a number between 0 (representing death) and 1 (perfect health). The most widely used measure of benefit in cost–utility analysis is the quality-adjusted life year, but other measures include disability-adjusted life years (DALYs) and healthy year equivalents (HYEs).

ISBN: 978-1-4731-4513-9

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. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Explanation and Exploration (PRESS E&E). Ottawa. CADTH, 2016. Available from: https://www.cadth.ca/sites/default/files/pdf/CP0015_PRESS_Update_Report_2016.pdf

3. 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

4. National Institute for Health and Care Excellence. The NICE Charter. 2020. Available from: https://www.nice.org.uk/about/who-we-are/our-charter Last accessed: 10/03/2020.

5. National Institute for Health and Care Excellence. The principles that guide the development of NICE guidance and standards. 2020. Available from: https://www.nice.org.uk/about/who-we-are/our-principles Last accessed: 10/03/2020.

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: 13/05/2021.

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/

ISBN: 978-1-4731-4513-9