National Institute for Health and Care Excellence

Draft

Chronic pain: assessment and management

NICE guideline: methods

NICE guideline

Methods

August 2020

Draft for Consultation

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

Disclaimer

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

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

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 All NICE guidance is subject to regular review and may be updated or withdrawn.

ISBN
[to be added on publication]

Contents

1

1 Development of the guideline

1.1 What is a NICE guideline?

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

NICE guidelines can:

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

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

We produce our guidelines using the following steps:

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

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

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

NICE Pathways brings together all connected NICE guidance.

1.2 Remit

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

The remit for this guideline is:

Chronic pain: assessment and management.

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 Nick Kosky in accordance with guidance from NICE.

The group met approximately every 6-8 weeks during the development of the guideline. At the start of the guideline development process all committee members declared interests including consultancies, fee-paid work, shareholdings, fellowships and support from the healthcare industry. At all subsequent committee meetings, members declared arising conflicts of interest.

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

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

1.3.1 What this guideline covers

This guideline will cover adults and young people (16 years and over) with chronic pain. It should be used alongside NICE guidance for specific conditions that cause pain, including headaches, low back pain and sciatica, rheumatoid arthritis, osteoarthritis, spondyloarthritis, endometriosis and irritable bowel syndrome It also includes recommendations on managing chronic primary pain (as defined in ICD-11) for which there is no other NICE guidance.

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

1.3.2 What this guideline does not cover

This guideline will not cover children and young people (under 16 years) with chronic pain.

Areas that will not be covered:

1 Specific management of chronic pain when this is covered by existing NICE guidance, for example, managing chronic pain in headaches, low back pain and sciatica, neuropathic pain, rheumatoid arthritis, osteoarthritis, spondyloarthritis, endometriosis and irritable bowel syndrome.

2 Pain management as part of palliative care.

1.3.3 Relationships between the guideline and other NICE guidance

Related NICE guidelines:

- Cannabis-based medicinal products. NICE guideline NG144 (2019).
- Post-traumatic stress disorder. NICE guideline NG116 (2018).

- Rheumatoid arthritis in adults: management. NICE guideline NG100 (2018).
- Endometriosis: diagnosis and management. NICE guideline NG73 (2017).
- Spondyloarthritis in over 16s: diagnosis and management. NICE guideline NG65 (2017).
- Neuropathic pain in adults: pharmacological management in non-specialist settings. NICE guideline CG173 (2017).
- Low back pain and sciatica in over 16s: assessment and management. NICE guideline NG59 (2016).
- Multimorbidity: clinical assessment and management. NICE guideline NG56 (2016).
- Palliative care for adults: strong opioids for pain relief. NICE guideline CG140 (2016).
- Controlled drugs: safe use and management. NICE guideline NG46 (2016).
- Transition from children's to adults' services for young people using health or social care services. NICE guideline NG43 (2016).
- Headaches in over 12s: diagnosis and management. NICE guideline CG150 (2015).
- Workplace health: management practices. NICE guideline NG13 (2015).
- Medicines optimisation. NICE guideline NG5 (2015).
- Osteoarthritis: care and management. NICE guideline CG177 (2014).
- Behaviour change: individual approaches. NICE public health guideline PH49 (2014).
- Physical activity: brief advice for adults in primary care. NICE public health guideline PH44 (2013).
- Patient experience in adult NHS services. NICE guideline CG138 (2012).
- Service user experience in adult mental health. NICE guideline CG136 (2011).
- Common mental health problems: identification and pathways to care. NICE guideline CG123 (2011).
- Depression in adults with a chronic physical health problem: recognition and management. NICE guideline CG91 (2009).
- Depression in adults: recognition and management. NICE guideline CG90 (2009).
- Medicines adherence. NICE guideline CG76 (2009).

Related NICE guidance currently in development:

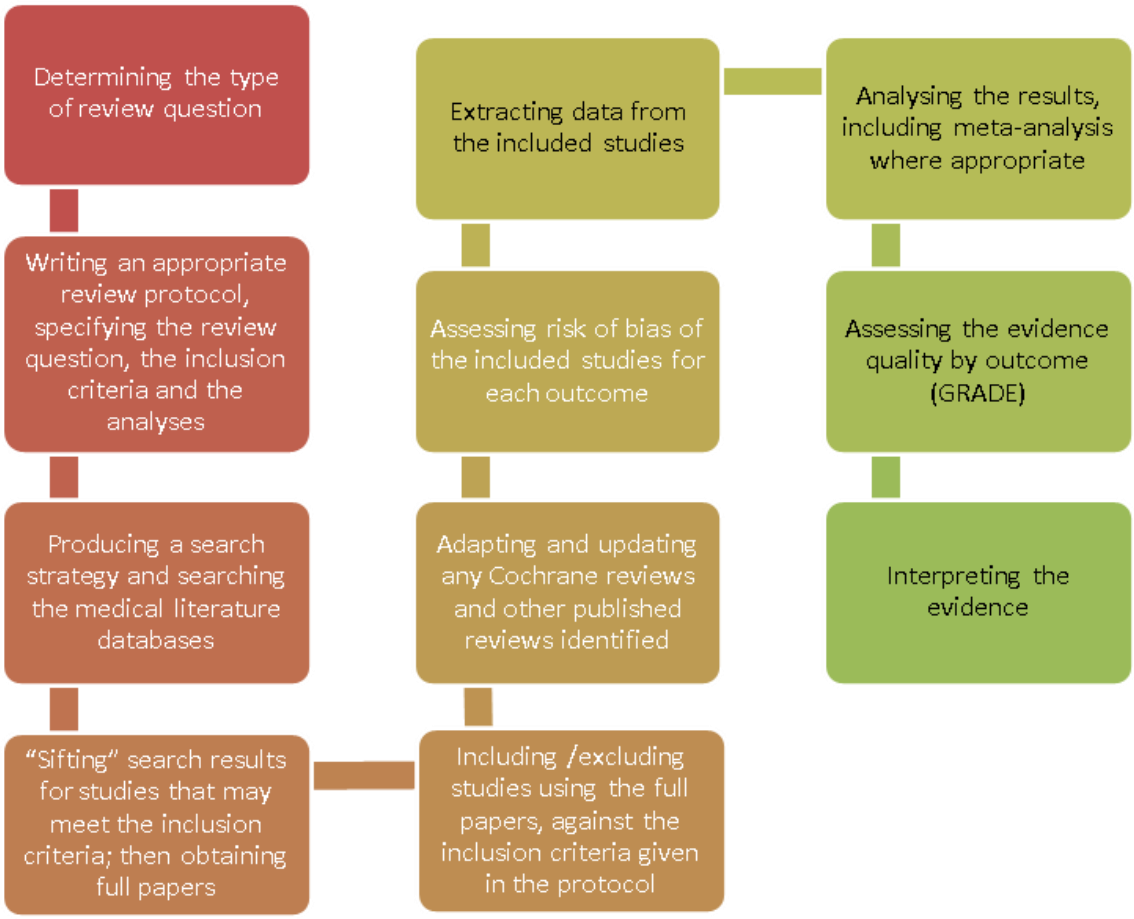
- Safe prescribing and withdrawal management of prescribed drugs associated with dependence and withdrawal. NICE guideline. Publication expected November 2021.

2 Methods

This report sets out in detail the methods used to review the evidence and to develop the recommendations that are presented in each of the evidence reviews for this guideline. This guidance was developed in accordance with the methods outlined in the NICE guidelines manual, 2014 version.[4]

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

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



2.1 Developing the review questions and outcomes

The questions were based on the key clinical areas and draft review questions identified in the scope. The review protocols were drafted by the NGC technical team and refined and validated by the committee and signed off by NICE. A total of 14 review questions were developed in this guideline and outlined in table 1.

Review questions were developed using:

- a PICO framework (population, intervention, comparison and outcome) for intervention reviews
- population, exposure and outcomes for prognostic reviews
- a framework of population, setting and context for qualitative reviews.

This use of a framework informed a more detail protocol that guided the literature searching process, critical appraisal and synthesis of evidence, and facilitated the development of recommendations by the guideline committee. 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
A	Prognostic	What psychological factors may be barriers to successfully managing chronic pain?	Critical outcomes: • Health related quality of life (including meaningful activity), measured using a validated scale e.g. EQ-5D, SF36, SF12 • Pain reduction, as reported by the studies Studies must report at least one of these outcomes in order to be included in the review.
A	Prognostic	What social factors may be barriers to successfully managing chronic pain?	Critical outcomes: • Health related quality of life (including meaningful activity), measured using a validated scale e.g. EQ-5D, SF36, SF12 • Pain reduction, as reported by the studies Studies must report at least one of these outcomes in order to be included in the review.
A	Prognostic	What biological factors may be barriers to successfully managing chronic pain?	Critical outcomes: • Health related quality of life (including meaningful activity), measured using a validated scale e.g. EQ-5D, SF36, SF12 • Pain reduction, as reported by the studies Studies must report at least one of these outcomes in order to be included in the review.
B	Qualitative	What are the best methods of communication between healthcare professionals and people with chronic pain?	Themes derived from the evidence identified for this review, and not pre-specified by the guideline committee in advance.
C	Intervention	What is the clinical and cost effectiveness of pain management programmes for the management of chronic pain?	Critical outcomes: • Health related quality of life (including meaningful activity) • Physical function (5 minute walk, sit to stand, Roland

Evidence report	Type of review	Review questions	Outcomes
			Morris Disability Questionnaire, Oswestry Disability Index, Canadian Occupational Performance Measure) • Psychological distress (depression/ anxiety) (preferably Hospital Anxiety and Depression Scale) • Pain interference (brief pain inventory interference subscale) • Pain self-efficacy (pain self-efficacy questionnaire). Important outcomes: • Use of healthcare services • Sleep • Discontinuation • Pain reduction (any validated scale) All outcomes extracted at the longest time point up to 3 months and at the longest time point after 3 months.
D	Intervention	What is the clinical and cost effectiveness of social interventions aimed at improving the quality of life of people with chronic pain?	Critical outcomes: • Health related quality of life (including meaningful activity). Important outcomes: • Physical function (5 minute walk, sit to stand, Roland Morris Disability Questionnaire, Oswestry Disability Index, Canadian Occupational Performance Measure) • Pain self-efficacy (pain self-efficacy questionnaire) • Use of healthcare services • Sleep • Discontinuation • Pain reduction (any validated scale) • Psychological distress (depression/ anxiety) (preferably Hospital Anxiety and Depression Scale) • Pain interference (brief pain inventory interference subscale).

Evidence report	Type of review	Review questions	Outcomes
			All outcomes extracted at the longest time point up to 3 months and at the longest time point after 3 months.
E	Intervention	What is the clinical and cost effectiveness of exercise interventions for the management of chronic primary pain?	Critical outcomes: • Pain reduction (any validated scale) • Health related quality of life (including meaningful activity) • Physical function (e.g. 5 minute walk, sit to stand, Roland Morris Disability Questionnaire, Oswestry Disability Index, Canadian Occupational Performance Measure) • Psychological distress (depression/anxiety) (preferably Hospital Anxiety and Depression Scale). Important outcomes: • Use of healthcare services • Sleep • Discontinuation. All outcomes extracted at the longest time point up to 3 months and at the longest time point after 3 months.
F	Intervention	What is the clinical and cost effectiveness of psychological therapy for the management of chronic primary pain?	Critical outcomes: • Health related quality of life (including meaningful activity) • Physical function (5 minute walk, sit to stand, Roland Morris Disability Questionnaire, Oswestry Disability Index, Canadian Occupational Performance Measure) • Psychological distress (depression/anxiety) (preferably Hospital Anxiety and Depression Scale) • Pain interference (brief pain inventory interference subscale) and pain self-efficacy (pain self-efficacy questionnaire). Important outcomes: • Use of healthcare services

Evidence report	Type of review	Review questions	Outcomes
			• Sleep • Discontinuation • Pain reduction (any validated scale). All outcomes extracted at the longest time point up to 3 months and at the longest time point after 3 months.
G	Intervention	What is the clinical and cost effectiveness of acupuncture or dry needling for the management of chronic primary pain?	Critical outcomes: • Pain reduction (any validated scale) • Health related quality of life (including meaningful activity) • Physical function (5 minute walk, sit to stand, Roland Morris Disability Questionnaire, Oswestry Disability Index, Canadian Occupational Performance Measure) • Psychological distress (depression/anxiety) (preferably Hospital Anxiety and Depression Scale) • Pain self-efficacy • Pain interference. Important outcomes: • Use of healthcare services • Sleep • Discontinuation. All outcomes extracted at the longest time point up to 3 months and at the longest time point after 3 months.
H	Intervention	What is the clinical and cost effectiveness of electrical physical modalities for the management of chronic primary pain?	Critical outcomes: • Pain reduction (any validated scale) • Health related quality of life (including meaningful activity) • Physical function (5 minute walk, sit to stand, Roland Morris Disability Questionnaire, Oswestry Disability Index, Canadian Occupational Performance Measure) • Psychological distress (depression/anxiety)

Evidence report	Type of review	Review questions	Outcomes
			(preferably Hospital Anxiety and Depression Scale) • Pain interference (brief pain inventory interference subscale) • Pain self-efficacy (pain self-efficacy questionnaire). Important outcomes: • Use of healthcare services • Sleep • Discontinuation. All outcomes extracted at the longest time point up to 3 months and at the longest time point after 3 months.
I	Intervention	What is the clinical and cost effectiveness of manual therapy for the management of chronic primary pain?	Critical outcomes: • Pain reduction (any validated scale) • Health related quality of life (including meaningful activity) • Physical function (5 minute walk, sit to stand, Roland Morris Disability Questionnaire, Oswestry Disability Index, Canadian Occupational Performance Measure) • Psychological distress (depression/anxiety) (preferably Hospital Anxiety and Depression Scale) • Pain interference (brief pain inventory interference subscale) • Pain self-efficacy (pain self-efficacy questionnaire). Important outcomes: • Use of healthcare services • Sleep • Discontinuation. All outcomes extracted at the longest time point up to 3 months and at the longest time point after 3 months.
J	Intervention	What is the clinical and cost effectiveness of pharmacological interventions for chronic primary pain?	Critical outcomes: • Pain reduction (any validated scale)

Evidence report	Type of review	Review questions	Outcomes
			• Health related quality of life (including meaningful activity) • Physical function (5 minute walk, sit to stand, Roland Morris Disability Questionnaire, Oswestry Disability Index, Canadian Occupational Performance Measure) • Psychological distress (depression/ anxiety) (preferably Hospital Anxiety and Depression Scale) • Discontinuation due to adverse events. Important outcomes: • Use of healthcare services • Sleep. All outcomes extracted at the longest time point up to 3 months and at the longest time point after 3 months.
J	Intervention	What is the long-term safety of opioids for the management of chronic pain?	Critical outcomes: Serious adverse events: • Cognitive impairment • Fractures and falls • Sexual dysfunction/endocrine impairment • Immune dysfunction • Sleep apnoea • Cardiovascular events • All-cause mortality • Self-harm/suicide • Dependence • Depressive symptoms/mood disturbances. Outcomes extracted at the longest time point up to 6 months, at the longest time point up to 1 year and at the longest time point after 1 year.
J	Intervention	What is the long-term safety of gabapentinoids for the management of chronic pain?	Critical outcomes: Serious adverse events: • Cognitive impairment • Gait disturbance/ataxia • Loss of balance • All-cause mortality

Evidence report	Type of review	Review questions	Outcomes
			• Dependence • Weight gain • Rash • Peripheral oedema • Tremor • Somnolence. Outcomes extracted at the longest time point up to 6 months, at the longest time point up to 1 year and at the longest time point after 1 year.

2.2 Searching for evidence

2.2.1 Clinical and health economics literature searches

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

Systematic literature searches were undertaken to identify all published clinical and health economic evidence relevant to the review questions. Searches were undertaken according to the parameters stipulated within the NICE guidelines manual.[4] Databases were searched using relevant medical subject headings, free-text terms and study-type filters where appropriate. Where possible, searches were restricted to papers published in English. Studies published in languages other than English were not reviewed. All searches were updated on 20 May 2020. Papers published or added to databases after this date were not considered. If new evidence, falling outside of the timeframe for the guideline searches, is identified, for example in consultation comments received from stakeholders, the impact on the guideline will be considered, and any further action agreed between NGC and NICE staff with a quality assurance role.

Prior to running, search strategies were quality assured using different approaches, checking for key paper retrieval and search strategies were peer reviewed by a second information specialist using a QA process based on the Peer Review of Electronic Search Strategies PRESS checklist.[3] 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 in the following databases

- The Cochrane Library (Wiley)
- Medline (Ovid)

Searching for unpublished literature was not undertaken.

2.3 Identifying and analysing evidence

The evidence for each review question was reviewed using the following process:

- Potentially relevant studies were identified from the search results by reviewing titles and abstracts. The full papers were then obtained.

- Full papers were evaluated against the pre-specified inclusion and exclusion criteria set out in the protocol to identify studies that addressed the review question. The review protocols are included in an appendix to each of the evidence reports.
- Relevant studies were critically appraised using the preferred study design checklist as specified in the NICE guidelines manual.[4] The checklist used is included in the individual review protocols in each of the evidence reports.
- Key information was extracted about interventional study methods and results into 'EviBase', NGC's purpose-built software. Summary evidence tables are produced from data entered into EviBase, including critical appraisal ratings. Key information about non-interventional study methods and results were manually extracted into standard Word evidence tables (evidence tables are included in an appendix to each of the evidence reports).
- Summaries of the evidence were generated by outcome. Where possible, outcome data were combined, analysed and reported according to study design:
  - Randomised data were meta-analysed where appropriate and reported in GRADE profile tables.
  - Non-randomised comparative data were meta-analysed where appropriate and reported in GRADE profile tables.
  - Data from non-comparative cohort studies were presented narratively in summary tables, with separate tables for study limitations assessments.
  - Prognostic data were meta-analysed where appropriate and reported in adapted GRADE profile tables.
  - Qualitative data were synthesised across studies using thematic analysis and presented as summary statements in GRADE CERQual tables.
- 10% of the abstracts were reviewed by two reviewers, with any disagreements resolved by discussion or, if necessary, a third independent reviewer.
- All of the evidence reviews were quality assured by a senior systematic reviewer. This included checking:
  - papers were included or excluded appropriately
  - a sample of the data extractions
  - a sample of the risk of bias assessments
  - correct methods were used to synthesise data.

  Discrepancies were identified and resolved through discussion (with a third reviewer where necessary).

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 an appendix to each of the evidence reports. The committee was consulted about any uncertainty regarding inclusion or exclusion.

Conference abstracts were not automatically excluded from any review. If the abstracts were included the authors were contacted for further information. Literature reviews, posters, letters, editorials, comment articles, unpublished studies and studies not in published in English language were excluded.

For the pharmacological intervention review, enriched enrolment trials (including a placebo run in phase) where participants were initiated on the drug or placebo prior to randomisation (and sometimes included/excluded based on response) were excluded. Evidence from trials employing this methodology was considered to be of lower quality due to the increased risk of participant blinding/performance bias and the limited applicability to the wider review

population. Therefore the committee decided to limit the study design to those that weren't enrichment trials.

For prognostic reviews, only studies that included interventions reviewed in this guideline were included. For example, studies assessing prognostic factors for successful pain management after surgery were excluded.

2.3.16 Saturation of qualitative studies

Data extraction in qualitative reviews is a thorough process and may require more time compared to intervention reviews. A common approach applied in systematic reviews of qualitative data is to stop extracting data once saturation has been reached. In an exploratory review, where themes are not predefined in the protocol, thematic or data extraction may be applied. For the purposes of this review, extraction of information from relevant studies was stopped when data saturation was reached, i.e. no new information was emerging for a specific theme. This includes; studies that don't report any new themes additional to those already identified in the review as well as not contributing additional information to the existing themes, as well as studies which report a new theme but data from other themes in the study doesn't contribute to the existing review themes. In the latter scenario only the new theme data is extracted. These studies are not specifically 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 data saturation having been reached, but that fit the inclusion criteria of the protocol, were listed in the table for studies 'identified but not extracted due to saturation' in an appendix to the qualitative evidence review.

2.3.17 Type of studies

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

For intervention reviews, randomised controlled trials (RCTs) were included where identified as because they are considered the most robust type of study design that can produce an unbiased estimate of the intervention effects. Non-randomised intervention studies were considered appropriate for inclusion in reviews of safety if there was insufficient randomised evidence for the committee to make a decision. Refer to the review protocols in each evidence report for full details on the study design of studies that were appropriate for each review question.

For prognostic review questions, prospective and retrospective cohort studies were included. Case–control studies were not included unless no cohort studies were identified.

Systematic reviews and meta-analyses conducted to the same methodological standards as the NICE reviews were included within the evidence reviews in preference to primary studies, where they were available and applicable to the review questions and updated or added to where appropriate to the guideline review question. Individual patient data (IPD) meta-analyses were preferentially included if meeting the protocol and methodological criteria.

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

2.3.12 Methods of combining clinical studies

2.3.2.21 Data synthesis for intervention reviews

Where possible, meta-analyses were conducted using Cochrane Review Manager (RevMan5)[8] software to combine the data given in all studies for each of the outcomes of interest for the review question.

2.3.2.1.61 Analysis of different types of data

Dichotomous outcomes

Fixed-effects (Mantel–Haenszel) techniques were used to calculate risk ratios (relative risk, RR) for the binary outcomes. 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 there were zero events in both arms, risk differences rather than risk ratios were calculated.

Continuous outcomes

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

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 and the committee have agreed that it is clinically meaningful and appropriate to do this); each different measure in each study was 'normalised' to the standard deviation value pooled between the intervention and comparator groups in that same study.

The means and standard deviations of continuous outcomes are required for meta-analysis. However, in cases where standard deviations were not reported, the standard error was calculated if the p values or 95% confidence intervals (95% CI) were reported, and meta-analysis was undertaken with the mean and standard error using the generic inverse variance method in Cochrane Review Manager (RevMan5[8] software.

2.3.2.2.92 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.[8] 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.

2.3.2.3.53 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 out for:

- type of chronic primary pain (chronic widespread pain, complex regional pain syndrome, chronic visceral pain, chronic orofacial pain, chronic primary musculoskeletal pain)
- cognitive impairment (vs. no cognitive impairment)

- learning difficulties (vs. no learning difficulties)
- first language not English (vs. first language English)
- sensory impairment (vs. no sensory impairment)
- homeless (vs. not homeless)
- age (16-25 years, >25 years)

These subgrouping strategies were applied in order of priority. Once a subgrouping strategy was found to explain heterogeneity from all derived subgroups, further subgrouping strategies were not used. If the subgroup analysis resolved heterogeneity within all of the derived subgroups, then each of the derived subgroups were adopted as separate outcomes. For example, instead of the single outcome of '*quality of life'*, this was separated into 2 outcomes '*quality of life in people with chronic widespread pain'* and '*quality of life in people with other types of pain'*. Assessments of potential differences in effect between subgroups were based on the chi-squared tests for heterogeneity statistics between subgroups. Any subgroup differences were interpreted with caution as separating the groups breaks the study randomisation and as such is subject to uncontrolled confounding.

If 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.

2.3.2.4 Complex analysis

Network meta-analysis was considered for the comparison of pharmacological treatments, but was not pursued because of insufficient data available for the relevant outcomes and lack of evidence of effectiveness for the majority of interventions.

2.3.3 Data synthesis for prognostic factor reviews

Adjusted odds ratios, risk ratios, hazard ratios, or beta coefficients with their 95% CIs, for the effect of the pre-specified prognostic factors were extracted from the studies. Studies were only included if the confounders pre-specified by the committee were either matched at baseline or were adjusted for in multivariate analysis. Prospective and retrospective cohort studies reporting multivariable analyses that adjusted for key confounders identified by the committee at the protocol stage for that outcome were the preferred study design.

Data were not combined in meta-analyses for prognostic studies, unless there was homogeneity in confounding factors adjusted for and measures of prognostic factors and outcomes.

2.3.4 Data synthesis for qualitative study reviews

The main findings for each included paper were identified and thematic analysis methods were used to synthesise this information into broad overarching themes which were summarised into the main review findings. The evidence was presented in the form of a narrative summary detailing the evidence from the relevant papers and how this informed the overall review finding plus a statement on the level of confidence for that review finding and an explanation of the quality assessment. 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.

2.3.13 Appraising the quality of evidence by outcomes

2.3.321 Intervention reviews

The evidence for outcomes from the included RCTs 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 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.

2.3.3.1451 Risk of bias

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

each have a score of −1 for that outcome, the overall score for that outcome would tend towards −1.

Table 3: Principle domains of bias in randomised controlled trials

Limitation	Explanation
Selection bias (sequence generation and allocation concealment)	If those enrolling patients are aware of the group to which the next enrolled patient will be allocated, either because of a non-random sequence that is predictable, or because a truly random sequence was not concealed from the researcher, this may translate into systematic selection bias. This may occur if the researcher chooses not to recruit a participant into that specific group because of: • knowledge of that participant's likely prognostic characteristics, and • a desire for one group to do better than the other.
Performance and detection bias (lack of blinding of patients and healthcare professionals)	Patients, caregivers, those adjudicating or recording outcomes, and data analysts should not be aware of the arm to which patients are allocated. Knowledge of the group can influence: • the experience of the placebo effect • performance in outcome measures • the level of care and attention received, and • the methods of measurement or analysis all of which can contribute to systematic bias.
Attrition bias	Attrition bias results from an unaccounted for loss of data beyond a certain level (a differential of 10% between groups). Loss of data can occur when participants are compulsorily withdrawn from a group by the researchers (for example, when a per-protocol approach is used) or when participants do not attend assessment sessions. If the missing data are likely to be different from the data of those remaining in the groups, and there is a differential rate of such missing data from groups, systematic attrition bias may result.
Selective outcome reporting	Reporting of some outcomes and not others on the basis of the results can also lead to bias, as this may distort the overall impression of efficacy.
Other limitations	For example: • Stopping early for benefit observed in randomised trials, in particular in the absence of adequate stopping rules. • Use of unvalidated patient-reported outcome measures. • Lack of washout periods to avoid carry-over effects in crossover trials. • Recruitment bias in cluster-randomised trials.

2.3.3.142 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.

2.3.3.113 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.

Heterogeneity or inconsistency amongst studies was also visually inspected. Where statistical heterogeneity as defined above was present or there was clear visual heterogeneity not captured in the I^2 value predefined subgrouping of studies was carried out according to the protocol. Where statistical heterogeneity was present, but the point estimates were all consistent with the same clinical interpretation, the outcome was not downgraded for inconsistency. See the review protocols for the subgrouping strategy.

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.

2.3.3.214 Imprecision

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

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

In the absence of values identified in the literature, the alternative approach to deciding on MID levels is to use the GRADE '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. There aren't established default values for ORs and the same values (0.8 and 1.25) are applied here but are acknowledged as arbitrary thresholds agreed by the committee.
- 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. Half of the median comparator group standard deviations were taken as the MID to maintain a consistent approach, as baseline values were not reported for all outcomes.
- 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.

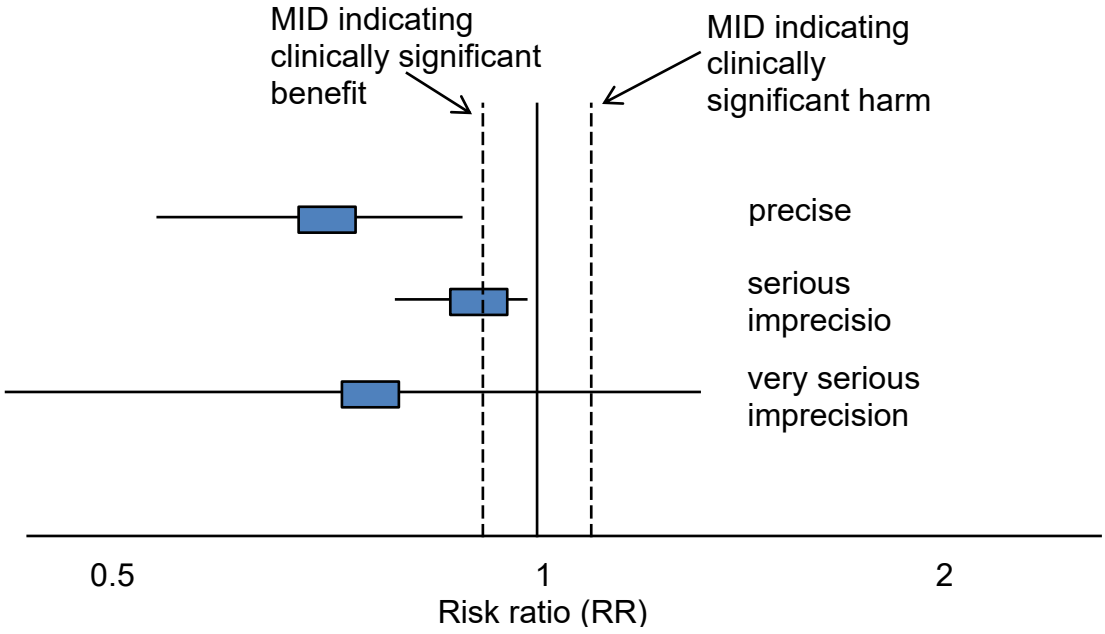
The default MID value was subject to amendment after discussion with the committee. If the committee decided that the MID level should be altered, after consideration of absolute as well as relative effects, this was allowed, provided that any such decision was not influenced by any bias towards making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

For this guideline, the following deviations from the default MIDs were used:

- SF-36 values published in the SF36v2 Health Survey Users manual.[2]
- 0.03 for the EQ-5D, this MID has been used in previous NGC guidance based on consensus.

These values were used to assess imprecision and clinical importance (see section 2.3.5 below). No appropriate MIDs for other continuous or dichotomous outcomes were found in the literature, and so the default method was adopted for these outcomes.

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



2.3.3.115 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.

Table 4: Overall quality of outcome evidence in GRADE

Level	Description
High	Further research is very unlikely to change our confidence in the estimate of effect
Moderate	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
Low	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
Very low	Any estimate of effect is very uncertain

2.3.322 Prognostic reviews

An adapted GRADE profile was used for quality assessment per outcome. 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.

2.3.3.211 Risk of bias

The risk of bias for prognostic studies was evaluated according to the QUIPS checklist, the main criteria are given in Table 5.

Table 5: Description of risk of bias criteria for prognostic studies

Risk of bias	Aim of section
Study participation	To judge selection bias (likelihood that relationship between the prognostic factor and outcome is different for participants and eligible non-participants)
Study attrition	To judge the risk of attrition bias (likelihood that relationship between prognostic factor and outcome are different for completing and non-completing participants).
Prognostic factor measurement	To judge the risk of measurement bias related to how the prognostic factor was measured (differential measurement of prognostic factor related to the baseline level of outcome).
Outcome measurement	To judge the risk of bias related to the measurement of outcome (differential measurement of outcome related to the baseline level of prognostic factor).
Study confounding	To judge the risk of bias due to confounding (i.e. the effect of the prognostic factor is distorted by another factor that is related to the prognostic factor and outcome).
Statistical Analysis and Reporting	To judge the risk of bias related to the statistical analysis and presentation of results.

2.3.3.252 Inconsistency

Inconsistency was assessed as for intervention studies.

2.3.3.273 Imprecision

In meta-analysed outcomes, or for non-pooled outcomes, 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.

2.3.3.224 Overall grading

Quality rating was assigned by study. However if there was more than 1 outcome involved in a study, then the quality rating of the evidence statements for each outcome was adjusted accordingly. For example, if one outcome was based on an invalidated measurement method, but another outcome in the same study was not, the second outcome would be graded 1 grade higher than the first outcome.

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 variable of interest then randomisation would be inappropriate as it can only be applied to 1 of the risk factors.

2.3.353 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 the Critical Appraisal Skills Programme (CASP) qualitative checklist.
Coherence	The extent to which the reviewer is able to identify a clear pattern across the studies included in the review, if there is variation present and whether this variation is explained by the contributing study authors.
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.

2.3.3.391 Methodological limitations

Each review finding had its methodological limitations assessed within each study first using the CASP checklist. Based on the degree of methodological limitations, studies were evaluated as having minor, moderate or severe limitations. A summary of the domains and questions covered is given below.

Table 7: Description of limitations assessed in the CASP checklist for qualitative studies

Domain	Aspects considered
Are the results valid?	• Was there a clear statement of the aims of the research? • Is qualitative methodology appropriate? • Was the research design appropriate to address the aims of the research? • Was the recruitment strategy appropriate to the aims of the research? • Was the data collected in a way that addressed the research issue? • Has the relationship between researcher and participants been adequately considered?
What are the results?	Have ethical issues been taken into consideration? Was the data analysis sufficiently rigorous? Is there a clear statement of findings?
Will the results help locally?	How valuable is the research?

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) was taken into account when giving an overall rating. The importance of each limitation was considered in relation to the extent to which it affected confidence in the review finding. For example, a

rating of minor limitations from CASP may lead to 'no or very minor limitations' or 'minor limitations' in the CERQual domain.

2.3.3.32 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 no or very minor, minor, moderate or major concerns about coherence.

2.3.3.33 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, timing) specified in the protocol. As such, relevance is dependent on the individual review and discussed with the guideline committee. Each review finding was given a rating of no or very minor, minor, moderate or major concerns about relevance.

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

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

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

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

Level	Description
High confidence	It is highly likely that the review finding is a reasonable representation of the phenomenon of interest.
Moderate confidence	It is likely that the review finding is a reasonable representation of the phenomenon of interest.
Low confidence	It is possible that the review finding is a reasonable representation of the phenomenon of interest.

Level	Description
Very low confidence	It is not clear whether the review finding is a reasonable representation of the phenomenon of interest.

2.3.14 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 as a starting point for the committee to begin their discussions, but then take into account the absolute effects, imprecision around this, as well as the quality of evidence and other factors such as size of the evidence base when forming the recommendations (see section 2.5). The absolute effect is considered for intervention reviews, and a consistent approach is applied 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 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. Established MIDs were found in the literature for the SF-36 and a consensus MID for the EQ-5D agreed for previous NICE guideline development was used. The committee were aware of a large body of literature reporting various MIDs for pain reduction. The values suggested vary considerably across the literature and the committee agreed that while evidence for acute pain was more consistent, there was not one consistently accepted value for a between group MID for chronic pain. It was noted that literature suggests this varies according to baseline pain and chronicity of the condition and therefore the GRADE default MID process was agreed as the most appropriate approach to take for pain outcomes in this guideline. This is consistent with choice of MID in previous NICE chronic pain guidelines.

The published or pre-agreed values used for imprecision and clinical importance are provided in Table 9. For continuous outcomes where the GRADE default MID has been used, the values for each outcome are provided in tables as an appendix in the relevant evidence review.

Table 9: MIDs

Outcome measure	MID	Source
EQ-5D	0.03	Consensus pragmatic MID used in previous NGC guidelines
SF36	Physical component summary: 2 Mental component summary: 3 Physical functioning: 3 Role-physical: 3 Bodily pain: 3 General health: 2 Vitality: 2 Social functioning: 3 Role-emotional: 4	User's manual for the SF-36v2 Health Survey, Third Edition[2]

Outcome measure	MID	Source
	Mental health: 3	

2.3.15 Clinical evidence statements

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

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

2.4 Identifying and analysing evidence of cost effectiveness

The committee is required to make decisions based on the best available evidence of both clinical effectiveness and cost effectiveness. Guideline recommendations should be based on the expected costs of the different options in relation to their expected health benefits (that is, their 'cost effectiveness') rather than the total implementation cost. However, the committee will also need to be increasingly confident in the cost effectiveness of a recommendation as the cost of implementation increases. Therefore, the committee may require more robust evidence on the effectiveness and cost effectiveness of any recommendations that are expected to have a substantial impact on resources; any uncertainties must be offset by a compelling argument in favour of the recommendation. The cost impact or savings potential of a recommendation should not be the sole reason for the committee's decision.[4]

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

- Undertook a systematic review of the published economic literature.
- Undertook new 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.[4]
- Extracted key information about the studies' methods and results into health economic evidence tables (which can be found in appendices to the relevant evidence reports).
- Generated summaries of the evidence in NICE health economic evidence profile tables (included in the relevant evidence report for each review question) – see below for details.

2.4.2 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 2002 and studies from non-OECD countries or the USA were also excluded, on the basis that the applicability of such studies to the present UK NHS context is likely to be too low for them to be helpful for decision-making.

Remaining health economic studies were prioritised for inclusion based on their relative applicability to the development of this guideline and the study limitations. For example, if a high quality, directly applicable UK analysis was available, or a study was felt to be of lower methodological quality, then other less relevant studies may not have been included. Where exclusions occurred on this basis, this is noted in the relevant evidence report.

For more details about the assessment of applicability and methodological quality see Table 10 below and the economic evaluation checklist (appendix H of the NICE guidelines manual[4]) and the health economics review protocol, which can be found in each of the evidence reports.

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

2.4.2.12 NICE health economic evidence profiles

NICE health economic evidence profile tables were used to summarise cost and cost-effectiveness estimates for the included health economic studies in each evidence review report. The health economic evidence profile shows an assessment of applicability and methodological quality for each economic study, with footnotes indicating the reasons for the assessment. These assessments were made by the health economist using the economic evaluation checklist from the NICE guidelines manual.[4] It also shows the incremental costs, incremental effects (for example, quality-adjusted life years [QALYs]) and incremental cost-effectiveness ratio (ICER) for the base case analysis in the study, as well as information about the assessment of uncertainty in the analysis. See Table 10 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.[7]

Table 10: 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)]

Item	Description
	• Minor limitations – the study meets all quality criteria, or fails to meet 1 or more quality criteria, but this is unlikely to change the conclusions about cost effectiveness. • Potentially serious limitations – the study fails to meet 1 or more quality criteria, and this could change the conclusions about cost effectiveness. • Very serious limitations – the study fails to meet 1 or more quality criteria, and this is highly likely to change the conclusions about cost effectiveness. Such studies would usually be excluded from the review.
Other comments	Information about the design of the study and particular issues that should be considered when interpreting it.
Incremental cost	The mean cost associated with one strategy minus the mean cost of a comparator strategy.
Incremental effects	The mean QALYs (or other selected measure of health outcome) associated with one strategy minus the mean QALYs of a comparator strategy.
Cost effectiveness	Incremental cost-effectiveness ratio (ICER): the incremental cost divided by the incremental effects (usually in £ per QALY gained).
Uncertainty	A summary of the extent of uncertainty about the ICER reflecting the results of deterministic or probabilistic sensitivity analyses, or stochastic analyses of trial data, as appropriate.

(a) Applicability and limitations were assessed using the economic evaluation checklist in appendix H of the NICE guidelines manual[4]

2.432 Undertaking new health economic analysis

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

The committee identified exercise as one of the priority areas for original health economic modelling. The guideline systematic review of the published clinical evidence showed a benefit of exercise compared to usual care in reducing pain and improving quality of life. Two economic evaluations were identified. One was a UK within-trial analysis with the intervention being a gym-based exercise program. The committee view was that this study was quite different to most of the other studies in the clinical review, which tended to be structured class-based interventions, generally group based, with varying frequency/intensity. This found that at follow up (30 months) exercise was not cost effective in the base case analysis using complete case data, but it was cost effective when using imputed data. A second Spanish economic evaluation was identified, which was a within trial analysis comparing 8 months of group pool-based exercised to usual care. This found exercise to be cost effective, although the staff costs were very low compared to UK costs so cost effectiveness was uncertain from this study. Pool-based exercises are not considered to be current practice in the UK because they have higher costs. Both studies had limitations regarding their generalisability because of the types of interventions analysed, and uncertainty remained around cost effectiveness. These factors, alongside a potential resource impact because of the variable use of exercise in practice and the population size, meant that this area was a modelling priority.

The second area identified as a priority was acupuncture. The guideline systematic review of the published clinical evidence showed a benefit of acupuncture compared to both sham and usual care in reducing pain and improving quality of life. Two economic evaluations were identified for this review. One UK-based within-trial economic analysis compared acupuncture in addition to usual care with usual care. This was in people with chronic neck pain, and had a 1 year follow up, although the intervention itself was around 5 months long (up to 12 x 50-minute treatments delivered once per week and then once every 2 weeks).

The study found that acupuncture was cost effective in the complete case analysis, but not when missing data was imputed (and 40% of data was missing in the acupuncture arm). Both ICERs had very large confidence intervals leading to uncertainty around cost effectiveness, although this would be the more relevant study as it is from a UK perspective. The costs of providing acupuncture seemed lower than current staff costs that might provide acupuncture in the NHS. A second study was a German within-trial analysis, comparing acupuncture to a waiting list control in people with chronic neck pain, with a 3 month follow-up. People in the acupuncture group received between 10 to 15 sessions of acupuncture. This paper found that acupuncture was cost effective compared to waiting list control. Although acupuncture costs were arbitrarily derived because acupuncture is not reimbursed by health insurance companies in Germany, and the costs per session seem lower than UK costs. Both studies had limitations regarding intervention costs potentially being underestimated, and uncertainty remained around cost effectiveness. Therefore, these reasons, alongside the fact that acupuncture for chronic pain is not currently used in the NHS and a recommendation could have a resource impact, meant that this area was prioritised for new economic modelling.

Note that where clinical evidence had both a sham and usual care comparison, only studies with a usual care comparison were used in the economic analysis, as the committee agreed trials versus usual care/no treatment (pragmatic trials), were the most appropriate comparator for the economic analysis as these would give the full benefit likely to be achieved in a real world scenario.

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.[4, 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 exercise are described in a separate economic analysis report.

2.4.3 Cost-effectiveness criteria

NICE's report 'Social value judgements: principles for the development of NICE guidance' sets out the principles that committees should consider when judging whether an intervention offers good value for money.[5] In general, an intervention was considered to be cost effective (given that the estimate was considered plausible) if either of the following criteria applied:

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

If the committee recommended an intervention that was estimated to cost more than £20,000 per QALY gained, or did not recommend one that was estimated to cost less than £20,000

per QALY gained, the reasons for this decision are discussed explicitly in 'The committee's discussion of the evidence' section of the relevant evidence report, with reference to issues regarding the plausibility of the estimate or to the factors set out in 'Social value judgements: principles for the development of NICE guidance'.[5]

When QALYs or life years gained are not used in the analysis, results are difficult to interpret unless 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.

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–J]).
- Evidence tables of the clinical and health economic evidence reviewed from the literature. All evidence tables can be found in appendices to the relevant evidence reports.
- Forest plots (in appendices to the relevant evidence reports).
- A description of the methods and results of the cost-effectiveness analysis undertaken for the guideline (in a separate economic analysis report).

Decisions on whether a recommendation could be made, and if so in which direction, were made 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. The net clinical benefit over harm (clinical effectiveness) was considered, focusing on the critical outcomes alongside the magnitude of the effect (or clinical importance), quality of evidence (including the uncertainty) and amount of evidence available. 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 the clinical harms were judged by the committee to outweigh any clinical benefits, they considered making a recommendation not to offer an intervention. This was dependant on whether the intervention had any reasonable prospect of providing cost-effective benefits to people using services and whether stopping the intervention was likely to cause harm for people already receiving it.

When clinical and health economic evidence was of poor quality, conflicting or absent, the committee decided on whether a recommendation could be made based on its expert opinion. The considerations for making consensus-based recommendations include the balance between potential harms and benefits, the economic costs compared to the economic benefits, current practices, recommendations made in other relevant guidelines, patient preferences and equality issues. The consensus recommendations were agreed

through discussions in the committee. The committee also considered whether the uncertainty was sufficient to justify delaying making a recommendation to await further research, taking into account the potential harm of failing to make a clear recommendation (see section 2.5.1 below).

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

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

- The actions health professionals need to take.
- The information readers need to know.
- The strength of the recommendation (for example the word 'offer' was used for strong recommendations and 'consider' for weaker recommendations).
- The involvement of patients (and their carers if needed) in decisions on treatment and care.
- Consistency with NICE's standard advice on recommendations about drugs, waiting times and ineffective interventions (see section 9.2 in the NICE guidelines manual[4]).

The classification 'Chronic Primary Pain' has been added in ICD-11 to reflect the pain conditions that are not able to be classified under the previous system (ICD-10). This includes some conditions that may appear aetiologically dissimilar. In the committee's view, for the majority of topics reviewed response to treatment could be assumed to be sufficiently similar for the conditions falling under the diagnostic construct of Chronic Primary Pain to allow recommendations to be made across all Chronic Primary Pain conditions, even where evidence was only available for one or more of these conditions. Where the committee thought there was reason to distinguish between Chronic Primary Pain conditions, this is reflected in the recommendations.

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.

2.5.12 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.

2.5.13 Updating the guideline

Following publication, and in accordance with the NICE guidelines manual, NICE will undertake a review of whether the evidence base has progressed significantly to alter the guideline recommendations and warrant an update.

2.5.14 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.

2.5.15 Funding

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

3 Additional information

3.1 Extracting outcome data from studies reporting multiple measures of similar outcome constructs

Where studies reported multiple measures of review protocol outcomes, a single measure was preferentially extracted according to a hierarchy which was agreed by the committee. This hierarchy is presented for critical outcomes in Table 11 below.

Where multicomponent measures were reported, individual sub scales were extracted in the absence of total scores. Where not all sub scales were reported, these were not extracted unless they were relevant to other protocol outcomes and the study did not report other measures that were higher in the hierarchy. For example, the SF36 is a validated measure of quality of life and was extracted for this outcome where physical and mental component summaries or all individual sub scales were reported. However, if only the physical function sub scale was reported in the absence of other measures of physical function in the hierarchy, then it would be extracted under the physical function protocol outcome.

Table 11: Hierarchy of outcome measures

Hierarchy	Quality of life	Pain scales	Physical function	Psychological distress
Highest to lowest ↓	EQ-5D	VAS/NRS/MPI/BPI	Pain/Oswestry/Neck disability index	HADS
	SF36	McGill pain questionnaire	Roland Morris Disability Questionnaire (RMDQ)	BDI/BAI/Hamilton depression rating scale/ Hamilton anxiety rating scale
	SF12	Symptom severity scores e.g. Prostatitis severity scale	5 minute walk/sit to stand	GAD-7/10/ PHQ9/5/DASS/Geriatric depression scale/ State-Trait Anxiety Inventory
	FIQ	Responder analyses (30% or 50% improvement)	Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)	Centre for epidemiological studies depression scale/General health questionnaire
			Canadian Occupational Performance Measure	Pain catastrophizing scale
			Northwick Park questionnaire	Profile of mood states

SF36; short form 36, SF12; short form 12, FIQ; fibromyalgia impact questionnaire, VAS; visual analogue scale, NRS; numeric rating scale, MPI; multidimensional pain inventory, BPI; brief pain inventory, HADS; Hospital anxiety and depression scale, BDI; Beck depression inventory, BAI; Beck anxiety inventory, GAD; General anxiety disorder-7 (or -10), PHQ9; Patient health questionnaire-9, DASS; depression and anxiety stress scales.

3.2 Concomitant interventions

Where other interventions were received in addition to the interventions/combination of interventions specified in the review protocols, these studies were excluded unless the additional interventions were received in both study groups and considered to be very low

intensity or part of usual care. For example, additional education, lifestyle advice or basic stretching exercises were not excluded.

3.3 Categorising exercise interventions

The committee pre-specified categories of exercise interventions within Evidence review E which assessed the clinical and cost effectiveness of exercise interventions for the management of chronic primary pain. However, sometimes interventions within studies were defined by study authors as one particular category but clearly from their description incorporated elements of other types of exercise. The technical team and committee considered these studies on a case-by-case basis and categorised each intervention based on the elements of exercise that made up a large component of the intervention. For example, if an intervention included a 5-minute cool down and stretching after a 30 minute aerobic exercise session, this would have been classed as aerobic exercise only.

4 Acronyms and abbreviations

2

Acronym or abbreviation	Description
ACT	Acceptance and commitment therapy
BNF	British National Formulary
BDI	Beck depression inventory
BPI	Brief pain inventory
CBT	Cognitive behavioural therapy
CI	Confidence interval
COMET	Core Outcome Measures in Effectiveness Trials
CUA	Cost-utility analysis
EMDR	Eye movement desensitisation reprocessing
EMG	Electromyogram
EQ-5D	EuroQol 5-dimension
FIQ	Fibromyalgia Impact Questionnaire
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HADS	Hospital Anxiety and Depression Scale
HR	Hazard ratio
ICER	Incremental cost-effectiveness ratio
IMMPACT	Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials
MPI	Multidimensional pain inventory
NGC	National Guideline Centre
NICE	National Institute for Health and Care Excellence
NIH-CPSI	NIH-Chronic Prostatitis Symptom Index
NSAID	Non-steroidal anti-inflammatory drug
OECD	Organisation for Economic Co-operation and Development
OR	Odds ratio
PENS	Percutaneous electrical nerve stimulation
QALY	Quality-adjusted life year
QoL	Quality of life
RCT	Randomised controlled trial
RR	Risk ratio
SF-12	12-Item Short Form Health Survey
SF-36	36-Item Short Form Health Survey
SMT	Spinal manipulation therapy
SNRI	Serotonin norepinephrine re-uptake inhibitor
SSRI	Selective serotonin reuptake inhibitor
TCA	Tricyclic antidepressant
TCBT	Telephone-delivered cognitive behaviour therapy
TDCS	Transcranial direct current stimulation
TENS	Transcutaneous electrical nerve stimulation
TMS	Transcranial magnetic stimulation

Acronym or abbreviation	Description
VAS	Visual analogue scale

5 Glossary

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

5.1 Guideline-specific terms

Term	Definition
Acceptance and commitment therapy (ACT)	An empirically-based psychological intervention that uses acceptance and mindfulness strategies, with commitment and behaviour change strategies, to increase psychological flexibility.
Acupuncture	A treatment derived from ancient Chinese medicine in which fine needles are inserted at certain sites in the body for therapeutic or preventative purposes.
Biofeedback	A process whereby electronic monitoring of a normally automatic bodily function is used to train someone to acquire voluntary control of that function.
Cannabinoid	This guideline covers the oral cannabinoids nabilone and nabixamols oromucosal spray.
Care plan	An agreement between patient and health or social care professional to support management of day to day health and symptoms by the patient and other healthcare professionals and/or to organise care. It can be a written document and/or something recorded in patient notes.
Chronic pain	Pain that persists or recurs for longer than 3 months.
Chronic primary pain	Defined in this guidance as chronic pain in one or more anatomical regions that is characterized by significant emotional distress (anxiety, anger/frustration or depressed mood) and functional disability (interference in daily life activities and reduced participation in social roles). The diagnosis is appropriate independently of identified biological or psychological contributors unless another diagnosis would better account for the presenting symptoms. Includes chronic widespread pain, complex regional pain syndrome, chronic visceral pain, chronic orofacial pain, chronic primary musculoskeletal pain other than orofacial.
Co-prescribing	The prescription of two or more medicine classes for the same indication.
Cognitive behavioural therapy (CBT)	Cognitive approaches are aimed at altering unhelpful or inappropriate beliefs as a basis for changing behaviour, such as fear-avoidance.
Dry needling	A technique similar to acupuncture (see above). Dry needling is sometimes also known as myofascial trigger point dry needling or intramuscular stimulation (IMS).
Electroacupuncture	A form of acupuncture (see above) where a small electric current is passed through the acupuncture needles.
Manipulation/mobilisation	Treatments involving moving joints, including spinal manipulation therapy (SMT) and Maitland technique.
Manual therapy	Includes a range of treatments aimed at improving the mobility of joints, muscles and soft tissue and decreasing pain. See also manipulation/mobilisation, and soft tissue technique.
Mindfulness	Therapy to make the patient aware of the present moment, and non-judgmentally to the unfolding of experience moment by moment to alter behaviours towards pain.

Term	Definition
Mixed modality manual therapy	In this guideline mixed modality manual therapy refer to soft tissue technique with or without traction; and with or without manipulation/mobilisation.
Pain management programme	In this guideline a pain management programme refers to any intervention that has two or more components including a physical and a psychological component delivered by trained people, with some interaction/coordination between the two.
Psychotherapy	The use of psychodynamically informed methods based on communication within a therapeutic alliance to help a person change their behaviour and overcome problems.
Social interventions/social prescribing	This guideline includes interventions with a social element that aim to improve quality of life for people with chronic pain, for example social prescribing, cultural commissioning, health training and coaching, case management, vocational rehabilitation, befriending and advocacy.
Soft tissue technique	Describes treatments where a direct physical pressure is applied to muscle and other soft tissues with the aim of improving mobility or circulation, for example massage, muscle energy technique, and myofascial/trigger point release.
Supervised group exercise programme	Encompasses varied group physical exercise programmes run by a professional.
Traction	The application of a sustained pull on a limb or muscle.

5.2 General terms

2

Term	Definition
Abstract	Summary of a study, which may be published alone or as an introduction to a full scientific paper.
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

Term	Definition
	works when it does not.) Bias can occur by chance, deliberately or as a result of systematic errors in the design and execution of a study. It can also occur at different stages in the research process, for example, during the collection, analysis, interpretation, publication or review of research data. For examples see selection bias, performance bias, information bias, confounding factor, and publication bias.
Blinding	A way to prevent researchers, doctors and patients in a clinical trial from knowing which study group each patient is in so they cannot influence the results. The best way to do this is by sorting patients into study groups randomly. The purpose of 'blinding' or 'masking' is to protect against bias. A single-blinded study is one in which patients do not know which study group they are in (for example whether they are taking the experimental drug or a placebo). A double-blinded study is one in which neither patients nor the researchers and doctors know which study group the patients are in. A triple blind study is one in which neither the patients, clinicians or the people carrying out the statistical analysis know which treatment patients received.
Carer (caregiver)	Someone who looks after family, partners or friends in need of help because they are ill, frail or have a disability.
Case–control study	A study to find out the cause(s) of a disease or condition. This is done by comparing a group of patients who have the disease or condition (cases) with a group of people who do not have it (controls) but who are otherwise as similar as possible (in characteristics thought to be unrelated to the causes of the disease or condition). This means the researcher can look for aspects of their lives that differ to see if they may cause the condition. For example, a group of people with lung cancer might be compared with a group of people the same age that do not have lung cancer. The researcher could compare how long both groups had been exposed to tobacco smoke. Such studies are retrospective because they look back in time from the outcome to the possible causes of a disease or condition.
Case series	Report of a number of cases of a given disease, usually covering the course of the disease and the response to treatment. There is no comparison (control) group of patients.
Clinical efficacy	The extent to which an intervention is active when studied under controlled research conditions.
Clinical effectiveness	How well a specific test or treatment works when used in the 'real world' (for example, when used by a doctor with a patient at home), rather than in a carefully controlled clinical trial. Trials that assess clinical effectiveness are sometimes called management trials. Clinical effectiveness is not the same as efficacy.
Clinician	A healthcare professional who provides patient care. For example, a doctor, nurse or physiotherapist.
Cochrane Review	The Cochrane Library consists of a regularly updated collection of evidence-based medicine databases including the Cochrane Database of Systematic Reviews (reviews of randomised controlled trials prepared by the Cochrane Collaboration).
Cohort study	A study with 2 or more groups of people – cohorts – with similar characteristics. One group receives a treatment, is exposed to a risk factor or has a particular symptom and the other group does not. The study follows their progress over time and records what happens. See also observational study.

Term	Definition
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	A group of people in a study who do not receive the treatment or test being studied. Instead, they may receive the standard treatment (sometimes called 'usual care') or a dummy treatment (placebo). The results for the control group are compared with those for a group receiving the treatment being tested. The aim is to check for any differences. Ideally, the people in the control group should be as similar as possible to those in the treatment group, to make it as easy as possible to detect any effects due to the treatment.
Cost–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

Term	Definition
	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-effectiveness plane	In health economics, the cost-effectiveness plane is used to visually represent the differences in costs and health outcomes between treatment alternatives in two dimensions, by plotting the costs against effects on a graph.
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
Discounting	Costs and perhaps benefits incurred today have a higher value than costs and benefits occurring in the future. Discounting health benefits reflects individual preference for benefits to be experienced in the present rather than the future. Discounting costs reflects individual preference for costs to be experienced in the future rather than the present.
Disutility	The loss of quality of life associated with having a disease or condition. See Utility
Dominance	A health economics term. When comparing tests or treatments, an option that is both less effective and costs more is said to be ‘dominated’ by the alternative.
Drop-out	A participant who withdraws from a trial before the end.
Economic evaluation	An economic evaluation is used to assess the cost effectiveness of healthcare interventions (that is, to compare the costs and benefits of a healthcare intervention to assess whether it is worth doing). The aim of an economic evaluation is to maximise the level of benefits – health effects – relative to the resources available. It should be used to inform and support the decision-making process; it is not supposed to replace the judgement of healthcare professionals. There are several types of economic evaluation: cost–benefit analysis, cost–consequences analysis, cost-effectiveness analysis, cost-minimisation analysis and cost–utility analysis. They use similar methods to define and evaluate costs, but differ in the way they estimate the benefits of a particular drug, programme or intervention.
Effect (as in effect measure, treatment effect, estimate of effect, effect size)	A measure that shows the magnitude of the outcome in one group compared with that in a control group. For example, if the absolute risk reduction is shown to be 5% and it is the outcome of interest, the effect size is 5%.

Term	Definition
	The effect size is usually tested, using statistics, to find out how likely it is that the effect is a result of the treatment and has not just happened by chance (that is, to see if it is statistically significant).
Effectiveness	How beneficial a test or treatment is under usual or everyday conditions, compared with doing nothing or opting for another type of care.
Efficacy	How beneficial a test, treatment or public health intervention is under ideal conditions (for example, in a laboratory), compared with doing nothing or opting for another type of care.
Epidemiological study	The study of a disease within a population, defining its incidence and prevalence and examining the roles of external influences (for example, infection, diet) and interventions.
EQ-5D (EuroQol 5 dimensions)	A standardised instrument used to measure health-related quality of life. It provides a single index value for health status.
Evidence	Information on which a decision or guidance is based. Evidence is obtained from a range of sources including randomised controlled trials, observational studies, expert opinion (of clinical professionals or patients).
Exclusion criteria (literature review)	Explicit standards used to decide which studies should be excluded from consideration as potential sources of evidence.
Exclusion criteria (clinical study)	Criteria that define who is not eligible to participate in a clinical study.
Extended dominance	If Option A is both more clinically effective than Option B and has a lower cost per unit of effect, when both are compared with a do-nothing alternative then Option A is said to have extended dominance over Option B. Option A is therefore cost effective and should be preferred, other things remaining equal.
Extrapolation	An assumption that the results of studies of a specific population will also hold true for another population with similar characteristics.
Follow-up	Observation over a period of time of an individual, group or initially defined population whose appropriate characteristics have been assessed in order to observe changes in health status or health-related variables.
Generalisability	The extent to which the results of a study hold true for groups that did not participate in the research. See also external validity.
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.

Term	Definition
Imprecision	Results are imprecise when studies include relatively few patients and few events and thus have wide confidence intervals around the estimate of effect.
Imputation	In statistics, imputation is the process of replacing missing data with substituted values.
Inclusion criteria (literature review)	Explicit criteria used to decide which studies should be considered as potential sources of evidence.
Incremental analysis	The analysis of additional costs and additional clinical outcomes with different interventions.
Incremental cost	The extra cost linked to using one test or treatment rather than another. Or the additional cost of doing a test or providing a treatment more frequently.
Incremental cost-effectiveness ratio (ICER)	The difference in the mean costs in the population of interest divided by the differences in the mean outcomes in the population of interest for one treatment compared with another.
Incremental net benefit (INB)	The value (usually in monetary terms) of an intervention net of its cost compared with a comparator intervention. The INB can be calculated for a given cost-effectiveness (willingness to pay) threshold. If the threshold is £20,000 per QALY gained then the INB is calculated as: (£20,000 × QALYs gained) − Incremental cost.
Indirectness	The available evidence is different to the review question being addressed, in terms of PICO (population, intervention, comparison and outcome).
Intention-to-treat analysis (ITT)	An assessment of the people taking part in a clinical trial, based on the group they were initially (and randomly) allocated to. This is regardless of whether or not they dropped out, fully complied with the treatment or switched to an alternative treatment. Intention-to-treat analyses are often used to assess clinical effectiveness because they mirror actual practice: that is, not everyone complies with treatment and the treatment people receive may be changed according to how they respond to it.
Intervention	In medical terms this could be a drug treatment, surgical procedure, diagnostic or psychological therapy. Examples of public health interventions could include action to help someone to be physically active or to eat a more healthy diet.
Kappa statistic	A statistical measure of inter-rater agreement that takes into account the agreement occurring by chance.
Licence	See ‘Product licence’.
Life years gained	Mean average years of life gained per person as a result of the intervention compared with an alternative intervention.
Likelihood ratio	The likelihood ratio combines information about the sensitivity and specificity. It tells you how much a positive or negative result changes the likelihood that a patient would have the disease. The likelihood ratio of a positive test result (LR+) is sensitivity divided by (1 minus specificity).
Long-term care	Residential care in a home that may include skilled nursing care and help with everyday activities. This includes nursing homes and residential homes.
Logistic regression or Logit model	In statistics, logistic regression is a type of analysis used for predicting the outcome of a binary dependent variable based on one or more predictor variables. It can be used to estimate the log of the odds (known as the ‘logit’).
Loss to follow-up	A patient, or the proportion of patients, actively participating in a clinical trial at the beginning, but whom the researchers were unable to trace or contact by the point of follow-up in the trial

Term	Definition
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.
Net monetary benefit (NMB)	The value in monetary terms of an intervention net of its cost. The NMB can be calculated for a given cost-effectiveness threshold. If the threshold is £20,000 per QALY gained then the NMB for an intervention is calculated as: (£20,000 × mean QALYs) − mean cost. The most preferable option (that is, the most clinically effective option to have an ICER below the threshold selected) will be the treatment with the highest NMB.
Non-randomised intervention study	A quantitative study investigating the effectiveness of an intervention that does not use randomisation to allocate patients (or units) to treatment groups. Non-randomised studies include observational studies, where allocation to groups occurs through usual treatment decisions or people's preferences. Non-randomised studies can also be experimental, where the investigator has some degree of control over the allocation of treatments. Non-randomised intervention studies can use a number of different study designs, and include cohort studies, case–control studies, controlled before-and-after studies, interrupted-time-series studies and quasi-randomised controlled trials.
Number needed to treat (NNT)	The average number of patients who need to be treated to get a positive outcome. For example, if the NNT is 4, then 4 patients would have to be treated to ensure 1 of them gets better. The closer the NNT is to 1, the better the treatment. For example, if you give a stroke prevention drug to 20 people before 1 stroke is prevented, the number needed to treat is 20. See also number needed to harm, absolute risk reduction.
Observational study	Individuals or groups are observed or certain factors are measured. No attempt is made to affect the outcome. For example, an observational study of a disease or treatment would allow 'nature' or usual medical care to take its course. Changes or differences in one characteristic (for example, whether or not people received a specific treatment or intervention) are studied without intervening. There is a greater risk of selection bias than in experimental studies.
Odds ratio	Odds are a way to represent how likely it is that something will happen (the probability). An odds ratio compares the probability of something in one group with the probability of the same thing in another. An odds ratio of 1 between 2 groups would show that the probability of the event (for example a person developing a disease, or a treatment working) is the same for both. An odds ratio greater than 1 means the event is more likely in the first group. An odds ratio less than 1 means that the event is less likely in the first group. Sometimes probability can be compared across more than 2 groups – in this case, one of the groups is chosen as the 'reference category', and the odds ratio is calculated for each group compared with the reference category. For example, to compare the risk of dying from lung cancer for non-smokers, occasional smokers and regular smokers, non-smokers could be used as the reference category. Odds ratios would be worked out for occasional smokers

Term	Definition
	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.
Parameterised	In economic evaluation, information on uncertainty around a point estimate is needed to create a distribution around the point estimate to run probabilistic analysis. Point estimates are parameterised using information on the uncertainty around a point estimate, such as standard deviation.
Placebo	A fake (or dummy) treatment given to participants in the control group of a clinical trial. It is indistinguishable from the actual treatment (which is given to participants in the experimental group). The aim is to determine what effect the experimental treatment has had – over and above any placebo effect caused because someone has received (or thinks they have received) care or attention.
Polypharmacy	The use or prescription of multiple medications.
Posterior distribution	In Bayesian statistics this is the probability distribution for a statistic based after combining established information or belief (the prior) with new evidence (the likelihood).
Postoperative	Pertaining to the period after patients leave the operating theatre, following surgery.
Post intervention	The measurement of outcomes at the end of an intervention. For example at intervention at 12 weeks that measured outcomes at 12 weeks.
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.

Term	Definition
Prior distribution	In Bayesian statistics this is the probability distribution for a statistic based on previous evidence or belief.
Primary care	Healthcare delivered outside hospitals. It includes a range of services provided by GPs, nurses, health visitors, midwives and other healthcare professionals and allied health professionals such as dentists, pharmacists and opticians.
Primary outcome	The outcome of greatest importance, usually the one in a study that the power calculation is based on.
Probabilistic analysis	In economic evaluation, this is an analysis that uses a probability distribution for each input. In contrast, see Deterministic analysis.
Product licence	An authorisation from the MHRA to market a medicinal product.
Prognosis	A probable course or outcome of a disease. Prognostic factors are patient or disease characteristics that influence the course. Good prognosis is associated with low rate of undesirable outcomes; poor prognosis is associated with a high rate of undesirable outcomes.
Prospective study	A research study in which the health or other characteristic of participants is monitored (or 'followed up') for a period of time, with events recorded as they happen. This contrasts with retrospective studies.
Publication bias	Publication bias occurs when researchers publish the results of studies showing that a treatment works well and don't publish those showing it did not have any effect. If this happens, analysis of the published results will not give an accurate idea of how well the treatment works. This type of bias can be assessed by a funnel plot.
Quality of life	See 'Health-related quality of life'.
Quality-adjusted life year (QALY)	A measure of the state of health of a person or group in which the benefits, in terms of length of life, are adjusted to reflect the quality of life. One QALY is equal to 1 year of life in perfect health. QALYS are calculated by estimating the years of life remaining for a patient following a particular treatment or intervention and weighting each year with a quality of life score (on a scale of 0 to 1). It is often measured in terms of the person's ability to perform the activities of daily life, freedom from pain and mental disturbance.
Randomisation	Assigning participants in a research study to different groups without taking any similarities or differences between them into account. For example, it could involve using a random numbers table or a computer-generated random sequence. It means that each individual (or each group in the case of cluster randomisation) has the same chance of receiving each intervention.
Randomised controlled trial (RCT)	A study in which a number of similar people are randomly assigned to 2 (or more) groups to test a specific drug or treatment. One group (the experimental group) receives the treatment being tested, the other (the comparison or control group) receives an alternative treatment, a dummy treatment (placebo) or no treatment at all. The groups are followed up to see how effective the experimental treatment was. Outcomes are measured at specific times and any difference in response between the groups is assessed statistically. This method is also used to reduce bias.
RCT	See 'Randomised controlled trial'.
Reporting bias	See 'Publication bias'.
Resource implication	The likely impact in terms of finance, workforce or other NHS resources.
Retrospective study	A research study that focuses on the past and present. The study examines past exposure to suspected risk factors for the disease or

Term	Definition
	condition. Unlike prospective studies, it does not cover events that occur after the study group is selected.
Review question	In guideline development, this term refers to the questions about treatment and care that are formulated to guide the development of evidence-based recommendations.
Risk ratio (RR)	The ratio of the risk of disease or death among those exposed to certain conditions compared with the risk for those who are not exposed to the same conditions (for example, the risk of people who smoke getting lung cancer compared with the risk for people who do not smoke). If both groups face the same level of risk, the risk ratio is 1. If the first group had a risk ratio of 2, subjects in that group would be twice as likely to have the event happen. A risk ratio of less than 1 means the outcome is less likely in the first group. The risk ratio is sometimes referred to as relative risk.
Secondary outcome	An outcome used to evaluate additional effects of the intervention deemed a priori as being less important than the primary outcomes.
Selection bias	Selection bias occurs if: a) The characteristics of the people selected for a study differ from the wider population from which they have been drawn, or b) There are differences between groups of participants in a study in terms of how likely they are to get better.
Sensitivity 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 ($p<0.05$).
Stakeholder	An organisation with an interest in a topic that NICE is developing a guideline or piece of public health guidance on. Organisations that register as stakeholders can comment on the draft scope and the draft guidance. Stakeholders may be: • manufacturers of drugs or equipment • national patient and carer organisations • NHS organisations • organisations representing healthcare professionals.
State transition model	See Markov model
Systematic review	A review in which evidence from scientific studies has been identified, appraised and synthesised in a methodical way according to predetermined criteria. It may include a meta-analysis.

Term	Definition
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).

1
2

References

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

2. Maruish ME. User's Manual for the SF-36v2 Health Survey. 3rd ed. Quality Metric Incorporated. 2011. Available from: https://books.google.co.uk/books?id=a0vYnQEACAAJ

3. 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

4. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual [updated October 2018]. London. National Institute for Health and Care Excellence, 2014. Available from: http://www.nice.org.uk/article/PMG20/chapter/1%20Introduction%20and%20overview

5. National Institute for Health and Clinical Excellence. Social value judgements: principles for the development of NICE guidance. 2nd ed. London. National Institute for Health and Clinical Excellence. 2008. Available from: https://www.nice.org.uk/Media/Default/About/what-we-do/Research-and-development/Social-Value-Judgements-principles-for-the-development-of-NICE-guidance.pdf

6. National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal 2013. London. National Institute for Health and Clinical Excellence, 2013. Available from: https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781

7. Organisation for Economic Co-operation and Development (OECD). Purchasing power parities (PPP). 2012. Available from: http://www.oecd.org/sdd/prices-ppp/ Last accessed: 18/01/2019

8. Review Manager (RevMan) [Computer program]. Version 5. Copenhagen. The Nordic Cochrane Centre, The Cochrane Collaboration, 2015. Available from: https://community.cochrane.org/help/tools-and-software/revman-5