National Institute for Health and Care Excellence

Draft for Consultation

Low back pain and sciatica in over 16s: assessment and management (Partial update)

NICE guideline: methods

NICE guideline NG59

Methods

July 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 All NICE guidance is subject to regular review and may be updated or withdrawn.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: TBC

Contents

1 Development of the guideline

1.1 Remit

NICE received the remit for this guideline from NHS England. NICE commissioned the National Guideline Centre to produce the guideline.

The remit for this guideline is:

To undertake a partial update of Low back pain and sciatica in over 16s: assessment and management (NG59).

1.2 What this guideline covers

The pharmacological management of sciatica.

1.3 What this guideline does not cover

- Assessment to identify low back pain and sciatica and any prognostic factors that could guide management.
- Lifestyle interventions:
  - self-management strategies, including education and advice
  - workplace interventions and return-to-work interventions (for example, occupational and ergonomic interventions).
- Pharmacological treatments for low back pain
- Non-pharmacological interventions:
  - exercise therapies
  - postural therapies
  - manual therapies
  - electrotherapy
  - orthotics and appliances
  - acupuncture
  - psychological interventions.
- Combined non-invasive therapies.
- Invasive procedures:
  - injection therapies
  - radiofrequency ablation procedures.
- Surgery:
  - indications for referral for surgery.
  - surgical interventions (fusion and disc replacement for low back pain and discectomy or laminectomy and decompression surgery for sciatica).

These areas are considered within the 2016 version of this guideline. Evidence and methods for these areas can be accessed at: https://www.nice.org.uk/guidance/ng59/evidence

1.4 Funding

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

2 Methods

This guideline was developed using the methods described in the 2014 NICE guidelines manual, updated October 2018.

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

Sections 2.1 to 2.3 describe the process used to identify and review evidence. Sections 2.1.1.1 and 2.4 describe the process used to identify and review the health economic evidence.

2.1 Developing the review questions and outcomes

The review question developed for this guideline was based on the key area identified in the surveillance review and detailed in the scope. It was drafted by the National Guideline Centre technical team and refined and validated by the committee and signed off by NICE. One review question was developed in this guideline update and outlined in Table 1.

The review question was based on the PICO framework:

Population, Intervention, Comparator and Outcome (PICO)

This use of a framework informed a more detailed protocol that guided the literature searching process, critical appraisal and synthesis of evidence, and facilitated the development of recommendations by the guideline committee. A full literature search, critical appraisal and evidence review was completed for the specified review question.

Table 1: Review questions

Type of review	Review questions	Outcomes
Intervention	What is the clinical and cost effectiveness of pharmacological treatment in the management of sciatica?	Critical outcomes: • Quality of life • Pain severity • Function • Psychological distress Important outcomes: • Healthcare utilisation • Adverse events (morbidity) • Adverse events (mortality) • Responder criteria (≥ 30% improvement in pain or function)

2.1.1 Stratification

Stratification is applied where the committee are confident the intervention will work differently in the groups and separate recommendations are required, therefore they should be reviewed separately. In this guideline all analyses were stratified for population (that is, people with low back pain, low back pain with or without sciatica, or sciatica), which meant that different studies with predominant population-groups in different population strata were not combined and analysed together.

This update only included evidence for the latter of the three strata – focussing on the treatment of sciatica. Evidence for low back pain in isolation and mixed low back pain and

sciatica populations is included within the 2016 iteration of this guideline in the review question for the pharmacological treatment of low back pain.

There was no further stratification within the review included in this update.

2.2 Searching for evidence

2.2.1 Clinical and health economics literature searches

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

Systematic literature searches were undertaken to identify all published clinical and health economic evidence relevant to the review questions. Searches were undertaken according to the parameters stipulated within the NICE guidelines manual.[4] Databases were searched using relevant medical subject headings, free-text terms and study-type filters where appropriate. Studies published in languages other than English were not reviewed, and where possible, searches were restricted to English language. All searches were conducted on 20 February 2020. If new evidence falls outside of the timeframe for the guideline searches, e.g. from stakeholder comments, the impact on the guideline will be considered, and any further action agreed between the developer and NICE staff with a quality assurance role.

Prior to running, searches were quality assured using different approaches, including checking key papers were retrieved. Medline search strategies were peer reviewed by a second information specialist using a QA processed based on PRESS checklist.[3] Additional studies were added by checking reference lists of relevant systematic reviews, and those highlighted by committee members.

2.3 Reviewing evidence

The evidence 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 were produced from data entered into EviBase, including critical appraisal ratings.
- Summaries of the evidence were generated by outcome. Outcome data were combined, analysed and reported according to study design:
    - Randomised data were meta-analysed where appropriate and reported in GRADE profile tables.
    - Data from non-randomised studies were meta-analysed where appropriate and reported in GRADE profile tables (NB. There were no non-randomised studies subsequently included in this review).
- A minimum of 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 Types of studies and inclusion and exclusion criteria

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

- People aged 16 years or above with sciatica.

The key population exclusion criteria were:

- Conditions of a non-mechanical nature, including;
  - inflammatory causes of back pain (for example, ankylosing spondylitis or diseases of the viscera)
  - serious spinal pathology (for example, neoplasms, infections or osteoporotic collapse)
  - neurological disorders (including cauda equina syndrome or mononeuritis)
  - adolescent scoliosis.
- People aged under 16 years.

Conference abstracts were not included in any of the reviews. Literature reviews, posters, letters, editorials, comment articles, unpublished studies and studies not in English were excluded.

2.3.2 Inclusion and exclusion criteria (from NG59)

For consistency with the 2016 guideline, the decision rules applied to the reviews in NG59 have been followed in this update. The relevant methodological criteria are detailed below.

2.3.1.3 Population

The following terms were considered appropriate to include to encompass sciatica:

- Sciatica/lumbago
- Radicular pain/Radiculopathy
- Pain radiating to the leg
- Neurogenic claudication
- Nerve root compression/irritation
- Spinal stenosis

Other than the excluded populations listed in the scope, the following exclusions were noted:

- Mixed populations for example, people with low back pain and neck pain (unless the results presented in the studies are split so data for people with low back pain or sciatica only is extractable).
- Pregnancy-related back pain
- Sacroiliac joint dysfunction

- Adjacent-segment disease
- Failed back surgery syndrome
- Spondylolisthesis
- Osteoarthritis.

Terms considered appropriate to encompass low back pain which are excluded from this review, and covered in NG59 include:

- Discogenic pain
- Degenerative disc disease
- Lumbar disc herniation
- Secondary to lumbar degenerative disease
- Facet joint pain.

2.3.1.12 Outcomes

Data presented in the reviews were agreed to be reported at 2 time-points; equal to or less than 4 months and greater than 4 months. For each time-point, where appropriate, data were pooled together. Where studies reported an outcome at multiple time-points within the 4 months' time-point for example, pain severity at 2 months and 3 months, the outcome closest to 4 months was extracted. Where studies reported multiple time-points at greater than 4 months, the outcome closest to 12 months was reported for example, between 6 months and 10 months, the 10 months data was extracted. However, in instances where outcomes greater than 12 months were reported, for example, 6 months and 18 months, 18 months data were extracted as this was the end of trial data and therefore more informative to the committee.

Outcomes measuring pain severity were pooled if they were on the same scale, i.e. numeric rating scale (NRS) and visual analogue scale (VAS) (both reported on a range of 0-10).

The Roland Morris Disability questionnaire (RMDQ) on a scale of 0-24 and Oswestry Disability index (ODI) on a scale of 0-100 were pooled together and presented as standardised mean difference where appropriate. In order to determine imprecision and clinical importance, the effect size was converted back on to the RMDQ 0-24 scale.

The health survey SF-36 was scored such that 8 scale scores are given: physical functioning, role physical, bodily pain, general health perceptions, vitality, social functioning, role emotional, and mental health. Two summary measures can be calculated from these scales; physical component score and the mental component score. It was agreed that where possible, all domains would be extracted for the evidence. If the individual domains were not reported, then just the two summary measures were extracted. A single overall score would not be extracted as it is not appropriate to combine the physical and mental domains.

2.3.2 Type of studies

Randomised trials or non-randomised intervention studies were included in the evidence reviews as appropriate.

For this intervention review, 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 if there was insufficient randomised evidence for the committee to make a decision. There were no relevant non-randomised studies included in this review therefore the following methodology relates to RCTs only. Refer to the review protocol in the evidence report for full details on the study design of studies that were appropriate for this review question.

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.

2.3.2 Methods of combining evidence

Data synthesis for intervention reviews

Where possible, meta-analyses were conducted using Cochrane Review Manager (RevMan5)[7] software

2.3.2.1 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 are zero events in both arms, the risk difference was calculated and reported instead.

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 for the same outcomes, standardised mean differences were used (providing all studies reported either change from baseline or final values rather than a mixture of both); each different measure in each study was 'normalised' to the standard deviation value pooled between the intervention and comparator groups in that same study.

The means and standard deviations of continuous outcomes are required for meta-analysis. However, in cases where standard deviations were not reported, the standard error was calculated if the p values or 95% confidence intervals (95% CI) were reported, and meta-analysis was undertaken with the mean and standard error using the generic inverse variance method in Cochrane Review Manager (RevMan5[7] software).

2.3.2.2 Appraising the quality of evidence by outcomes

Intervention reviews

The evidence for outcomes from the included RCTs were evaluated and presented using 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 bias is tested for when there are more than 5 studies for an outcome. This was not relevant to any outcome in this review.

2.3.263 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 using the appropriate checklist for the study design (Cochrane RoB 2 for RCTs). 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. An overall rating is calculated across all studies 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 participants 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

Limitation	Explanation
	• a desire for one group to do better than the other.
Performance and detection bias (lack of blinding)	Patients, caregivers, those adjudicating or recording outcomes, and data analysts should not be aware of the arm to which the participants 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 at least 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.

2.3.214 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. An overall rating is calculated across all studies by taking into account the weighting of studies according to 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.

2.3.255 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. Statistical heterogeneity was assessed for each meta-analysis estimate by an I-squared (I^2) inconsistency statistic.

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. See the review protocols for the subgrouping strategy.

When heterogeneity existed within an outcome (I^2>50%), but no plausible explanation could be found, the quality of evidence for that outcome was downgraded. Inconsistency for that outcome was given a 'serious' score of −1 if the I^2 was 50–74%, and a 'very serious' score of −2 if the I^2 was 75% or more.

If inconsistency could be explained based on pre-specified subgroup analysis (that is, each subgroup had an I^2<50%) then each of the derived subgroups were presented separately (providing at least 1 study remained in each subgroup). The committee took this into account and considered whether to make separate recommendations based on the variation in effect across subgroups within the same outcome. In such a situation the quality of evidence was not downgraded.

If all predefined strategies of subgrouping were unable to explain statistical heterogeneity, 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. If, however, the committee considered the heterogeneity was so large that meta-analysis was inappropriate, then the results were not pooled and were described narratively.

2.3.286 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 1.

The value / 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.

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

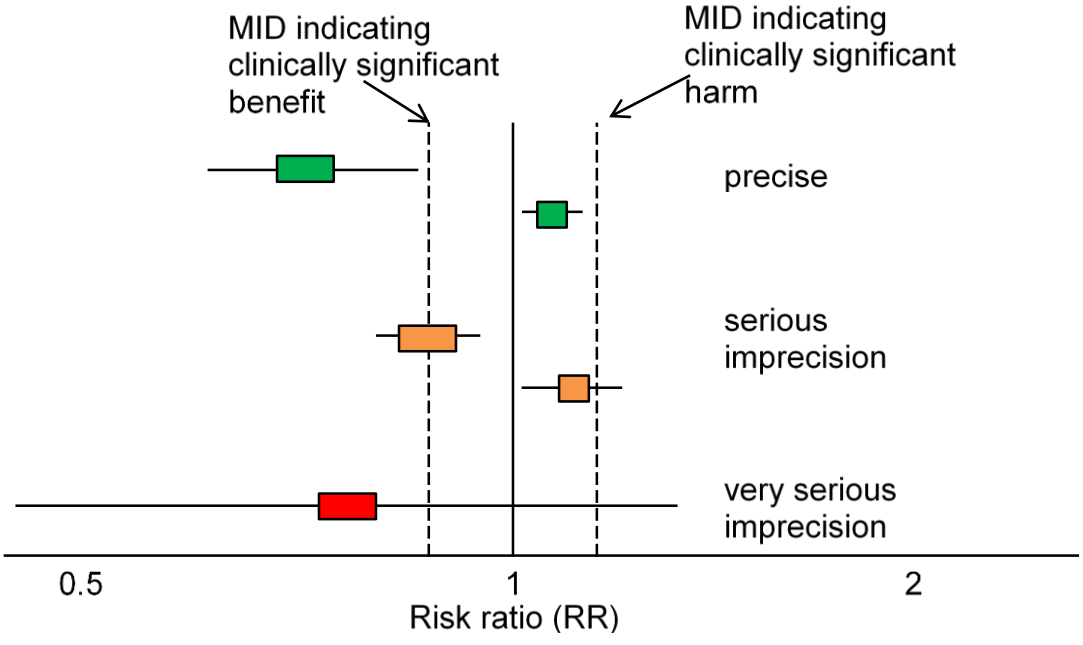
- For dichotomous outcomes the MIDs were taken to be RRs of 0.8 and 1.25. For 'positive' outcomes such as 'patient satisfaction', the RR of 0.8 is taken as the line denoting the boundary between no clinically important effect and a clinically important harm, whilst the RR of 1.25 is taken as the line denoting the boundary between no clinically important effect and a clinically important benefit. For 'negative' outcomes such as 'bleeding', the opposite occurs, so the RR of 0.8 is taken as the line denoting the boundary between no clinically important effect and a clinically important benefit, whilst the RR of 1.25 is taken as the line denoting the boundary between no clinically important effect and a clinically important harm. There are no 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 mortality any change was considered to be clinically important and the imprecision was assessed on the basis of the whether the confidence intervals crossed the line of no effect; that is whether the result was consistent with both benefit and harm.
- For continuous outcome variables the MID was taken as half the median baseline standard deviation of that variable, across all studies in the meta-analysis. Hence the MID denoting the minimum clinically important benefit was positive for a 'positive' outcome (for example, a quality of life measure where a higher score denotes better health), and negative for a 'negative' outcome (for example, a visual analogue scale [VAS] pain score). Clinically important harms will be the converse of these. If baseline values were unavailable, then half the median comparator group standard deviation of that variable was be taken as the MID. As these vary for each outcome per review, details of the values used are reported in the summary GRADE tables in the evidence report.
- If standardised mean differences were used, where the GC are able to specify a priority measure, the results would be back-converted to a mean difference on that scale for the assessment of imprecision and clinical importance. If it is not deemed appropriate to back-convert to a single scale, then the MID was set at the absolute value of +0.5. Standardised mean differences were not used within this review.

For this guideline, MIDs were found in the literature for the continuous health related quality of life outcome SF-36[2] which were used to assess imprecision and clinical importance (see section 2.3.2.8 below). The MIDs that had been specified by the NG59 committee were followed for the relevant outcomes. For the outcomes where an MID had not been agreed by the previous committee, the default values were used as described above for imprecision only, and clinical importance was determined by consideration of clinical importance based the point estimate, baseline values (for continuous outcomes), control event rate and absolute effect.

Figure 1: 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)



2.3.217 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. RCTs start at High, 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 are specified in the footnotes of the GRADE tables.

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

Table 4: Overall quality of outcome evidence in GRADE

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

2.3.2258 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 values used for imprecision and clinical importance are provided in Table 5.

The assessment of clinical benefit favouring intervention or comparator, or no benefit was based on the point estimate of absolute effect for intervention studies, which was standardised across the reviews. The committee used MIDs to determine clinical importance. Where there was no published MID in the literature (as discussed in section 2.3.2.6), the MIDs determined by consensus by the NG59 committee were adopted. The MIDs to assess clinical importance used in NG59 were based on an improvement of 10% as a measure of clinical benefit e.g. 1 point decrease on a 0-10 scale for pain severity. It was agreed that for the EQ-5D scale, a value of 0.03 should be used to be consistent with the published SF-36 values. The values used for imprecision and clinical importance are provided in Table 5.

Table 5: MIDs for assessing between group differences

Outcome	MID for imprecision	MID for clinical importance	Source
Pain measures including VAS & NRS (0-10 scale)	GRADE default	1	NG59
RMDQ (0-24 scale)	GRADE Default	2	NG59
ODI (0-100 scale)	GRADE Default	10	NG59

Outcome	MID for imprecision	MID for clinical importance	Source
SF-36 (0-100 scale)	Physical component summary: 2 Mental component summary: 3 Physical functioning: 3 Role-physical: 3 Bodily pain: 3 General health: 2 Vitality: 2 Social functioning: 3 Role-emotional: 4 Mental health: 3		User's manual for the SF-36v2 Health Survey, Third Edition[2]
EQ5D (0.0-1.0 scale)	Default	0.03	NG59
Other continuous outcomes	Default	10% of scale	NG59

VAS = visual analogue scale, NRS = numeric rating scale, RMDQ = Roland Morris Disability Questionnaire, ODI = Oswestry Disability Index

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

2.4 Identifying and analysing evidence of cost effectiveness

The committee is required to make decisions based on the best available evidence of both clinical effectiveness and cost effectiveness. Guideline recommendations should be based on the expected costs of the different options in relation to their expected health benefits (that is, their 'cost effectiveness') rather than the total implementation cost. However, the committee will also need to be increasingly confident in the cost effectiveness of a recommendation as the cost of implementation increases. Therefore, the committee may require more robust evidence on the effectiveness and cost effectiveness of any recommendations that are expected to have a substantial impact on resources; any 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.

2.4.1 Literature review

The health economists:

- Identified potentially relevant studies for the 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.

2.4.15 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 evidence report.

For more details about the assessment of applicability and methodological quality see Table 6 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 the evidence report.

2.4.16 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 the 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 6 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.[6]

Table 6: 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.

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

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

2.432 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.

2.443 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

The committee was presented with:

- Summaries of clinical and health economic evidence and quality (as presented in the evidence report).
- Evidence tables of the clinical and health economic evidence reviewed from the literature. All evidence tables can be found in the appendices to the evidence report.
- Forest plots (in the appendices to the evidence 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.

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

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

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

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

2.5.11 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 2-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.

2.6 Glossary

2.621 General terms methodological terms

3

Term	Definition
Abstract	Summary of a study, which may be published alone or as an introduction to a full scientific paper.
Algorithm (in guidelines)	A flow chart of the clinical decision pathway described in the guideline, where decision points are represented with boxes, linked with arrows.
Allocation concealment	The process used to prevent advance knowledge of group assignment in an RCT. The allocation process should be impervious to any influence by the individual making the allocation, by being administered by someone who is not responsible for recruiting participants.
Applicability	How well the results of a study or NICE evidence review can answer a clinical question or be applied to the population being considered.
Arm (of a clinical study)	Subsection of individuals within a study who receive one particular intervention, for example placebo arm.
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.
Bias	Influences on a study that can make the results look better or worse than they really are. (Bias can even make it look as if a treatment works when it does not.) Bias can occur by chance, deliberately or as a result of systematic errors in the design and execution of a study. It can also occur at different stages in the research process, for example, during the collection, analysis, interpretation, publication or review of research data. For examples see selection bias, performance bias, information bias, confounding factor, and publication bias.
Blinding	A way to prevent researchers, doctors and patients in a clinical trial from knowing which study group each patient is in so they cannot influence the results. The best way to do this is by sorting patients into study groups randomly. The purpose of 'blinding' or 'masking' is to protect against bias. A single-blinded study is one in which patients do not know which study group they are in (for example whether they are taking the experimental drug or a placebo). A double-blinded study is one in which neither patients nor the researchers and doctors know which study group the patients are in. A triple blind study is one in which neither the patients, clinicians or the people carrying out the statistical analysis know which treatment patients received.
Carer (caregiver)	Someone who looks after family, partners or friends in need of help because they are ill, frail or have a disability.
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.

Term	Definition
Cochrane Review	The Cochrane Library consists of a regularly updated collection of evidence-based medicine databases including the Cochrane Database of Systematic Reviews (reviews of randomised controlled trials prepared by the Cochrane Collaboration).
Cohort study	A study with 2 or more groups of people – cohorts – with similar characteristics. One group receives a treatment, is exposed to a risk factor or has a particular symptom and the other group does not. The study follows their progress over time and records what happens. See also observational study.
Comorbidity	A disease or condition that someone has in addition to the health problem being studied or treated.
Comparability	Similarity of the groups in characteristics likely to affect the study results (such as health status or age).
Confidence interval (CI)	A range of values for an unknown population parameter with a stated 'confidence' (conventionally 95%) that it contains the true value. The interval is calculated from sample data, and generally straddles the sample estimate. The 'confidence' value means that if the method used to calculate the interval is repeated many times, then that proportion of intervals will actually contain the true value.
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-effectiveness model	An explicit mathematical framework, which is used to represent clinical decision problems and incorporate evidence from a variety of sources in order to estimate the costs and health outcomes.
Cost–utility analysis (CUA)	Cost–utility analysis is one of the tools used to carry out an economic evaluation. The benefits are assessed in terms of both quality and duration of life, and expressed as quality-adjusted life years (QALYs). See also utility.
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

Term	Definition
	reflects individual preference for benefits to be experienced in the present rather than the future. Discounting costs reflects individual preference for costs to be experienced in the future rather than the present.
Disutility	The loss of quality of life associated with having a disease or condition. See Utility
Dominance	A health economics term. When comparing tests or treatments, an option that is both less effective and costs more is said to be 'dominated' by the alternative.
Drop-out	A participant who withdraws from a trial before the end.
Economic evaluation	An economic evaluation is used to assess the cost effectiveness of healthcare interventions (that is, to compare the costs and benefits of a healthcare intervention to assess whether it is worth doing). The aim of an economic evaluation is to maximise the level of benefits – health effects – relative to the resources available. It should be used to inform and support the decision-making process; it is not supposed to replace the judgement of healthcare professionals. There are several types of economic evaluation: cost–benefit analysis, cost–consequences analysis, cost-effectiveness analysis, cost-minimisation analysis and cost–utility analysis. They use similar methods to define and evaluate costs, but differ in the way they estimate the benefits of a particular drug, programme or intervention.
Effect (as in effect measure, treatment effect, estimate of effect, effect size)	A measure that shows the magnitude of the outcome in one group compared with that in a control group. For example, if the absolute risk reduction is shown to be 5% and it is the outcome of interest, the effect size is 5%. The effect size is usually tested, using statistics, to find out how likely it is that the effect is a result of the treatment and has not just happened by chance (that is, to see if it is statistically significant).
Effectiveness	How beneficial a test or treatment is under usual or everyday conditions, compared with doing nothing or opting for another type of care.
Efficacy	How beneficial a test, treatment or public health intervention is under ideal conditions (for example, in a laboratory), compared with doing nothing or opting for another type of care.
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.
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

Term	Definition
	system uses a common, sensible and transparent approach to grading the quality of evidence. The results of applying the GRADE system to clinical trial data are displayed in a table known as a GRADE profile.
Harms	Adverse effects of an intervention.
Health economics	Study or analysis of the cost of using and distributing healthcare resources.
Health-related quality of life (HRQoL)	A measure of the effects of an illness to see how it affects someone's day-to-day life.
Heterogeneity or Lack of homogeneity	The term is used in meta-analyses and systematic reviews to describe when the results of a test or treatment (or estimates of its effect) differ significantly in different studies. Such differences may occur as a result of differences in the populations studied, the outcome measures used or because of different definitions of the variables involved. It is the opposite of homogeneity.
Imprecision	Results are imprecise when studies include relatively few patients and few events and thus have wide confidence intervals around the estimate of effect.
Inclusion criteria (literature review)	Explicit criteria used to decide which studies should be considered as potential sources of evidence.
Incremental analysis	The analysis of additional costs and additional clinical outcomes with different interventions.
Incremental cost	The extra cost linked to using one test or treatment rather than another. Or the additional cost of doing a test or providing a treatment more frequently.
Incremental cost-effectiveness ratio (ICER)	The difference in the mean costs in the population of interest divided by the differences in the mean outcomes in the population of interest for one treatment compared with another.
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.
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.
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
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.
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

Term	Definition
	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	A measure of treatment effectiveness. The odds of an event happening in the treatment group, expressed as a proportion of the odds of it happening in the control group. The 'odds' is the ratio of events to non-events.
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, or more extreme results by chance. By convention, if the p value is below 0.05 (that is, there is less than a 5% probability that the results occurred by chance) it is considered that there probably is a real difference between treatments. If the p value is 0.001 or less (less than a 1% probability that the results occurred by chance), the result is seen as highly significant. If the p value shows that there is likely to be a difference between treatments, the confidence interval describes how big the difference in effect might be.
Placebo	A fake (or dummy) treatment given to participants in the control group of a clinical trial. It is indistinguishable from the actual treatment (which is given to participants in the experimental group). The aim is to determine what effect the experimental treatment has had – over and

Term	Definition
	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).
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.
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.
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'.

Term	Definition
Resource implication	The likely impact in terms of finance, workforce or other NHS resources.
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.
Stratification	When a different estimate effect is thought to underlie two or more groups based on the PICO characteristics. The groups are therefore kept separate from the outset and are not combined in a meta-analysis, for example; children and adults. Specified a priori in the protocol.

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

2.6.12 Guideline-specific terms

Term	Definition
Acute	Symptoms with a duration of less than 3 months.
Chronic	Duration of symptoms 3 months or more
Epidural injections	An injection into the epidural space within the spine, using either corticosteroids or anti-TNF agents for their anti-inflammatory and immunosuppressant properties.
Flare	A transient increase in back pain or sciatica symptoms
Lumbar disc prolapse/ herniation	Partial or complete protrusion of the nucleus pulposus through the annulus fibrosus of the intervertebral discs which can lead to compression of spinal nerve roots causing symptoms associated with sciatica.
Lumbosacral radiculopathy	Compression of nerve roots in the lower back causing symptoms including pain and numbness.
Neuropathic pain	Pain caused by a lesion or disease of the somatosensory nervous system.
Non-specific low back pain	Pain in the back between the bottom of the rib cage and the buttock creases.
Pharmacological interventions	Oral/sublingual, rectal, intra-muscular and transdermal drug treatments to relieve sciatica.
Sciatica	Irritation of the sciatica nerve causing pain and numbness in the body parts supplied by the nerve (lower back, hip, and outer side of the leg). May also be used to describe any pain starting in the lower back going down the leg.
Self-management	Programmes to assist people with low back pain and sciatica returning to normal activities. This includes education and advice for staying active.
Shared decision making	A collaborative process through which a healthcare professional supports a person to reach a decision about their care, now or in the future involving healthcare professionals working together with people who use services and their families and carers to choose treatments based on evidence and personal informed preferences, health beliefs and values.

Term	Definition
Spinal stenosis	Narrowing of the spinal canal causing compression of the spinal cord the can lead to persistent pain and numbness.

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: 21/04/2020

2. Maruish M. E. (Ed.). User's manual for the SF-36v2 Health Survey. 3rd ed. Lincoln, RI. QualityMetric Incorporated,.

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 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. London. National Institute for Health and Clinical Excellence, 2008. Available from: https://www.ncbi.nlm.nih.gov/books/NBK395865/

6. Organisation for Economic Co-operation and Development (OECD). Purchasing power parities (PPP). 2020. Available from: https://doi.org/10.1787/1290ee5a-en Last accessed: 21/04/2020

7. 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