National Institute for Health and Care Excellence

Draft for consultation

Joint replacement (primary): hip, knee and shoulder

NICE guideline: methods

NICE guideline

Methods

October 2019

Draft for Consultation

This evidence review was 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 All NICE guidance is subject to regular review and may be updated or withdrawn.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

Contents

1 Development of the guideline

1.1 What is a NICE guideline?

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

NICE guidelines can:

- provide recommendations for the treatment and care of people by health professionals
- be used to develop standards to assess the clinical practice of individual health professionals
- be used in the education and training of health professionals
- help patients to make informed decisions
- improve communication between patient and health 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: Primary hip, knee and shoulder joint replacement

ISBN

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 Adam Firth in accordance with guidance from NICE.

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

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

Staff from the NGC provided methodological support and guidance for the development process. The team working on the guideline included a project manager, systematic reviewers (research fellows), health economists and information specialists. They undertook systematic searches of the literature, appraised the evidence, conducted 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 covers adults referred for consideration of primary elective hip, knee or shoulder joint replacement. The clinical areas covered are:

- Information needs and shared decision-making.
- Preoperative rehabilitation.
- Anaesthesia.
- Tranexamic acid to reduce blood loss.
- Preventing infection through the use of wound irrigation and the type of operating room ventilation.
- Reducing incorrect implant selection errors.
- Aspects of surgery specific to hip, knee or shoulder joint replacement.
- Inpatient and outpatient postoperative rehabilitation.
- Long-term follow-up and surveillance.

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

1.3.2 What this guideline does not cover

This guideline does not cover the following groups:

- Children
- Adults having joint replacement as immediate treatment following fracture
- Adults having revision joint replacement
- Adults having joint replacement as treatment for primary or secondary cancer affecting the bones.

ISBN

This guideline does not cover the follow topics:

- Indications for referral for joint replacement.
- Diagnosis.
- Revision of joint replacement.

1.3.35 Relationships between the guideline and other NICE guidance

Related NICE technology appraisals:

- Autologous chondrocyte implantation for treating symptomatic articular cartilage defects of the knee (2017) NICE technology appraisal guidance TA477.
- Total hip replacement and resurfacing arthroplasty for end-stage arthritis of the hip (2014) NICE technology appraisal guidance TA304
- Apixaban for the prevention of venous thromboembolism after total hip or knee replacement in adults (2012) NICE technology appraisal guidance TA245.
- Rivaroxaban for the prevention of venous thromboembolism after total hip or total knee replacement in adults (2009) NICE technology appraisal guidance TA170.
- Dabigatran etexilate for the prevention of venous thromboembolism after hip or knee replacement surgery in adults (2008) NICE technology appraisal guidance TA157.

Related NICE interventional procedures guidance:

- Biodegradable subacromial spacer insertion for rotator cuff tears (2016) NICE interventional procedures guidance IPG558.
- Microstructural scaffold (patch) insertion without autologous cell implantation for repairing symptomatic chondral knee defects (2016) NICE interventional procedures guidance IPG560.
- Implantation of a shock or load absorber for mild to moderate symptomatic medial knee osteoarthritis (2015) NICE interventional procedures guidance IPG512.
- Platelet-rich plasma injections for osteoarthritis of the knee (2014) NICE interventional procedures guidance IPG491.
- Partial replacement of the meniscus of the knee using a biodegradable scaffold (2012) NICE interventional procedures guidance IPG430.
- Arthroscopic femoro–acetabular surgery for hip impingement syndrome (2011) NICE interventional procedures guidance IPG408.
- Mini-incision surgery for total knee replacement (2010) NICE interventional procedures guidance IPG345.
- Minimally invasive total hip replacement (2010) NICE interventional procedures guidance IPG363.
- Shoulder resurfacing arthroplasty (2010) NICE interventional procedures guidance IPG354.
- Individually magnetic resonance imaging-designed unicompartmental interpositional implant insertion for osteoarthritis of the knee (2009) NICE interventional procedures guidance IPG317.
- Arthroscopic knee washout, with or without debridement, for the treatment of osteoarthritis (2007) NICE interventional procedures guidance IPG230.
- Artificial trapeziometacarpal joint replacement for end-stage osteoarthritis (2005) NICE interventional procedures guidance IPG111.
- Artificial metacarpophalangeal and interphalangeal joint replacement for end-stage arthritis (2005) NICE interventional procedures guidance IPG110.

Related NICE guidelines:

- Blood transfusion (2015). NICE guideline NG24.
- Venous thromboembolism in over 16s: reducing the risk of hospital-acquired deep vein thrombosis or pulmonary embolism (2018) NICE guideline NG89.
- Hip fracture in adults (2017) NICE quality standard QS16.
- Spondyloarthritis in over 16s (2017) NICE guideline NG65.
- Osteoarthritis (2015) NICE quality standard QS87.
- Rheumatoid arthritis in adults: management (2018). NICE guideline NG100.
- The OSCAR 3 ultrasonic arthroplasty revision instrument for removing bone cement during prosthetic joint revision (2014) NICE medtech innovation briefing MIB13.
- Osteoarthritis: care and management (2014) NICE guideline CG177.
- Hip fracture: management (2011) NICE guideline CG124.
- The EOS 2D/3D imaging system (2011) NICE diagnostics guidance DG1.

Related NICE guidance currently in development:

- Perioperative care in adults. NICE guideline. Publication expected May 2020.

NICE guidance about the experience of people using NHS services

NICE has produced the following guidance on the experience of people using the NHS. This guideline will not include additional recommendations on these topics unless there are specific issues related to joint replacement:

- Patient experience in adult NHS services (2012) NICE guideline CG138

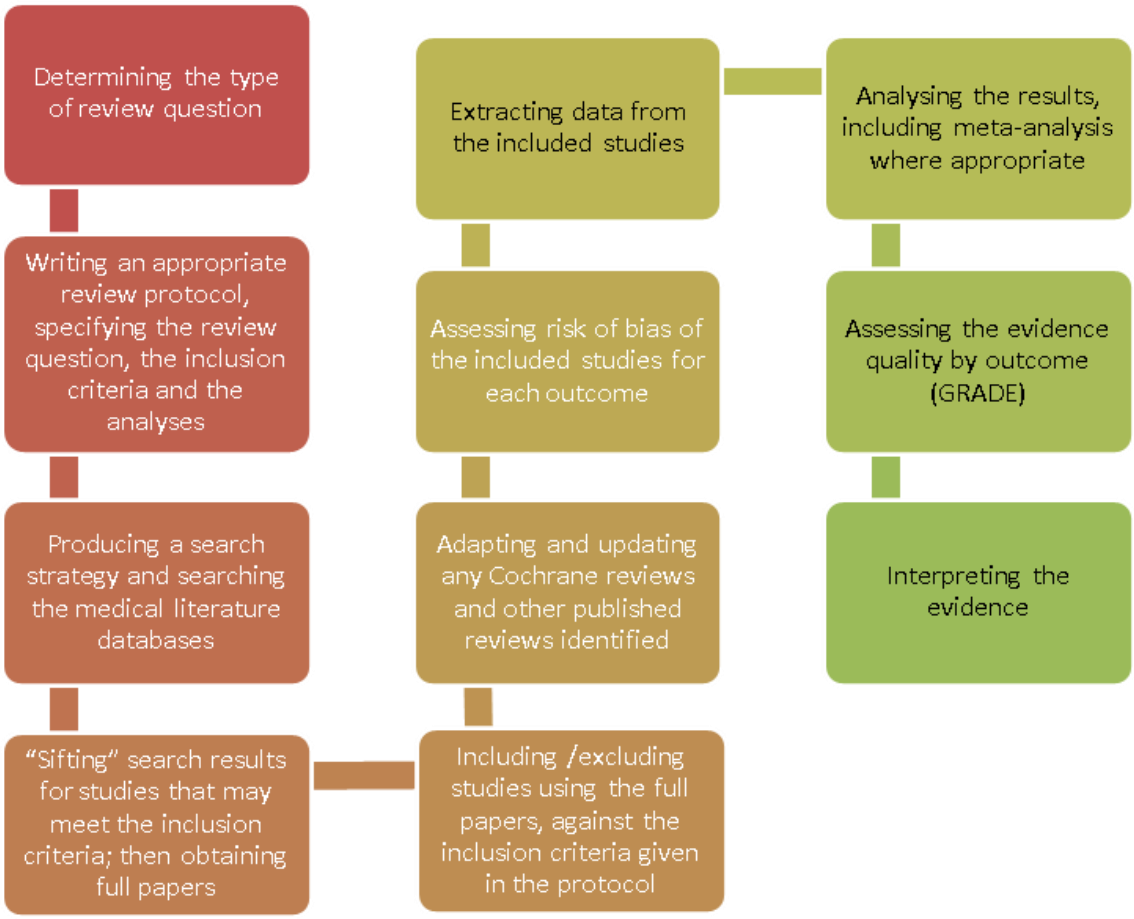
ISBN

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 which was updated in 2018.[2]

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

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



2.1 Developing the review questions and outcomes

Review questions were developed using a PICO framework (population, intervention, comparison and outcome) for intervention reviews; and using a framework of population, setting and context for qualitative reviews.

This use of a framework guided the literature searching process, critical appraisal and synthesis of evidence, and facilitated the development of recommendations by the guideline committee. The review questions were drafted by the NGC technical team and refined and validated by the committee. The questions were based on the key clinical areas identified in the scope.

A total of 17 review questions were identified.

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

Table 1: Review questions

Evidence report	Type of review	Review questions	Outcomes
A	Qualitative	What information would those having primary elective joint replacement surgery like to have prior to surgery?	• Synthesis of qualitative research with results presented in narrative format. Themes will be identified across studies.
B	Intervention and Qualitative	How useful are decision aids in helping people who are referred for primary elective joint replacement make decisions about their treatment (for example, the type of procedure, timing and implant choice)?	**Intervention review:** Critical outcomes • Quality of life • Patient Reported Outcome Measures (PROMs) • Patient-clinician communication • Participation in decision making • Accurate risk perceptions • Knowledge of the surgery • Decisional Conflict Scale • Satisfaction with care/decision-making Important outcomes • Proportion undecided • Adherence to chosen option **Qualitative review:** Themes will be identified across studies. Synthesis of qualitative research with results presented in narrative format
C	Intervention	Is preoperative rehabilitation clinically and cost effective for people having primary elective joint replacement?	Critical outcomes • Quality of life within 6 to 24 months for example EQ-5D, EQ-VAS • Patient Reported Outcome Measures (PROMs) within 6 to 24 months • Revision of joint replacement • Depression within 2 years • Disability within 6 to 24 months Important outcomes • Hospital readmissions: within 90 days • Muscle atrophy within 2 years • Length of stay

Evidence report	Type of review	Review questions	Outcomes
			To be extracted when not included within a PROM: • Function / ADL / return to work within 6 to 24 months • Pain within 2 years
D	Intervention	In adults having primary elective hip joint replacement, what is the clinical and cost effectiveness of intraoperative anaesthetic approaches: regional anaesthesia or general anaesthesia, with or without nerve blocks and local infiltration analgesia, compared with each other or in combination?	Critical outcomes • Mortality: within 90 days • Quality of life within 30 days • Postoperative pain within 30 days • Postoperative neurocognitive decline within 30 days • Thromboembolic complications within 90 days • Hospital readmission within 30 days Important outcomes • Postoperative use of analgesia • Length of stay • Nausea within 30 days • Mobilisation within 24 hours after surgery
E	Intervention	In adults having primary elective knee joint replacement, what is the clinical and cost effectiveness of regional anaesthesia or general anaesthesia, with or without nerve blocks and local infiltration analgesia, compared with each other or in combination?	Critical outcomes • Mortality: within 90 days • Quality of life within 30 days • Postoperative pain within 30 days • Postoperative neurocognitive decline within 30 days • Thromboembolic complications within 90 days • Hospital readmission within 30 days Important outcomes • Postoperative use of analgesia • Length of stay • Nausea within 30 days • Mobilisation within 24 hours after surgery
F	Intervention	In adults having primary elective shoulder joint replacement, what is the most clinical and cost effective intraoperative anaesthetic approach?	Critical outcomes • Mortality: within 90 days • Quality of life within 30 days • Postoperative pain within

Evidence report	Type of review	Review questions	Outcomes
			30 days • Hospital readmission within 30 days • Adverse events o Postoperative neurocognitive decline within 30 days o Thromboembolic complications within 90 days o Phrenic nerve injury within 90 days o Brachial plexus injury within 90 days Important outcomes • Postoperative use of analgesia • Length of stay • Nausea within 30 days • Mobilisation within 24 hours after surgery
G	Intervention	In adults having primary elective joint replacement, what is the clinical and cost effectiveness of tranexamic acid (TXA) for minimising blood loss from surgery?	Critical outcomes • Mortality: 30 day • Blood (allogeneic or autologous) transfusion • Adverse events o Acute myocardial infarction o Postoperative thrombosis • Quality of life within 6 weeks • Surgical bleeding Important outcomes • Postoperative anaemia • Postoperative bleeding • Length of stay
H	Intervention	In adults having primary elective joint replacement, what is the clinical and cost effectiveness of antibiotic or antiseptic wound lavage during the procedure?	Critical outcomes • Mortality 30 day • Quality of life • Superficial Surgical site infection • Deep surgical site infection Important outcomes • Return to theatre • Allergic reaction • Adverse antibiotic reactions • Hospital readmission • Pain • Length of stay

ISBN

Evidence report	Type of review	Review questions	Outcomes
I	Intervention	In adults having primary elective joint replacement or orthopaedic surgery utilising metallic implants, what is the clinical and cost effectiveness of using ultra clean-air theatres?	Critical outcomes • Mortality: 30 day • Quality of life • Deep surgical site infection • Superficial surgical site infection Important outcomes • Return to theatre • Hospital readmission • Length of stay
J	Intervention	What interventions would reduce the number of intraoperative implant selection errors, including systems and processes for selection, in adults having primary elective joint replacement?	Critical outcomes • Incorrect implant use • Revision rate • Revision surgery • Mortality: life expectancy • Mortality: 30 day • Quality of life Important outcomes • Hospital readmission • Length of stay • Enhanced follow up – recommend blood tests, cross sectional imaging
K	Intervention	In adults having primary elective knee replacement, what is the clinical and cost effectiveness of total knee replacement versus partial knee replacement?	Critical outcomes • Mortality: life expectancy • Mortality: 30 day • Quality of life at 6 weeks or earlier, later than 6 weeks up to 1 year, at least 2 years • Patient Reported Outcome Measures (PROMs) at 6 weeks or earlier, later than 6 weeks up to 1 year, at least 2 years • Major revision: revision of the tibia femoral compartment • Minor – polyethylene liner/polyethylene exchange Important outcomes • Deep surgical site infection • Superficial surgical site infection • Length of stay • Reoperation: excluding revision • Major adverse events as described by the studies: for example, VTE,

Evidence report	Type of review	Review questions	Outcomes
			myocardial infarction To be extracted when not included within a PROM: • Function at 6 weeks or earlier, later than 6 weeks up to 1 year, at least 2 years • Pain at 6 weeks or earlier, later than 6 weeks up to 1 year, at least 2 years
L	Intervention	In adults having primary elective knee replacement, what is the clinical and cost effectiveness of total knee replacement with patella resurfacing versus total knee replacement without patella resurfacing versus total knee replacement with selective resurfacing?	Critical outcomes • Mortality: life expectancy • Mortality: 30 day • Quality of life at 6 weeks or earlier, later than 6 weeks up to 1 year, at least 2 years • Patient Reported Outcome Measures (PROMs) at 6 weeks or earlier, later than 6 weeks up to 1 year, at least 2 years • Major revision: revision of the tibia femoral compartment • Minor – polyethylene liner/polyethylene exchange Important outcomes • Deep surgical site infection • Superficial surgical site infection • Length of stay • Reoperation (excluding revision) at 6 weeks or earlier, later than 6 weeks up to 1 year, at least 2 years • Major adverse events as described by the studies (For example, VTE, myocardial infarction)
M	Intervention	In adults having primary elective hip replacement, what is the most clinical and cost-effective approach: posterior, direct anterior, anterolateral, direct superior or SuperPATH?	Critical outcomes • Mortality: life expectancy • Mortality: 30 day • Quality of life at 6 weeks or earlier, later than 6 weeks up to 1 year, at least 2 years • Patient Reported Outcome Measures (PROMs) at 6 weeks or earlier, later than 6 weeks up to 1 year, at least 2 years

Evidence report	Type of review	Review questions	Outcomes
			• Revision rate of joint replacement Important outcomes • Deep surgical site infection • Superficial surgical site infection • Length of stay • Reoperation/dislocation rate • Intraoperative complications (for example nerve damage) • Surgery time To be extracted when not included within a PROM: • Function at 6 weeks or earlier, later than 6 weeks up to 1 year, at least 2 years • Pain at 6 weeks or earlier, later than 6 weeks up to 1 year, at least 2 years
N	Intervention	In adults having primary elective shoulder replacement for osteoarthritis with an intact rotator cuff, what is the clinical and cost effectiveness of humeral hemiarthroplasty versus conventional total shoulder arthroplasty versus reverse total shoulder arthroplasty?	Critical outcomes • Mortality: life expectancy • Mortality: 30 day • Quality of life at 6 weeks or earlier, later than 6 weeks up to 1 year, at least 2 years • Patient Reported Outcome Measures (PROMs) at 6 weeks or earlier, later than 6 weeks up to 1 year, at least 2 years • Revision of joint replacement • Reoperation at 6 weeks or earlier, later than 6 weeks up to 1 year, at least 2 years Important outcomes • Component failure • Dislocations within 1 year, after 1 year • Return to activity/sports • Deep surgical site Infection • Superficial surgical site infection • Length of stay • Major adverse events (including nerve injury, MI,

Evidence report	Type of review	Review questions	Outcomes
			VTE)
O	Intervention	In adults having primary elective shoulder replacement for pain and functional loss after a previous proximal humeral fracture (not acute trauma), what is the clinical and cost effectiveness of reverse total shoulder arthroplasty versus humeral hemiarthroplasty versus conventional total shoulder arthroplasty?	Critical outcomes • Mortality: life expectancy • Mortality: 30 day • Quality of life at 6 weeks or earlier, later than 6 weeks up to 1 year, at least 2 years • Patient Reported Outcome Measures (PROMs) at 6 weeks or earlier, later than 6 weeks up to 1 year, at least 2 years • Revision of joint replacement (time to event) • Reoperation Patient Reported Outcome Measures (PROMs) at 6 weeks or earlier, later than 6 weeks up to 1 year, at least 2 years Important outcomes • Component failure • Dislocations within 1 year, after 1 year • Return to activity/sports • Deep surgical site Infection • Superficial surgical site infection • Length of stay • Major adverse events (including nerve injury, MI, VTE)
P	Intervention	In adults who have undergone primary elective hip or knee replacement, what is the most clinical and cost-effective timing and duration for inpatient rehabilitation?	Critical outcomes • Quality of life within 6 weeks for example EQ-5D, EQ-VAS. • Patient Reported Outcome Measures (PROMs) within 6 weeks • Revision of joint replacement • Reoperation including dislocation within 6 weeks Important outcomes • Deep Surgical site infection within 6 weeks • Superficial surgical site infection within 6 weeks • Hospital readmissions: within 90 days

Evidence report	Type of review	Review questions	Outcomes
			• Thromboembolic events within 90 days • Length of stay To be extracted when not included within a PROM: • Function within 6 weeks • Pain within 6 weeks
Q	Intervention	In adults who have undergone primary elective shoulder replacement, what is the most clinical and cost-effective timing and duration for inpatient rehabilitation?	Critical outcomes • Quality of life within 6 weeks for example EQ-5D, EQ-VAS. • Patient Reported Outcome Measures (PROMs) within 6 weeks • Revision of joint replacement • Reoperation including dislocation within 6 weeks Important outcomes • Deep Surgical site infection within 6 weeks • Superficial surgical site infection within 6 weeks • Hospital readmissions: within 90 days • Thromboembolic events within 90 days • Length of stay To be extracted when not included within a PROM: • Function within 6 weeks Pain within 6 weeks
R	Intervention	In adults who have undergone primary elective hip or knee replacement, what is the clinical and cost effectiveness of self-directed outpatient rehabilitation versus supervised outpatient rehabilitation?	Critical outcomes • Quality of life within 6 weeks for example EQ-5D, EQ-VAS. • Patient Reported Outcome Measures (PROMs) within 6 weeks • Revision of joint replacement • Reoperation including dislocation within 6 weeks Important outcomes • Deep Surgical site infection within 6 weeks • Superficial surgical site infection within 6 weeks • Hospital readmissions: within 90 days • Thromboembolic events within 90 days

Evidence report	Type of review	Review questions	Outcomes
			• Length of stay To be extracted when not included within a PROM: • Function within 6 weeks Pain within 6 weeks
S	Intervention	In adults who have undergone primary elective shoulder replacement, what is the clinical and cost effectiveness of supervised outpatient rehabilitation versus self-directed outpatient rehabilitation?	Critical outcomes • Quality of life within 6 weeks for example EQ-5D, EQ-VAS. • Patient Reported Outcome Measures (PROMs) within 6 weeks • Revision of joint replacement • Reoperation including dislocation within 6 weeks Important outcomes • Deep Surgical site infection within 6 weeks • Superficial surgical site infection within 6 weeks • Hospital readmissions: within 90 days • Thromboembolic events within 90 days • Length of stay To be extracted when not included within a PROM: • Function within 6 weeks Pain within 6 weeks
T	Intervention	• In adults having primary elective joint replacement, what would be the optimal timing of follow-up or surveillance appointments? • In adults having primary elective joint replacement, who should carry out follow-up or surveillance appointments? • In adults having primary elective joint replacement, should x-rays be undertaken for all follow-up or surveillance appointments?	Critical outcomes • Emergency reoperation • Mortality: life expectancy • Quality of life at 2 years after surgery, at the longest time point (at least 4 years after surgery) • Patient Reported Outcome Measures (PROMs) at 2 years after surgery, at the longest time point • Reoperation (including revision) at 2 years after surgery, at the longest time point To be extracted when not included within a PROM: • Function at 2 years after surgery, at the longest time point • Pain at 2 years after surgery, at the longest time point

2.2 Searching for evidence

2.2.1 Clinical and health economics literature searches

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.[2] 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 01 May 2019. 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 a variety of approaches. Medline search strategies were checked by a second information specialist before being run. Searches were cross-checked with reference lists of highly relevant papers, searches in other systematic reviews were analysed, and committee members were requested to highlight additional studies.

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

- American Joint Replacement Registry (https://www.ajrr.net/)
- Australian Orthopaedic Association - National Joint Replacement Registry (https://aoanjrr.sahmri.com/)
- British Orthopaedic Association (https://www.boa.ac.uk/)
- California Joint Replacement Registry (http://staging.caljrr.org/)
- European Hip Society (http://www.europeanhipsociety.com/)
- European Society for Sports Traumatology, Knee Surgery and Arthroscopy (ESSKA) (https://www.esska.org/)
- Guidelines International Network database (www.g-i-n.net)
- National Guideline Clearing House (www.guideline.gov)
- National Institute for Health and Care Excellence (NICE) (www.nice.org.uk)
- National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) (https://www.niams.nih.gov/)
- National Joint Registry (http://www.njrcentre.org.uk/njrcentre/default.aspx)
- Royal College of Occupational Therapists (https://www.rcot.co.uk/)
- United States Bone and Joint Initiative (USBJI) (https://www.usbji.org/)

Searching for unpublished literature was not undertaken. The NGC and NICE do not have access to drug manufacturers' unpublished clinical trial results, so the clinical evidence considered by the committee for pharmaceutical interventions may be different from that considered by the MHRA and European Medicines Agency for the purposes of licensing and safety regulation.

Detailed search strategies can be found as an appendix to each evidence review.

ISBN

2.3 Identifying and analysing evidence of effectiveness

Research fellows conducted the tasks listed below, which are described in further detail in the rest of this section:

- Identified potentially relevant studies for each review question from the relevant search results by reviewing titles and abstracts. Full papers were then obtained.
- Reviewed full papers against prespecified inclusion and exclusion criteria to identify studies that addressed the review question in the appropriate population, and reported on outcomes of interest (review protocols are included in an appendix to each of the evidence reports).
- Critically appraised relevant studies using the appropriate study design checklist as specified in the NICE guidelines manual.[2] Qualitative studies were critically appraised using the GRADE CERQual approach for rating confidence in the body of evidence as a whole and using an NGC checklist for the methodological limitations section of the quality assessment.
- Extracted key information about interventional study methods and results using 'Evibase', NGC's purpose-built software. Evibase produces summary evidence tables, including critical appraisal ratings. Key information about non-interventional study methods and results was manually extracted onto standard evidence tables and critically appraised separately (evidence tables are included in an appendix to each of the evidence reports).
- Generated summaries of the evidence by outcome. Outcome data were combined, analysed and reported according to study design:
  - Randomised data were meta-analysed where appropriate and reported in GRADE profile tables.
  - Data from non-randomised studies were presented separately in GRADE profile tables or meta-analysed if appropriate.
  - Qualitative data were synthesised across studies and presented as summary statements with accompanying GRADE CERQual ratings for each review finding.
- A sample of a minimum of 10% of the abstract lists of the first 3 sifts by new reviewers and those for complex review questions (for example, prognostic reviews) were double-sifted by a senior research fellow and any discrepancies were rectified. All of the evidence reviews were quality assured by a senior research fellow. This included checking:
  - papers were included or excluded appropriately
  - a sample of the data extractions
  - correct methods were used to synthesise data
  - a sample of the risk of bias assessments.

2.3.1 Inclusion and exclusion criteria

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

The key population inclusion criterion was:

- Adults having primary elective knee, hip or shoulder replacement surgery, or two or more of these combined.

The key population exclusion criterion was:

- Adults having joint replacement as immediate treatment following fracture.

ISBN

- Adults having revision joint replacement.
- Adults having joint replacement as treatment for primary or secondary cancer affecting the bones.

Literature reviews, posters, letters, editorials, comment articles, unpublished studies and studies not in English were excluded.

2.3.1.17 Saturation of qualitative studies

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

2.3.2 Type of studies

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

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

Where data from non-randomised studies were included, the results for each outcome were presented separately for each study or meta-analysed if appropriate.

2.3.3 Methods of combining clinical studies

2.3.3.1 Data synthesis for intervention reviews

Where possible, meta-analyses were conducted using Cochrane Review Manager (RevMan5)[6] software to combine the data given in all studies for each of the outcomes of interest for the review question.

2.3.3.1.1 Analysis of different types of data

Dichotomous outcomes

Fixed-effects (Mantel–Haenszel) techniques (using an inverse variance method for pooling) were used to calculate risk ratios (relative risk, RR) for the binary outcomes, which included:

- mortality
- revision of joint replacement

- surgical site infection thromboembolic events

The absolute risk difference was also calculated using GRADEpro[1] software, using the median event rate in the control arm of the pooled results.

For binary variables where there were zero events in either arm or a less than 1% event rate, Peto odds ratios, rather than risk ratios, were calculated. Peto odds ratios are more appropriate for data with a low number of events.

Continuous outcomes

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

- heath-related quality of life (HRQoL)
- patient Reported Outcome Measures (PROMs)
- length of stay in hospital
- function
- pain

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

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)[6] software. Where p values were reported as 'less than', a conservative approach was undertaken. For example, if a p value was reported as 'p≤0.001', the calculations for standard deviations were based on a p value of 0.001. If these statistical measures were not available then the methods described in section 16.1.3 of the Cochrane Handbook (version 5.1.0, updated March 2018) were applied.

2.3.3.1.29 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. [6] If the control event rate was reported this was used to generate the absolute risk difference in GRADEpro.[1] If multivariate analysis was used to derive the summary statistic but no adjusted control event rate was reported no absolute risk difference was calculated.

2.3.3.1.35 Heterogeneity

Statistical heterogeneity was assessed for each meta-analysis estimate by considering the chi-squared test for significance at $p<0.1$ or an I-squared (I^2) inconsistency statistic (with an I-squared value of more than 50% indicating significant heterogeneity) as well as the distribution of effects. Where significant heterogeneity was present, predefined subgrouping of studies was carried as specified in each protocol. These subgrouping strategies were applied independently, so subunits of subgroups were not created, unlike the situation with strata.

If the subgroup analysis resolved heterogeneity within all of the derived subgroups, then each of the derived subgroups would have been adopted as separate outcomes (providing at least 1 study remained in each subgroup. Assessments of potential differences in effect between subgroups were based on the chi-squared tests for heterogeneity statistics between

subgroups. Any subgroup differences were interpreted with caution as separating the groups breaks the study randomisation and as such is subject to uncontrolled confounding.

If 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 could have been described narratively.

2.3.3.2 Network meta-analysis

Network meta-analysis was considered for the comparison of interventional treatments. In most cases it was not considered appropriate due to either insufficient data available in relevant outcomes or because there were too few interventions to elicit any benefit.

An original network meta-analysis (NMA) and cost comparison was conducted for the tranexamic acid review question by the heath economist in the technical team.

This type of analysis simultaneously compares multiple treatments in a single meta-analysis, preserving the randomisation of RCTs included in the reviews of direct comparisons trials. The aim of an NMA is to include all relevant evidence in order both to answer questions on the clinical effectiveness of interventions when no direct comparison was available and to give a ranking of treatments in terms of efficacy.

2.3.3.3 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. 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.

2.3.4 Appraising the quality of evidence by outcomes

2.3.4.1 Intervention reviews

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

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

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

Quality element	Description

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.

2.3.4.1.1 Risk of bias

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

Table 3: Principle domains of bias in randomised controlled trials

Limitation	Explanation
Selection bias (sequence generation and allocation concealment)	If those enrolling patients are aware of the group to which the next enrolled patient will be allocated, either because of a non-random sequence that is predictable, or because a truly random sequence was not concealed from the researcher, this may translate into systematic selection bias. This may occur if the researcher chooses not to recruit a participant into that specific group because of: • knowledge of that participant's likely prognostic characteristics, and • a desire for one group to do better than the other.
Performance and detection bias	Patients, caregivers, those adjudicating or recording outcomes, and data analysts should not be aware of the arm to which patients are allocated.

Limitation	Explanation
(lack of blinding of patients and healthcare professionals)	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.

The assessment of risk of bias differs for non-randomised intervention studies, as they are inherently at high risk of selection bias. For this reason, GRADE requires that non-randomised evidence is initially downgraded on the basis of study design, starting with a rating of −2. This accounts for selection bias and so non-randomised intervention studies are not downgraded any further on that domain. Non-randomised evidence was assessed against the remaining domains used for RCTs in Table 3, and downgraded further as appropriate.

2.3.4.1.2 Indirectness

Indirectness refers to the extent to which the populations, interventions, comparisons and outcome measures are dissimilar to those defined in the inclusion criteria for the reviews. Indirectness is important when these differences are expected to contribute to a difference in effect size, or may affect the balance of harms and benefits considered for an intervention. As for the risk of bias, each outcome had its indirectness assessed within each study first. For each study, if there were no sources of indirectness, indirectness was given a rating of 0. If there was indirectness in just 1 source (for example in terms of population), indirectness was given a 'serious' rating of −1, but if there was indirectness in 2 or more sources (for example, in terms of population and treatment) the indirectness was given a 'very serious' rating of −2. A weighted average score was then calculated across all studies contributing to the outcome by taking into account study precision. For example, if the most precise studies tended to have an indirectness score of −1 each for that outcome, the overall score for that outcome would tend towards −1.

2.3.4.1.3 Inconsistency

Inconsistency refers to an unexplained heterogeneity of results for an outcome across different studies. When estimates of the treatment effect across studies differ widely, this suggests true differences in the underlying treatment effect, which may be due to differences in populations, settings or doses. When heterogeneity existed within an outcome (chi-squared $p<0.1$, or $I^2>50\%$), but no plausible explanation could be found, the quality of evidence for that outcome was downgraded. Inconsistency for that outcome was given a

'serious' score of −1 if the I^2 was 50–74%, and a 'very serious' score of −2 if the I^2 was 75% or more.

If inconsistency could be explained based on prespecified subgroup analysis (that is, each subgroup had an I^2<50%), the committee took this into account and considered whether to make separate recommendations on new outcomes based on the subgroups defined by the assumed explanatory factors. In such a situation the quality of evidence was not downgraded for those emergent outcomes.

Since the inconsistency score was based on the meta-analysis results, the score represented the whole outcome and so weighted averaging across studies was not necessary.

2.3.4.1.4 Imprecision

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

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

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

- For categorical outcomes the MIDs were taken to be RRs of 0.8 and 1.25. For 'positive' outcomes such as 'patient satisfaction', the RR of 0.8 is taken as the line denoting the boundary between no clinically important effect and a clinically significant harm, whilst the RR of 1.25 is taken as the line denoting the boundary between no clinically important effect and a clinically significant benefit. For 'negative' outcomes such as 'bleeding', the opposite occurs, so the RR of 0.8 is taken as the line denoting the boundary between no clinically important effect and a clinically significant benefit, whilst the RR of 1.25 is taken as the line denoting the boundary between no clinically important effect and a clinically significant harm.
- For mortality any change was considered to be clinically important and the imprecision was assessed on the basis of the whether the confidence intervals crossed the line of no effect that is whether the result was consistent with both benefit and harm.

- For continuous outcome variables the MID was taken as half the median baseline standard deviation of that variable, across all studies in the meta-analysis. Hence the MID denoting the minimum clinically significant benefit was positive for a 'positive' outcome (for example, a quality of life measure where a higher score denotes better health), and negative for a 'negative' outcome (for example, a visual analogue scale [VAS] pain score). Clinically significant harms will be the converse of these. If baseline values are unavailable, then half the median comparator group standard deviation of that variable will be taken as the MID.
- If standardised mean differences have been used, then the MID will be set at the absolute value of +0.5. This follows because standardised mean differences are mean differences normalised to the pooled standard deviation of the 2 groups, and are thus effectively expressed in units of 'numbers of standard deviations'. The 0.5 MID value in this context therefore indicates half a standard deviation, the same definition of MID as used for non-standardised mean differences.

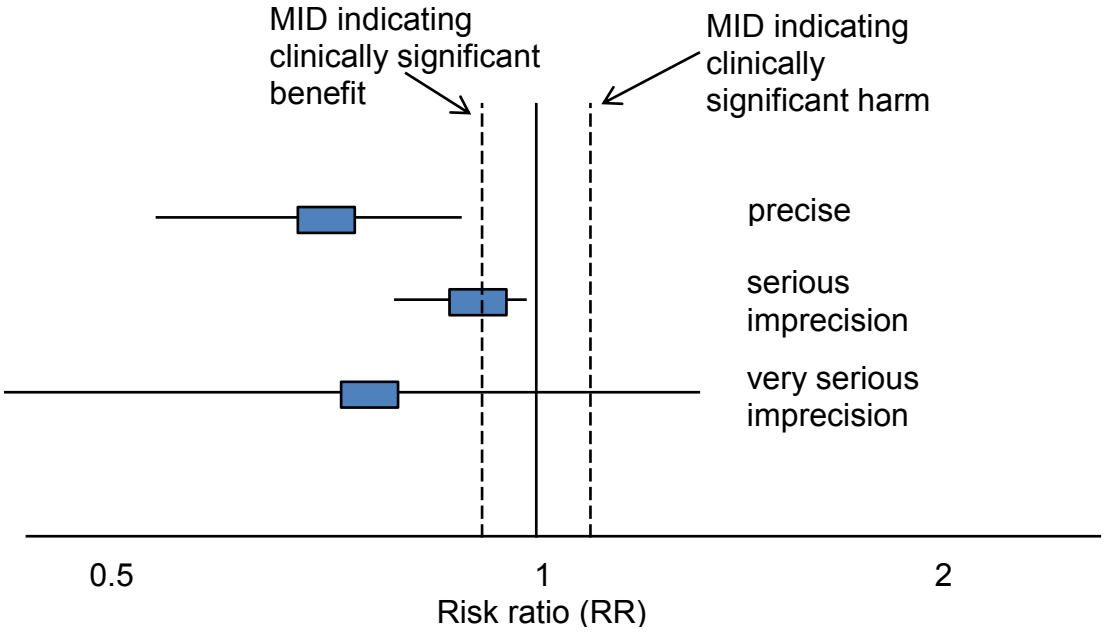
The default MID value was subject to amendment after discussion with the committee. If the committee decided that the MID level should be altered, after consideration of absolute as well as relative effects, this was allowed, provided that any such decision was not influenced by any bias towards making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

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

- SF-36 health survey[7]
  - mean difference of 2 for the Physical Component Summary (PCS)
  - mean difference of 3 for the Mental Component Summary (MCS)
- 10% difference on the Constant Murley scale was considered a clinically important difference. This was formulated through guideline committee consensus.

ISBN

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



*2.3.4.1.5*1 *Overall grading of the quality of clinical evidence*

2 Once an outcome had been appraised for the main quality elements, as above, an overall
3 quality grade was calculated for that outcome. The scores (0, −1 or −2) from each of the
4 main quality elements were summed to give a score that could be anything from 0 (the best
5 possible) to −8 (the worst possible). However scores were capped at −3. This final score was
6 then applied to the starting grade that had originally been applied to the outcome by default,
7 based on study design. All RCTs started as High and the overall quality became Moderate,
8 Low or Very Low if the overall score was −1, −2 or −3 points respectively. The significance of
9 these overall ratings is explained in Table 4. The reasons for downgrading in each case were
10 specified in the footnotes of the GRADE tables.

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

14 **Table 4: Overall quality of outcome evidence in GRADE**

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

2.3.4.21 Qualitative reviews

Review findings from the included qualitative studies were evaluated and presented using the 'Confidence in the Evidence from Reviews of Qualitative Research' (CERQual) Approach developed by the GRADE-CERQual Project Group, a subgroup of the GRADE Working Group.

The CERQual Approach assesses the extent to which a review finding is a reasonable representation of the phenomenon of interest (the focus of the review question). Each review finding was assessed for each of the 4 quality elements listed and defined below in Table 5.

Table 5: Description of quality elements in GRADE-CERQual for qualitative studies

Quality element	Description
Methodological limitations	The extent of problems in the design or conduct of the included studies that could decrease the confidence that the review finding is a reasonable representation of the phenomenon of interest. Assessed at the study level using an NGC checklist.
Coherence	The extent to which the reviewer is able to identify a clear pattern across the studies included in the review.
Relevance	The extent to which the body of evidence from the included studies is applicable to the context (study population, phenomenon of interest, setting) specified in the protocol.
Adequacy	The degree of the confidence that the review finding is being supported by sufficient data. This is an overall determination of the richness (depth of analysis) and quantity of the evidence supporting a review finding or theme.

Details of how the 4 quality elements (methodological limitations, coherence, relevance and adequacy) were appraised for each review finding are given below.

2.3.4.2.12 Methodological limitations

Each review finding had its methodological limitations assessed within each study first using an NGC checklist. Based on the degree of methodological limitations studies were evaluated as having minor, moderate or severe limitations. The questions to be answered in the checklist below included:

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

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

ISBN

2.3.4.2.2 Coherence

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

2.3.4.2.3 Relevance

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

2.3.4.2.4 Adequacy

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

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

GRADE-CERQual is used to assess the body of evidence as a whole through a confidence rating representing the extent to which a review finding is a reasonable representation of the phenomenon of interest. The 4 components (methodological limitations, coherence, relevance and adequacy) are used in combination to form an overall judgement. GRADE-CERQual uses 4 levels of confidence: high, moderate, low and very low confidence. The significance of these overall ratings is explained in Table 6. Each review finding starts at a high level of confidence and is downgraded based on the concerns identified in any 1 or more of the 4 components. Quality assessment of qualitative reviews is a subjective judgement by the reviewer based on the concerns that have been noted. A detailed explanation of how such a judgement had been made was included in the narrative summary.

Table 6: Overall level of confidence for a review finding in GRADE-CERQual

Level	Description
High confidence	It is highly likely that the review finding is a reasonable representation of the phenomenon of interest.
Moderate confidence	It is likely that the review finding is a reasonable representation of the phenomenon of interest.
Low confidence	It is possible that the review finding is a reasonable representation of the phenomenon of interest.
Very low confidence	It is not clear whether the review finding is a reasonable representation of the phenomenon of interest.

2.3.5 Assessing clinical importance

The committee assessed the evidence by outcome in order to determine if there was, or potentially was, a clinically important benefit, a clinically important harm or no clinically important difference between interventions. Where possible this was done in relation to published minimally important differences (MIDs).

Where MIDs were not available a collaborative method was used; combining the absolute effect and the relative effect with committee consensus. 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.

. For dichotomous outcomes the technical team made an initial judgement based on the relative effect utilising default MIDs. The committee utilised the reported absolute effect of the intervention with the relative effect and their experience and understanding of the outcome to make a conclusive assessment of the clinical importance.

For continuous outcomes if the mean difference was greater than the minimally important difference (MID) then this represented a clinical benefit or harm. The final judgement of clinical importance was combined with the committee's experience and understanding of the outcome.

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

2.3.6 Clinical evidence statements

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

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

2.4 Identifying and analysing evidence of cost effectiveness

The committee is required to make decisions based on the best available evidence of both clinical effectiveness and cost effectiveness. Guideline recommendations should be based on the expected costs of the different options in relation to their expected health benefits (that is, their 'cost effectiveness') rather than the total implementation cost. However, the committee will also need to be increasingly confident in the cost effectiveness of a recommendation as the cost of implementation increases. Therefore, the committee may require more robust evidence on the effectiveness and cost effectiveness of any recommendations that are expected to have a substantial impact on resources; any uncertainties must be offset by a compelling argument in favour of the recommendation. The cost impact or savings potential of a recommendation should not be the sole reason for the committee's decision.[2]

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.

2.4.15 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.[2]
- Extracted key information about the studies' methods and results into health economic evidence tables (which can be found in appendices to the relevant evidence reports).
- Generated summaries of the evidence in NICE health economic evidence profile tables (included in the relevant evidence report for each review question) – see below for details.

2.4.1.17 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 2003 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, 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 7 below and the economic evaluation checklist (appendix H of the NICE guidelines manual[2]) and the health economics review protocol, which can be found in each of the evidence reports.

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

2.4.1.20 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.[2] 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 7 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.[5]

Table 7: Content of NICE health economic evidence profile

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

(a) Applicability and limitations were assessed using the economic evaluation checklist in appendix H of the NICE guidelines manual[2]

2.4.20 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 tranexamic acid as the highest priority area for original health economic modelling. The rationale for this was that other areas that were originally noted as higher priority areas could not be modelled. Furthermore, there was uncertainty around the clinical and cost effectiveness of the different ways of administering tranexamic acid.

The following general principles were adhered to in developing the network meta-analysis and cost analysis:

- Methods were consistent with the NICE reference case for interventions with health outcomes in NHS settings.[2, 4]
- 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 network meta-analysis and cost analysis for tranexamic acid 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.[3] 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'.[3]

If a study reported the cost per life year gained but not QALYs, the cost per QALY gained was estimated by multiplying by an appropriate utility estimate to aid interpretation. The estimated cost per QALY gained is reported in the health economic evidence profile with a footnote detailing the life-years gained and the utility value used. 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 reviews [A–B]).
- Evidence tables of the clinical and health economic evidence reviewed from the literature. All evidence tables can be found in appendices to the relevant evidence reports.
- Forest plots (in appendices to the relevant evidence reviews).
- A description of the methods and results of the health economic analysis undertaken for the guideline (in a separate economic analysis report).

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

When clinical and health economic evidence was of poor quality, conflicting or absent, the committee drafted recommendations based on its expert opinion. The considerations for making consensus-based recommendations include the balance between potential harms and benefits, the economic costs compared to the economic benefits, current practices, recommendations made in other relevant guidelines, patient preferences and equality issues. The consensus recommendations were agreed through discussions in the committee. 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[2]).

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

2.5.1 Research recommendations

When areas were identified for which good evidence was lacking, the committee considered making recommendations for future research. Decisions about the inclusion of a research recommendation were based on factors such as:

- the importance to patients or the population
- national priorities
- potential impact on the NHS and future NICE guidance
- ethical and technical feasibility.

2.5.2 Validation process

This guidance is subject to a 6-week public consultation and feedback as part of the quality assurance and peer review of the document. All comments received from registered stakeholders are responded to in turn and posted on the NICE website.

2.5.3 Updating the guideline

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

2.5.4 Disclaimer

Healthcare providers need to use clinical judgement, knowledge and expertise when deciding whether it is appropriate to apply guidelines. The recommendations cited here are a guide and may not be appropriate for use in all situations. The decision to adopt any of the recommendations cited here must be made by practitioners in light of individual patient circumstances, the wishes of the patient, clinical expertise and resources.

The National Guideline Centre disclaims any responsibility for damages arising out of the use or non-use of this guideline and the literature used in support of this guideline.

2.5.5 Funding

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

31 Acronyms and abbreviations

Acronym or abbreviation	Description
ACB	Adductor canal block
ADL	Activities of daily living
AKSS	American Knee Society Score
ASA	American Society of Anesthesiologists
BKS	Bristol Knee Score
BNF	British National Formulary
BOA	British Orthopaedic Association
CEA	Cost-effectiveness analysis
CI	Confidence interval
CUA	Cost–utility analysis
DVT	Deep vein thrombosis
FN	False negative
FNB	Femoral nerve block
FP	False positive
GC	Guideline Committee
GIRFT	Getting It Right First Time
GRADE	Grading of recommendations assessment, development and evaluation
HAAS	High Activity Arthroplasty Score
HEPA	High Efficiency Particulate Air
HES	Hospital Episode Statistics
HRG	Healthcare resource group
HSS score	Hospital for Special Surgery score
IA	Intra-articular
ICER	Incremental cost-effectiveness ratio
ISB	Interscalene brachial plexus block
IU	International unit
IV	Intravenous
KOS	Knee Outcome Scale
KOOS	Knee injury and Osteoarthritis Outcome Score
KSS	Knee Society Score
LIA	Local infiltration analgesia
MCS	Mental component score
NGC	National Guideline Centre
NICE	National Institute of Health and Care Excellence
NIHR	The National Institute for Health Research
NJR	National Joint Registry
NPV	Negative predictive value
NRS	Numerical rating scale
OA	Osteoarthritis
ODEP	Orthopaedic Data Evaluation Panel
OECD	Organisation for Economic Co-operation and Development
OKS	Oxford Knee Score

ISBN

Acronym or abbreviation	Description
PCS	Physical component score
PE	Pulmonary embolism
PHE	Public Health England
PKA	Partial knee arthroplasty
PKR	Partial knee replacement
PO	Per os (by mouth, orally)
PPV	Positive predictive value
PROM	Patient reported outcome measure
QALY	Quality-adjusted life year
QOL	Quality of life
RA	Rheumatoid arthritis
RCT	Randomised control trial
ROC	Receiver operating characteristic
RR	Relative risk
RSA	Reverse total shoulder arthroplasty
SF-12	12-Item Short Form Health Survey
SF-36	36-Item Short Form Health Survey
SNB	Sciatic nerve block
SSI	Surgical site infection
SuperPATH approach	Supercapsular Percutaneously Assisted Total Hip approach
THA	Total hip arthroplasty
THR	Total hip replacement
TKA	Total knee arthroplasty
TKR	Total knee replacement
TN	True negative
TP	True positive
TSA	Total shoulder arthroplasty
TSR	Total shoulder replacement
TXA	Tranexamic acid
UCA	Ultra clean-air
UKA	Unicompartmental knee arthroplasty
UKR	Unicompartmental knee replacement
VAS	Visual Analogue Scale
WBC	White blood cell
WC	Whole cell
WOMAC	The Western Ontario and McMaster Universities Arthritis Index

1

4 Glossary

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

4.1 Guideline-specific terms

Term	Definition
Arthroplasty	The surgical reconstruction or replacement of a joint.
Arthrotomy	Surgical exploration of a joint, which should include inspection of the cartilage, intra-articular structures, joint capsule, and ligaments.
ASA Physical Status Classification System	A system for assessing the fitness of patients before surgery
Bilateral joint replacement	Both of a person's hips, knees or shoulders will be replaced during the same surgery.
Cemented joint prosthesis/implant	Polymer bone cement used to affix a joint prosthesis to bone
Conventional total shoulder replacement	A conventional shoulder replacement mimics the normal anatomy of the shoulder. The surgery involves replacing the ball part of the joint with a new ball and the socket part of the joint with anew socket.
Deep surgical site infection	Infection beneath the fascial layers involving the muscle, joint and bone. These infections can present with pus or an abscess, fever with tenderness of the wound, or a separation of the edges of the incision exposing the deeper tissues.
Elective surgery	Surgery that is planned, not urgent and subject to choice.
General anaesthesia	A combination of medications that put a person in a sleep-like state before a surgery or other medical procedure.
Humeral Hemiarthroplasty	Replacement of the humeral head with a humeral prosthesis
Joint replacement implant	Artificial components of a joint replacement
Joint school	Joint school is an educational session where a series of presentations are given by members of the orthopaedic team on the specific stages of care. This could include the pre-assessment nurse, surgeon, anaesthetist, physiotherapist, occupational therapist and pharmacist.
Inflammatory arthritis	Inflammatory arthritis are a group of systemic diseases characterised by inflammation of the joints and often other tissues. These include rheumatoid arthritis, psoriatic arthritis, ankylosing spondylitis, juvenile idiopathic arthritis and systemic lupus erythematosus (lupus) among others.
Joint replacement	The surgical replacement of a joint.
Laminar flow	A continuous flow of highly filtered ultra clean-air (CA) of less than 10 colony-forming units per metre cubed of bacteria.
Local infiltration analgesia	Use of local anaesthetic in the tissues around the surgical field to provide pain relief
Nerve block	Use of a local anaesthetic injection to provide pain relief along a specific distribution of nerve or nerves.
Partial knee replacement/arthroplasty	Only a portion of the knee is replaced. Either the medial or lateral tibiofemoral compartments.
Patella resurfacing	Conducted as part of a total knee replacement. A separate patella implant is attached to the back of the kneecap to articulate and fit smoothly with the femoral implant
Preoperative rehabilitation	Preoperative preparation leading up to the operation that can include physical therapy, occupational therapy, nutritional counselling, acupuncture, transcutaneous electrical nerve stimulation, hydrotherapy or education interventions

ISBN

Term	Definition
Primary joint replacement/arthroplasty	Primary joint replacement is the first joint replacement operation on a particular joint.
Prosthesis	An artificial component used in joint replacement surgery
Regional anaesthesia	Regional anaesthesia makes a specific part of the body numb to relieve pain or allow surgical procedures to be done. Types of regional anaesthesia include spinal anaesthesia, epidural anaesthesia, and nerve blocks.
Reverse total shoulder replacement/arthroplasty	A reverse total shoulder replacement does not mimic the normal anatomy of the shoulder. It involves replacing the ball part of the joint with a socket and replacing the socket part with a ball. . The relationship is therefore reversed.
Total blood loss after joint replacement surgery	The amount of blood-loss in people who have undergone joint replacement surgery
Total knee replacement/arthroplasty	Surgery involves the replacement of both sides of the knee joint
Total hip replacement/arthroplasty	Surgery involves replacing the ball and the socket parts of the hop joint with artificial parts
Tourniquet	Compressive devices that occlude venous and arterial blood flow to limbs
Ultra clean-air ventilation	A controlled column of clean air is delivered down and over the operating table, surgical team and equipment
Unicompartmental knee replacement/arthroplasty	Only a portion of the knee is replaced. Either the medial or lateral tibiofemoral compartments.
Unilateral joint replacement	One joint is replaced as opposed to both being done as in bilateral joint replacement (which is rare)
Wound lavage	Wound lavage is the intermittent irrigation of an open wound with solution to remove deeper debris, prevent infection, and to assist with the visual examination.

1

4.22 General terms

Term	Definition
Abstract	Summary of a study, which may be published alone or as an introduction to a full scientific paper.
Algorithm (in guidelines)	A flow chart of the clinical decision pathway described in the guideline, where decision points are represented with boxes, linked with arrows.
Allocation concealment	The process used to prevent advance knowledge of group assignment in an RCT. The allocation process should be impervious to any influence by the individual making the allocation, by being administered by someone who is not responsible for recruiting participants.
Applicability	How well the results of a study or NICE evidence review can answer a clinical question or be applied to the population being considered.
Arm (of a clinical study)	Subsection of individuals within a study who receive one particular intervention, for example placebo arm.
Association	Statistical relationship between 2 or more events, characteristics or other variables. The relationship may or may not be causal.
Base case analysis	In an economic evaluation, this is the main analysis based on the most plausible estimate of each input. In contrast, see Sensitivity analysis.
Baseline	The initial set of measurements at the beginning of a study (after run-

Term	Definition
	in period where applicable), with which subsequent results are compared.
Bayesian analysis	A method of statistics, where a statistic is estimated by combining established information or belief (the 'prior') with new evidence (the 'likelihood') to give a revised estimate (the 'posterior').
Before-and-after study	A study that investigates the effects of an intervention by measuring particular characteristics of a population both before and after taking the intervention, and assessing any change that occurs.
Bias	Influences on a study that can make the results look better or worse than they really are. (Bias can even make it look as if a treatment works when it does not.) Bias can occur by chance, deliberately or as a result of systematic errors in the design and execution of a study. It can also occur at different stages in the research process, for example, during the collection, analysis, interpretation, publication or review of research data. For examples see selection bias, performance bias, information bias, confounding factor, and publication bias.
Blinding	A way to prevent researchers, doctors and patients in a clinical trial from knowing which study group each patient is in so they cannot 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.

ISBN

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).
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.

ISBN

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

Term	Definition
	to replace the judgement of healthcare professionals. There are several types of economic evaluation: cost–benefit analysis, cost–consequences analysis, cost-effectiveness analysis, cost-minimisation analysis and cost–utility analysis. They use similar methods to define and evaluate costs, but differ in the way they estimate the benefits of a particular drug, programme or intervention.
Effect (as in effect measure, treatment effect, estimate of effect, effect size)	A measure that shows the magnitude of the outcome in one group compared with that in a control group. For example, if the absolute risk reduction is shown to be 5% and it is the outcome of interest, the effect size is 5%. The effect size is usually tested, using statistics, to find out how likely it is that the effect is a result of the treatment and has not just happened by chance (that is, to see if it is statistically significant).
Effectiveness	How beneficial a test or treatment is under usual or everyday conditions, compared with doing nothing or opting for another type of care.
Efficacy	How beneficial a test, treatment or public health intervention is under ideal conditions (for example, in a laboratory), compared with doing nothing or opting for another type of care.
Epidemiological study	The study of a disease within a population, defining its incidence and prevalence and examining the roles of external influences (for example, infection, diet) and interventions.
EQ-5D (EuroQol 5 dimensions)	A standardised instrument used to measure health-related quality of life. It provides a single index value for health status.
Evidence	Information on which a decision or guidance is based. Evidence is obtained from a range of sources including randomised controlled trials, observational studies, expert opinion (of clinical professionals or patients).
Exclusion criteria (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.
Gold standard	A method, procedure or measurement that is widely accepted as being the best available to test for or treat a disease.
GRADE, GRADE profile	A system developed by the GRADE Working Group to address the shortcomings of present grading systems in healthcare. The GRADE system uses a common, sensible and transparent approach to grading the quality of evidence. The results of applying the GRADE system to clinical trial data are displayed in a table known as a GRADE profile.
Harms	Adverse effects of an intervention.

Term	Definition
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.
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.
Intraoperative	The period of time during a surgical procedure.
Kappa statistic	A statistical measure of inter-rater agreement that takes into account the agreement occurring by chance.
Length of stay	The total number of days a participant stays in hospital.
Licence	See 'Product licence'.
Life years gained	Mean average years of life gained per person as a result of the intervention compared with an alternative intervention.
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	In statistics, logistic regression is a type of analysis used for

Term	Definition
Logit model	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
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.

Term	Definition
	Sometimes probability can be compared across more than 2 groups – in this case, one of the groups is chosen as the 'reference category', and the odds ratio is calculated for each group compared with the reference category. For example, to compare the risk of dying from lung cancer for non-smokers, occasional smokers and regular smokers, non-smokers could be used as the reference category. Odds ratios would be worked out for occasional smokers compared with non-smokers and for regular smokers compared with non-smokers. See also confidence interval, risk ratio.
Opportunity cost	The loss of other healthcare programmes displaced by investment in or introduction of another intervention. This may be best measured by the health benefits that could have been achieved had the money been spent on the next best alternative healthcare intervention.
Outcome	The impact that a test, treatment, policy, programme or other intervention has on a person, group or population. Outcomes from interventions to improve the public's health could include changes in knowledge and behaviour related to health, societal changes (for example, a reduction in crime rates) and a change in people's health and wellbeing or health status. In clinical terms, outcomes could include the number of patients who fully recover from an illness or the number of hospital admissions, and an improvement or deterioration in someone's health, functional ability, symptoms or situation. Researchers should decide what outcomes to measure before a study begins.
P value	The p value is a statistical measure that indicates whether or not an effect is statistically significant. For example, if a study comparing 2 treatments found that one seems more effective than the other, the p value is the probability of obtaining these results by chance. By convention, if the p value is below 0.05 (that is, there is less than a 5% probability that the results occurred by chance) it is considered that there probably is a real difference between treatments. If the p value is 0.001 or less (less than a 1% probability that the results occurred by chance), the result is seen as highly significant. If the p value shows that there is likely to be a difference between treatments, the confidence interval describes how big the difference in effect might be.
Perioperative	The period from admission through surgery until discharge, encompassing the preoperative and postoperative periods.
Placebo	A fake (or dummy) treatment given to participants in the control group of a clinical trial. It is indistinguishable from the actual treatment (which is given to participants in the experimental group). The aim is to determine what effect the experimental treatment has had – over and above any placebo effect caused because someone has received (or thinks they have received) care or attention.
Polypharmacy	The use or prescription of multiple medications.
Posterior distribution	In Bayesian statistics this is the probability distribution for a statistic based after combining established information or belief (the prior) with new evidence (the likelihood).
Postoperative	Pertaining to the period after patients leave the operating theatre, following surgery.
Post-test probability	In diagnostic tests: The proportion of patients with that particular test result who have the target disorder (post-test odds/[1 plus post-test odds]).
Power (statistical)	The ability to demonstrate an association when one exists. Power is related to sample size; the larger the sample size, the greater the

Term	Definition
	power and the lower the risk that a possible association could be missed.
Preoperative	The period before surgery commences.
Pre-test probability	In diagnostic tests: The proportion of people with the target disorder in the population at risk at a specific time point or time interval. Prevalence may depend on how a disorder is diagnosed.
Prevalence	See Pre-test probability.
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.

ISBN

Term	Definition
RCT	See ‘Randomised controlled trial’.
Receiver operated characteristic (ROC) curve	A graphical method of assessing the accuracy of a diagnostic test. Sensitivity is plotted against 1 minus specificity. A perfect test will have a positive, vertical linear slope starting at the origin. A good test will be somewhere close to this ideal.
Reference standard	The test that is considered to be the best available method to establish the presence or absence of the outcome – this may not be the one that is routinely used in practice.
Reporting bias	See ‘Publication bias’.
Resource implication	The likely impact in terms of finance, workforce or other NHS resources.
Retrospective study	A research study that focuses on the past and present. The study examines past exposure to suspected risk factors for the disease or condition. Unlike prospective studies, it does not cover events that occur after the study group is selected.
Review question	In guideline development, this term refers to the questions about treatment and care that are formulated to guide the development of evidence-based recommendations.
Risk ratio (RR)	The ratio of the risk of disease or death among those exposed to certain conditions compared with the risk for those who are not exposed to the same conditions (for example, the risk of people who smoke getting lung cancer compared with the risk for people who do not smoke). If both groups face the same level of risk, the risk ratio is 1. If the first group had a risk ratio of 2, subjects in that group would be twice as likely to have the event happen. A risk ratio of less than 1 means the outcome is less likely in the first group. The risk ratio is sometimes referred to as relative risk.
Secondary outcome	An outcome used to evaluate additional effects of the intervention deemed a priori as being less important than the primary outcomes.
Selection bias	Selection bias occurs if: a) The characteristics of the people selected for a study differ from the wider population from which they have been drawn, or b) There are differences between groups of participants in a study in terms of how likely they are to get better.
Sensitivity 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 ($p<0.05$).

ISBN

Term	Definition
Stakeholder	An organisation with an interest in a topic that NICE is developing a guideline or piece of public health guidance on. Organisations that register as stakeholders can comment on the draft scope and the draft guidance. Stakeholders may be: • manufacturers of drugs or equipment • national patient and carer organisations • NHS organisations • organisations representing healthcare professionals.
State transition model	See Markov model
Systematic review	A review in which evidence from scientific studies has been identified, appraised and synthesised in a methodical way according to predetermined criteria. It may include a meta-analysis.
Time horizon	The time span over which costs and health outcomes are considered in a decision analysis or economic evaluation.
Transition probability	In a state transition model (Markov model), this is the probability of moving from one health state to another over a specific period of time.
Treatment allocation	Assigning a participant to a particular arm of a trial.
Univariate	Analysis 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

ISBN

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: 02/08/2019

2. 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

3. National Institute for Health and Clinical Excellence. Social value judgements: principles for the development of NICE guidance. London. National Institute for Health and Clinical Excellence, 2008. Available from: https://www.nice.org.uk/media/default/about/what-we-do/research-and-development/social-value-judgements-principles-for-the-development-of-nice-guidance.pdf

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

5. Organisation for Economic Co-operation and Development (OECD). Purchasing power parities (PPP). 2012. Available from: http://www.oecd.org/std/ppp Last accessed: 17/06/2019

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

7. Ware J, Kosinski M, Bjorner J, Turner-Bowker D, Gandek B, Maruish M. User's Manual for the SF-36v2® Health Survey. Lincoln (RI). QualityMetric Incorporated, 2007. Available from: https://eprovide.mapi-trust.org/instruments/sf-36-health-survey

ISBN