National Institute for Health and Care Excellence

Final

Type 2 diabetes in adults: management (Medicines update)

NICE guideline: methods

NICE guideline

Methods

February 2026

Final

Developed by NICE

Disclaimer

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

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

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 All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

ISBN: 978-1-4731-9273-7

Contents

1 Development of the guideline

1.1 Remit

NICE received the remit for this guideline from NHS England.

The remit for this guideline is: to update the area of drug treatment in the NICE guideline Type 2 diabetes in adults: management (NG28).

1.2 What this guideline covers

The methods outlined in this document relate to drug treatment for Type 2 diabetes in adults: management. This guideline update will include looking at the evidence for and consider making new recommendations or updating existing recommendations on:

- Drug treatment including a refresh of the type 2 diabetes insulin-based treatment recommendations (recommendations 1.7.24 - 1.7.32).

1.3 What this guideline does not cover

This guideline update does not include the following sections:

- Individualised care
- Education
- Dietary advice and bariatric surgery
- Diagnosing and managing hypertension
- Antiplatelet therapy
- Blood glucose management
- Managing complications

The evidence for these sections relates to the methods from previous versions of the guideline. For details of the methods used for these reviews please refer to the following sources:

- For 2022 recommendations developed in accordance with the methods outlined in the NICE Guidelines Manual 2020: Appendix B of the evidence review document.
- For 2015 recommendations developed in accordance with the methods outlined in the NICE Guidelines Manual 2012: Chapter 3 of the full guideline document

2 Methods

This guideline was developed using the methods described in the NICE guidelines manual update 2022.

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

2.1 Developing the review questions and outcomes

The review questions developed for this guideline were based on the key areas and draft review questions identified in the guideline scope. It was drafted by the technical team, refined and validated by the guideline committee and signed off by NICE. A total of 2 review questions were developed in this guideline and outlined in Table 1.

The review questions were based on the following frameworks:

- population, intervention, comparator and outcome (PICO) for reviews of interventions

This use of a framework informed a more detailed protocol that guided the literature searching process, critical appraisal and synthesis of evidence, and facilitated the development of recommendations by the guideline committee. 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
E Initial therapy	Intervention	For different population subgroups, which individual and/or combinations of pharmacological therapies are most clinically and cost effective as initial treatment for the management of type 2 diabetes?	• Health-related quality of life • All-cause mortality • Cardiovascular mortality • Major cardiovascular events (MACE) • Non-fatal stroke • Non-fatal myocardial infarction • Unstable angina • Hospitalisation for heart failure • Acute kidney injury • Persistent signs of worsening kidney disease • Development of end-stage kidney disease • Death from renal causes • Cardiac arrhythmia • Diabetic ketoacidosis • Falls requiring hospitalisation • Progression of liver disease • Remission • Hypoglycaemia episodes • At night hypoglycaemic episodes • Severe hypoglycaemic episodes

Evidence report	Type of review	Review questions	Outcomes
			• HbA1c change • Weight change • BMI change
F Subsequent therapy	Intervention	Which pharmacological therapies are most clinically and cost effective for the management of type 2 diabetes when current treatment has not given adequate response, including: • medicines within the following classes biguanides, DPP-4 inhibitors, GLP-1 receptor agonist, insulin, sulfonylureas, SGLT2 inhibitors, and thiazolidinediones (but not limited to these), • approaches to optimise treatment (including combination treatment, switching to different therapies, de-escalation and stopping previous therapies), and • consideration of different population subgroups?	• Health-related quality of life • All-cause mortality • Cardiovascular mortality • Major cardiovascular events (MACE) • Non-fatal stroke • Non-fatal myocardial infarction • Unstable angina • Hospitalisation for heart failure • Acute kidney injury • Persistent signs of worsening kidney disease • Development of end-stage kidney disease • Death from renal causes • Cardiac arrhythmia • Diabetic ketoacidosis • Falls requiring hospitalisation • Progression of liver disease • Remission • Hypoglycaemia episodes • At night hypoglycaemic episodes • Severe hypoglycaemic episodes • HbA1c change • Weight change • BMI change

For these two review questions, the below criteria were used:

- If the population of an included study was ≥75% of people not taking antihyperglycaemic medications at the start of the trial (this could be treatment naive, or it could be people taking medications but with a wash-out period of at least 6 weeks before the trial treatment is started) then it was included in review question 1.1 initial therapy.
- If the population of an included study was ≥75% of people taking antihyperglycaemic medications at the start of the trial then it was included in review question 1.2 subsequent therapy.
- If the population of an included study was <75% of people not taking antihyperglycaemic medications at the start of the trial and <75% of people taking antihyperglycaemic medications at the start of the trial, then the study could not be classified for either review question and was excluded, unless results were reported separately for the two populations.

2.1.1 Stratification

Stratification is applied where the committee are confident the intervention will work differently in the groups and separate recommendations are required, therefore they should be reviewed separately. In this guideline all analyses were stratified for people with type 2 diabetes into the following five population models:

- Model 1: People with type 2 diabetes mellitus and heart failure
- Model 2: People with type 2 diabetes mellitus and atherosclerotic cardiovascular disease
- Model 3: People with type 2 diabetes mellitus and chronic kidney disease
- Model 4: People with type 2 diabetes mellitus and lower risk of developing cardiovascular disease
- Model 5: People with type 2 diabetes mellitus and higher risk of developing cardiovascular disease

Where trials reported a mix of populations across strata, a threshold of 80% with the condition of interest was agreed with the committee as a minimum cut off for what would be acceptable to constitute a predominant group. As such, trials were categorised into one of 4 groups, those:

- with the condition of interest (e.g. ≥80% of participants have heart failure)
- without the condition of interest (e.g. ≥80% of people do not have heart failure)
- with a mixed population (e.g. >20% have heart failure and >20% do not have heart failure)
- where there is insufficient information to determine strata (e.g. not stated/unclear).

Trials were included in the evidence for models 1-3 when the trial was classified as ≥80% of participants with the condition of interest, or if there was a subgroup analysis within the trial which gave results for people with the condition of interest. For model 5, classification of trials was difficult due to trials not reporting independent participant level data in order to determine the proportion of participants in the trial at higher CV risk. However, the committee agreed and prespecified that, due to the background level of CV risk in a high proportion of people with type 2 diabetes, and looking at the average age of the trials and the average presence of comorbidities and risk factors in the trials, that trials falling into the mixed population, or the not stated/unclear classification, were actually likely to represent evidence for people with type 2 diabetes and higher risk of developing cardiovascular disease, and therefore these trials were included in the evidence for model 5. No evidence was identified for the model 4 population.

The following criteria were used to classify each trial into the appropriate strata:

Table 2: Criteria for population stratification

Condition of interest	Accepted definitions
Heart failure	• People with heart failure with or without preserved ejection fraction; • New York Heart Association (NYHA) Class II, III, or IV; • American College of Cardiology Foundation/American Heart Association (ACCF/AHA) Stages, B, C, and D.
Atherosclerotic heart disease	• Ischaemic stroke; • Transient ischaemic attack; • Coronary heart disease, including ST-elevation and non=ST-elevation myocardial infarction (STEMI and NSTEMI); • Peripheral arterial disease;

	• Coronary and non-coronary revascularization procedures.
Chronic kidney disease	• People with chronic kidney disease (as defined in study); • Presence of nephropathy.
Higher cardiovascular risk	• QRISK2>10% in adults aged≥40 years old • Presence of one or more cardiovascular risk factors in adult aged <40 years old.

2.1.2 Extra considerations

2.1.2.1 Initial treatment review

To be defined as initial therapy, it was agreed that one of the following had to be met:

- The trial population had to have received no pharmacological therapy for type 2 diabetes before entering the study (be truly drug naïve)
- The trial population had to have participated in a washout of all previous type 2 diabetes drug therapy for a period of at least 6 weeks before entering the treatment period of the study
- The trial had to report another satisfactory method by which participants would be effectively drug naïve entering the study, that could be verified as acceptable by the committee

Based on this, a sensitivity analysis was developed to investigate whether trials where participants were not truly drug naïve had a significant effect on the results, specifically whether these provided an enrichment effect.

Studies were grouped into the following categories:

1. Including only responders
    a. Run-in treatment washout, excluding those that get better/stay the same when medication withdrawn
    b. Run-in treatment washout, for example, including only those whose HbA1c increases on withdrawal
    c. Run-in period with test treatment, inclusion criteria are only those who respond to the treatment
    d. Inclusion criteria – known previous responders to the treatment being tested
2. Excluding non-responders (but not exclusively including only responders)
    a. Run-in treatment washout, excluding those that get better when the treatment is withdrawn
    b. Exclusion criteria – previous non-response/intolerance to the treatment under investigation
3. Selection of specific population
    a. Inclusion criteria – using a specific therapy, exclusion criteria – responding to a treatment.
    b. Exclusion criteria – those who are currently well controlled with minimal adverse events.
    c. Those who have not responded to adequate trials of one or more other treatments.
    d. Run-in placebo period, excluding people that respond to placebo.
    e. Response criteria (either via run-in period, washout or inclusion criteria) related to a treatment not being tested (e.g. DPP-4 inhibitor responders randomised to metformin versus placebo).
4. Unclear

    a. Previous response to treatment is unclear, some implicit suggestion that selection based on response may have taken place
5. All treatment naïve
    a. All participants are naïve to the treatment being tested
6. No response criteria
    a. Population washed out of previous treatment but no inclusion criteria based on withdrawal response
7. Mixed population
    a. Inclusion criteria for people already receiving treatment fits into the above, but treatment naïve participants also included (and these accounted for >20% of the population)
8. Not reported
    a. No reference to selection based on previous treatment or washout period. If it is not possible to infer from the study report then population can be considered as treatment naïve but study should be highlighted as high risk of bias.

When the analysis was conducted, a sensitivity analysis was partaken where only studies included in group 5 (all treatment naïve) were included in the analysis. More information can be found in section 2.5.1.3.

2.1.2.2 Subsequent treatment review

In this review, an additional factor considered in the analysis was the strategy used when managing treatment. This included whether:

- Treatment was added to existing treatment (for example: adding an SGLT-2 inhibitor to metformin)
- Treatment was being switched from another treatment (for example: switching a GLP-1 agonist for a DPP-4 inhibitor)
- Treatment was being stopped (for example: stopping pioglitazone, but continuing insulin and metformin)

The different strategies were separated in the analysis and the differences between them were considered by the committee as a part of the discussion of the results.

2.2 Searching for evidence

2.2.1 Clinical and health economics literature searches

The full strategy including population terms, intervention terms, study types applied, the databases searched, and the years covered can be found in Appendix B of the evidence reviews for initial therapy (E1 initial) and subsequent therapy (F2 subsequent).

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.

Databases were searched using relevant medical subject headings, free-text terms and study-type filters where appropriate. Searches were restricted to papers published in English. Studies published in languages other than English were not reviewed. All searches were updated on 12th October 2023 and 22nd March 2024. Papers published or added to databases after this date were not considered. Where new evidence was identified, for example in consultation comments received from stakeholders, the impact on the guideline was considered and the action agreed between the technical team and NICE staff with a quality assurance role.

Search strategies were quality assured by the following approaches. Medline search strategies were checked by a second information specialist. Searches were cross-checked with reference lists of relevant papers, searches in other systematic reviews were analysed, and committee members were requested to highlight key studies.

Searching for unpublished literature was not undertaken. 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.

2.3 Reviewing evidence

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

- Potentially relevant studies were identified from the search results by reviewing titles and abstracts. The full papers were then obtained.
- Full papers were evaluated against the pre-specified inclusion and exclusion criteria set out in the protocol to identify studies that addressed the review question. The review protocols are included in an appendix to each of the evidence reports.
- Relevant studies were critically appraised using the preferred study design checklist as specified in the NICE guidelines manual and reported in the review protocol.
- Key information was extracted about interventional study methods and results into EPPI reviewer version 5 and Microsoft Excel. Summary evidence tables were produced from data entered into EPPI Reviewer.
- Summaries of the evidence were generated by outcome. Outcome data from the randomised trials were meta-analysed where appropriate using Review Manager version 5.3.
- Summaries of the quality of outcomes were generated using the GRADE process. This was done using an automated process with data inputs generated from Review Manager and Microsoft Power BI being imported into the GRADE Buddy R shiny tool and generating GRADE tables. Additional quality assurance of all GRADE tables took place during this guideline to ensure that this was conducted accurately.
- A minimum of 10% of the abstracts were reviewed by two analysts, with any disagreements resolved by discussion or, if necessary, a third independent analyst.
- All of the evidence reviews were quality assured by a senior technical analyst. This included checking:
    - papers were included or excluded appropriately
    - a sample of at least 10% of the data extractions for each analyst
    - a sample of at least 10% of the risk of bias assessments for each analyst
    - correct methods were used to synthesise data.

Discrepancies if identified, were resolved through discussion (with a third analyst where necessary).

- Studies identified during the rerun searches were checked for whether they would add any additional information that would change the results of the analysis. This was discussed with the guideline development team and the committee. If it was agreed that they would, then they underwent data extraction and further analysis as relevant. If it was agreed that they likely would not, then they were noted down in the excluded study section and agreed that they would be extracted in a future update of the guideline instead.

2.3.1 Types of studies and inclusion and exclusion criteria

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

Conference abstracts were not considered for inclusion. Literature reviews, posters, letters, editorials, comment articles, unpublished studies and studies not in published in English language were excluded.

2.3.1.1 Type of studies

Randomised controlled trials (RCTs) were included in the evidence reviews where identified because they are considered the most robust type of study design that can produce an unbiased estimate of the intervention effects.

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

2.4 Methods of combining evidence

2.4.1 Data synthesis for intervention reviews

Meta-analyses were conducted using Cochrane Review Manager (RevMan5) software[8]. Network meta-analyses (NMAs) were conducted either in RStudio version 2023.09.0 Build 463, R version 4.3.3, (2024-02-29 ucrt) using the multinma package version 0.6.1.9003, or in WinBUGS, version 1.4.3[9], using standard and adapted TSD codes (WinBUGS code).

2.4.1.1 Analysis of different types of data

Dichotomous outcomes

Fixed-effects (Mantel–Haenszel) techniques were used to calculate risk ratios (relative risk, RR) for the binary outcomes. The absolute risk difference was also calculated using the methods in the GRADE handbook in the GRADE Buddy R shiny tool app[1], 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 as they are more appropriate for data with a low number of events. Where there are zero events in both arms, the risk difference was calculated and reported instead.

Time to event data

Where sufficient information was provided, hazard ratios were reported in addition to risk ratios for dichotomous outcomes. Both hazard ratios and risk ratios were presented. If there

were differences in effect estimates between the two measures, potential reasons for this were considered in the interpretation of the evidence.

Continuous outcomes

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

Where the studies within a single meta-analysis reported either change from baseline or final values these were pooled, with a preference for the change scores if a single study reported both.

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) software[8].

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. If the control event rate was reported this was used to generate the absolute risk difference in the GRADE Buddy shiny tool app. If multivariate analysis was used to derive the summary statistic but no adjusted control event rate was reported no absolute risk difference was calculated.

2.4.1.2 Network meta-analysis

For detailed methods on the network meta-analyses, see methods in reports F9-F13 for methods specific to each network meta-analysis.

2.5 Appraising the quality of evidence by outcomes

2.5.1 Intervention reviews

The evidence for outcomes from the included RCTs were evaluated and presented using the 'Grading of Recommendations Assessment, Development and Evaluation (GRADE) toolbox' developed by the international GRADE working group[1]. The software (GRADEpro) 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 3.

Table 3: Description of quality elements in GRADE for intervention studies

Quality element	Description
Risk of bias	Limitations in the study design and implementation may bias the estimates of the treatment effect. Major limitations in studies decrease the confidence in the estimate of the effect. Examples of such limitations are selection bias (often due to poor allocation concealment), performance and detection bias (often due to a lack of blinding of the patient, healthcare professional or assessor) and attrition bias (due to missing data causing systematic bias in the analysis).
Indirectness	Indirectness refers to differences in study population, intervention, comparator and outcomes between the available evidence and the review question.
Inconsistency	Inconsistency refers to an unexplained heterogeneity of effect estimates between studies in the same meta-analysis.

Quality element	Description
Imprecision	Results are imprecise when studies include relatively few patients and few events (or highly variable measures) and thus have wide confidence intervals around the estimate of the effect relative to clinically important thresholds. 95% confidence intervals denote the possible range of locations of the true population effect at a 95% probability, and so wide confidence intervals may denote a result that is consistent with conflicting interpretations (for example a result may be consistent with both clinical benefit AND clinical harm) and thus be imprecise.
Publication bias	Publication bias is a systematic underestimate or overestimate of the underlying beneficial or harmful effect due to the selective publication of studies. A closely related phenomenon is where some papers fail to report an outcome that is inconclusive, thus leading to an overestimate of the effectiveness of that outcome.
Other issues	Sometimes randomisation may not adequately lead to group equivalence of confounders, and if so this may lead to bias, which should be taken into account. Potential conflicts of interest, often caused by excessive pharmaceutical company involvement in the publication of a study, should also be noted.

Details of how the 4 main quality elements (risk of bias, indirectness, inconsistency and imprecision) were appraised for each outcome are given below. Publication bias was considered with the committee, but was not suspected to be present in any of the analyses.

2.5.1.1 Risk of bias

The main domains of bias for RCTs are listed in Table 4. Each outcome had its risk of bias assessed within each study first using the appropriate checklist for the study design (Cochrane RoB 2 for RCTs, or ROBIS for systematic reviews). For each study, if there was no risk of bias in any domain, the risk of bias was given a rating of 'low risk of bias'. An overall judgment of 'some concerns' was made if some concerns were present in at least one domain for this outcome, and no domain was judged to be at high risk of bias. An overall judgment of 'high risk of bias' was made if high risk of bias was present in at least one domain or there were 'some concerns' for multiple domains in a way that substantially lowers confidence in the result. An overall rating of; not serious, serious or very serious, is applied in GRADEpro across all studies combined in a meta-analysis by taking into account the weighting of studies according to study precision.

Table 4: Principle domains of bias in randomised controlled trials

Limitation	Explanation
Selection bias (sequence generation and allocation concealment)	If those enrolling participants are aware of the group to which the next enrolled patient will be allocated, either because of a non-random sequence that is predictable, or because a truly random sequence was not concealed from the researcher, this may translate into systematic selection bias. This may occur if the researcher chooses not to recruit a participant into that specific group because of: • knowledge of that participant's likely prognostic characteristics, and • a desire for one group to do better than the other.
Performance and detection bias (lack of blinding)	Patients, caregivers, those adjudicating or recording outcomes, and data analysts should not be aware of the arm to which the participants are allocated. Knowledge of the group can influence: • the experience of the placebo effect • performance in outcome measures • the level of care and attention received, and • the methods of measurement or analysis all of which can contribute to systematic bias.

Limitation	Explanation
Attrition bias	Attrition bias results from an unaccounted for loss of data beyond a certain level (a differential of at least 10% between groups). Loss of data can occur when participants are compulsorily withdrawn from a group by the researchers (for example, when a per-protocol approach is used) or when participants do not attend assessment sessions. If the missing data are likely to be different from the data of those remaining in the groups, and there is a differential rate of such missing data from groups, systematic attrition bias may result.
Selective outcome reporting	Reporting of some outcomes and not others on the basis of the results can also lead to bias, as this may distort the overall impression of efficacy.
Other limitations	For example: • Stopping early for benefit observed in randomised trials, in particular in the absence of adequate stopping rules. • Use of unvalidated patient-reported outcome measures. • Lack of washout periods to avoid carry-over effects in crossover trials. • Recruitment bias in cluster-randomised trials.

2.5.1.2 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 'directly applicable'. If there was indirectness in just 1 source (for example in terms of population), indirectness was given a rating of partially applicable, but if there was indirectness in 2 or more sources (for example, in terms of population and treatment) the indirectness was given an 'indirectly applicable' rating. An overall rating of; not serious, serious or very serious, was applied in GRADEpro across all studies combined in a meta-analysis by taking into account the weighting of studies according to study precision.

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

Heterogeneity or inconsistency amongst studies was also visually inspected. Where statistical heterogeneity as defined above was present or there was clear visual heterogeneity not captured in the I^2 value, predefined subgrouping of studies was carried out according to the protocol provided there were at least 4 studies included in the forest plot to allow for meaningful subgroup analyses. See the review protocols for the subgrouping strategy. Heterogeneity was also investigated with sensitivity analyses as follows:

- investigating the removal of studies at high risk of bias or very high risk of bias;
- for 1.1 initial therapy, removal of studies where the population was not treatment naïve at the start of the trial.

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

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

If all predefined strategies of subgrouping were unable to explain statistical heterogeneity, then a random effects (DerSimonian and Laird) model was employed to the entire group of studies in the meta-analysis. A random-effects model assumes a distribution of populations, rather than a single population. This leads to a widening of the confidence interval around the overall estimate.

2.5.1.4 Imprecision

The criteria applied for imprecision were based on the 95% CIs for the pooled estimate of effect, and the minimal important differences (MID) for the outcome. The MIDs are the threshold for appreciable benefits and harms, separated by a zone either side of the line of no effect where there is assumed to be no clinically important effect. If either end of the 95% CI of the overall estimate of effect crossed 1 of the MID lines, imprecision was regarded as serious in the GRADEpro rating. 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. This was because the overall result was consistent with all 3 interpretations defined by the MID (no clinically important effect, clinical benefit and clinical harm). This is illustrated in Figure 1.

The value / position of the MID lines is ideally determined by values reported in the literature. ‘Anchor-based’ methods aim to establish clinically meaningful changes in a continuous outcome variable by relating or ‘anchoring’ them to patient-centred measures of clinical effectiveness that could be regarded as gold standards with a high level of face validity. For example, a MID for an outcome could be defined by the minimum amount of change in that outcome necessary to make patients feel their quality of life had ‘significantly improved’. MIDs in the literature may also be based on expert clinician or consensus opinion concerning the minimum amount of change in a variable deemed to affect quality of life or health.

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

- For dichotomous outcomes the MIDs were taken to be RRs of 0.8* and 1.25. For ‘positive’ outcomes such as ‘patient satisfaction’, the RR of 0.8 is taken as the line denoting the boundary between no clinically important effect and a clinically important harm, whilst the RR of 1.25 is taken as the line denoting the boundary between no clinically important effect and a clinically important benefit. For ‘negative’ outcomes such as ‘bleeding’, the opposite occurs, so the RR of 0.8 is taken as the line denoting the boundary between no clinically important effect and a clinically important benefit, whilst the RR of 1.25 is taken as the line denoting the boundary between no clinically important effect and a clinically important harm. There aren’t established default values for ORs and the same values (0.8 and 1.25) are applied here but are acknowledged as arbitrary thresholds agreed by the committee.
    - In cases where there are zero events in one arm of a single study, or some or all of the studies in one arm of a meta-analysis, the same process was followed as for dichotomous outcomes. However if there are no events in either arm in a meta-analysis (or in a single unpooled study) the sample size was used to determine imprecision using the following rule of thumb:
        - No imprecision: sample size ≥350

  - Serious imprecision: sample size ≥70 but <350
  - Very serious imprecision: sample size <70.
- When there was more than one study in an analysis and zero events occurred in both groups for some but not all of the studies across both arms, the optimum information size was used to determine imprecision using the following guide:
  - No imprecision: >90% power
  - Serious imprecision: 80-90% power
  - Very serious imprecision: <80% power.

- Time to event data: there are no established default values for HRs so the same values as dichotomous outcomes are applied here (0.8 and 1.25). However, these are acknowledged as arbitrary thresholds agreed by the committee.
- For continuous outcome variables the MID was taken as half the median baseline standard deviation of that variable, across all studies in the meta-analysis. Hence the MID denoting the minimum clinically important benefit was positive for a 'positive' outcome (for example, a quality of life measure where a higher score denotes better health), and negative for a 'negative' outcome (for example, a visual analogue scale [VAS] pain score). Clinically important harms will be the converse of these. If baseline values were unavailable, then half the median comparator group standard deviation of that variable was taken as the MID. As these vary for each outcome per review, details of the values used are reported in the footnotes of the relevant GRADE summary table.

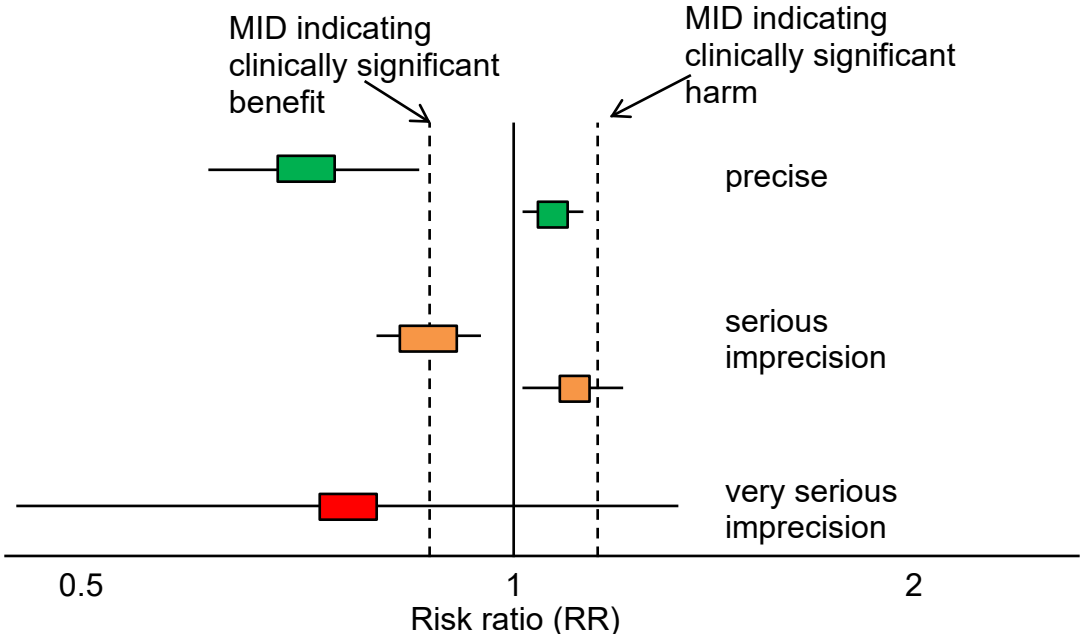
*To note: GRADE report the default values as 0.75 and 1.25. These are consensus values. This guideline follows NICE process to use modified values of 0.8 and 1.25 as they are symmetrical on a relative risk scale.

For this guideline, the following MIDs for continuous or dichotomous outcomes were found in the literature or agreed by consensus and adopted for use:

Table 5: Published or pre-agreed MIDs

Outcome measure	MID	Source
EQ-5D	±0.03	Consensus pragmatic MID used in previous NICE guidelines
SF-36	Physical component summary: ±2 Mental component summary: ±3 Physical functioning: ±3 Role-physical: ±3 Bodily pain: ±3 General health: ±2 Vitality: ±2 Social functioning: ±3 Role-emotional: ±4 Mental health: ±3	User's manual for the SF-36v2 Health Survey, Third Edition[2]
HbA1c change	±0.5% or ±5.5 mmol/mol	Consensus pragmatic MID used in previous NICE guidelines
Weight change	±3% or 2.4 kg	Consensus pragmatic MID agreed by committee. 2.4 kg agreed based on the average baseline weight of participants in all trials.
BMI change	±3% or 0.8 kg/m^2	Consensus pragmatic MID agreed by committee 0.8 kg/m^2 agreed based on the average baseline BMI of participants in all trials.

Figure 1: Illustration of precise and imprecise outcomes based on the 95% CI of dichotomous outcomes in a forest plot (Note that all 3 results would be pooled estimates, and would not, in practice, be placed on the same forest plot)



2.5.1.5 Overall grading of the quality of clinical evidence

Once an outcome had been appraised for the main quality elements, as above, an overall quality grade was calculated for that outcome. The ratings from each of the main quality elements were summed to give an overall rating from high to very low. The evidence for each outcome started at High, and the overall quality (or confidence in the evidence) remained High if there were no reasons for downgrading, or became Moderate, Low or Very Low according to the number of independent reasons for downgrading. The significance of these overall ratings is explained in Table 6. The reasons for downgrading in each case are specified in the footnotes of the GRADE tables.

Table 6: 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

2.5.2 Intervention studies – NMA

Risk of bias and indirectness for the included studies was assessed as described above for the pairwise analyses and was presented in a summary table for the studies and outcomes included in the NMA.

Inconsistency in the network was assessed as described in section 3.3 of report F10 on network meta-analysis.

2.5.2.1 Modified GRADE for intervention studies analysed using NMA

A modified version of the standard GRADE approach for pairwise interventions was used to assess the quality of evidence across the network meta-analyses. While most criteria for pairwise meta-analyses still apply, it is important to adapt some of the criteria to take into consideration additional factors, such as how each 'link' or pairwise comparison within the network applies to the others. As a result, the following was used when modifying the GRADE framework to a network meta-analysis. It is designed to provide a single overall quality rating for an NMA to judge the overall strength of evidence.

Table 7: Rationale for downgrading quality of evidence for network meta-analysis

GRADE criteria	Reasons for downgrading quality
Risk of bias	Not serious: If fewer than 33.3% of the studies in the network meta-analysis were at moderate or high risk of bias Serious: If greater than 33.3% of the studies in the network meta-analysis were at moderate or high risk of bias Very serious: If greater than 33.3% of the studies in the network meta-analysis were at high risk of bias
Indirectness	Not serious: If fewer than 33.3% of the studies in the network meta-analysis were partially indirect or indirect Serious: If greater than 33.3% of the studies in the network meta-analysis were partially indirect or indirect Very serious: If greater than 33.3% of the studies in the network meta-analysis were indirect
Inconsistency	N/A: If there were no links in the network where data from multiple studies (either direct or indirect) were synthesised. Serious: if the DIC for an inconsistency model was more than 3 points lower than the corresponding consistency model or, for a random effects model, the between studies standard deviation was meaningfully lower for the inconsistency model than the corresponding consistency model.
Imprecision	95% Credible intervals are used to assess imprecision. Not serious: The data were sufficiently precise to allow the committee to draw conclusions from the results of the NMA. Serious: Imprecision had a moderate impact on the ability of the committee to draw conclusions from the results of the NMA. Very serious: Imprecision had a substantial impact on the committee to draw conclusions from the results of the NMA.

2.6 Assessing clinical importance

The committee assessed the evidence by outcome in order to determine if there was, or potentially was, a clinically important benefit, a clinically important harm or no clinically important difference between interventions. To facilitate this, binary outcomes were converted into absolute risk differences (ARDs) using the approach in the GRADE handbook using the GRADE Buddy R shiny app: the median control group risk across studies was used to calculate the ARD and its 95% CI from the pooled risk ratio. For calculation of ARDs from hazard ratios estimated from the NMA, see sections 3.1.5 and 3.2 of report F10.

The assessment of clinical benefit, harm, or no benefit or harm was based on the point estimate of absolute effect for intervention studies, which was standardised across the reviews. The committee considered a number of factors when determining a minimally important difference for clinical importance for each outcome including: the baseline rate of the outcome, the relative importance of the percentage reduction from baseline and the impact of the event on the person with the condition (Table 8). Based on this they agreed

that some outcomes required lower thresholds for clinical importance due to their potential impact (for example: mortality outcomes and development of end-stage renal disease), while others may have higher thresholds based on the events being more common at baseline and so a larger absolute effect being required (for example: 3-item MACE). The baseline rates for the outcomes were determined from the trial populations included in the review. In some cases, this was not possible due to heterogeneity in the reported values, where this was the case values from related outcomes or default values were used instead. The minimally important differences used can be found in Table 9.

Table 8: Criteria used by the committee to consider the threshold for clinical importance for each outcome

How rare is the event at baseline?	How high is the absolute value?	Percentage reduction from baseline	Impact of event
>1 in 10 (common)	**Higher** values may be achievable (30+ per 1000)	Less useful	Useful
1 in 10-1 in 100 (moderate)	**Moderate** values may be achievable (10 per 1000)	Case-by-case, may be useful	Useful
<1 in 100 (rare)	**Small** values may be achievable (5 per 1000)	Useful	Useful

What is the impact of the event if untreated?
High - leading to death, very significant impact on quality of life, or introducing significant health impact into daily living
Moderate - impacts on quality of life but not as significantly, introduces a health impact into daily life – but not with a significant constraint
Minor - minor impact on quality of life, introduces minimal health impact into daily life

Table 9: Minimally important differences agreed for dichotomous outcomes with corresponding baseline rate category and impact

Outcome	Baseline rate (in people with type 2 diabetes at higher risk of cardiovascular disease unless otherwise stated)	Impact	Minimally important difference
All-cause mortality	1 in 10-1 in 100 (most groups)/>1 in 10 (people with type 2 diabetes and heart failure)	High	1 per 1,000
Cardiovascular mortality	60 out of 1,000 (1 in 10-1 in 100)	High	1 per 1,000
3-item MACE	122 out of 1,000 (>1 in 10)	Moderate	25 per 1,000

4-item MACE	132 out of 1,000 (>1 in 10)	Moderate	30 per 1,000
5-item MACE	154 out of 1,000 (>1 in 10)	Moderate	40 per 1,000
Non-fatal myocardial infarction	68 out of 1,000 (1 in 10-1 in 100)	Low-Moderate	15 per 1,000
Non-fatal stroke	38 out of 1,000 (1 in 10-1 in 100)	Low-Moderate	10 per 1,000
Unstable angina	27 out of 1,000 (1 in 10-1 in 100)	Low-Moderate	10 per 1,000
Hospitalisation for heart failure	53 out of 1,000 (1 in 10-1 in 100) for people with type 2 diabetes at higher risk of cardiovascular disease to 130 out of 1,000 (>1 in 10) for people with type 2 diabetes with heart failure	Moderate-High	10 per 1,000 for most groups 30 per 1,000 for people with pre-existing heart failure
Acute kidney injury	20 out of 1,000	Moderate	10 per 1,000
Persistent signs of worsening kidney disease	38 out of 1,000	Low-Moderate	15 per 1,000
Development of end-stage kidney disease	9 out of 1,000 (<1 in 100) for people with type 2 diabetes at higher risk of cardiovascular disease to 75 out of 1,000 (1 in 10-1 in 100) for people with type 2 diabetes and chronic kidney disease	High	1 per 1,000
Death from renal causes	<1 in 100	High	1 per 1,000
Cardiac arrhythmia	30 out of 1,000 (1 in 10-1 in 100)	Low	15 per 1,000
Diabetic ketoacidosis	1 out of 1,000 (<1 in 100)	Moderate	2 per 1,000
Falls requiring hospitalisation	4 out of 1,000 (<1 in 100)	Moderate	2 per 1,000
Progression of liver disease	10 out of 1,000 (1 in 10-1 in 100)	Low-Moderate	2 per 1,000
Remission	Not estimable	Low	100 per 1,000 (default value for a beneficial effect)

Hypoglycaemia episodes	155 out of 1,000 (>1 in 10)	Low	25 per 1,000
At night hypoglycaemic episodes	Difficult to determine from data sample available	Low	25 per 1,000
Severe hypoglycaemic episodes	18 out of 1,000 (1 in 10-1 in 100)	Moderate	10 per 1,000

For continuous outcomes if the mean difference was greater than the minimally important difference (MID) then this represented a clinical benefit or harm. Established MIDs found in the literature and were agreed to be used for the SF-36. For other continuous outcomes, the MIDs provided in Table 5 were used.

2.7 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.[3]

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

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

2.7.1 Literature review

The health economists:

- Identified potentially relevant studies for each review question from the health economic search results by reviewing titles and abstracts. Full papers were then obtained.
- Reviewed full papers against prespecified inclusion and exclusion criteria to identify relevant studies (see below for details).
- Critically appraised relevant studies using economic evaluations checklists as specified in the NICE guidelines manual.[3]
- Extracted key information about the studies' methods and results into health economic evidence tables (which can be found in appendices to the relevant evidence reports).
- Generated summaries of the evidence in NICE health economic evidence profile tables (included in the relevant evidence report for each review question) – see below for details.

2.7.1.1 Inclusion and exclusion criteria

Full economic evaluations (studies comparing costs and health consequences of alternative courses of action: cost–utility, cost-effectiveness, cost–benefit and cost–consequences

analyses) and comparative costing studies that addressed the review question in the relevant population were considered potentially includable as health economic evidence.

Studies that only reported cost per hospital (not per patient), or only reported average cost effectiveness without disaggregated costs and effects were excluded. Literature reviews, abstracts, posters, letters, editorials, comment articles, unpublished studies and studies not in English were excluded. Studies published before 2008 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 8 below and the economic evaluation checklist (appendix H of the NICE guidelines manual[3]) and the health economics review protocol, which can be found in each of the evidence reports.

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

2.7.1.2 NICE health economic evidence profiles

NICE health economic evidence profile tables were used to summarise cost and cost-effectiveness estimates for the included health economic studies in each evidence review report. The health economic evidence profile shows an assessment of applicability and methodological quality for each economic study, with footnotes indicating the reasons for the assessment. These assessments were made by the health economist using the economic evaluation checklist from the NICE guidelines manual.[3] It also shows the incremental costs, incremental effects (for example, quality-adjusted life years [QALYs]) and incremental cost-effectiveness ratio (ICER) for the base case analysis in the study, as well as information about the assessment of uncertainty in the analysis. See Table 8 for more details.

When a non-UK study was included in the profile, the results were converted into pounds sterling using the appropriate purchasing power parity.[7]

Table 10: Content of NICE health economic evidence profile

Item	Description
Study	Surname of first author, date of study publication and country perspective with a reference to full information on the study.
Applicability	An assessment of applicability of the study to this guideline, the current NHS situation and NICE decision-making:[(a)] • Directly applicable – the study meets all applicability criteria, or fails to meet 1 or more applicability criteria but this is unlikely to change the conclusions about cost effectiveness. • Partially applicable – the study fails to meet 1 or more applicability criteria, and this could change the conclusions about cost effectiveness. • Not applicable – the study fails to meet 1 or more of the applicability criteria, and this is likely to change the conclusions about cost effectiveness. Such studies would usually be excluded from the review.
Limitations	An assessment of methodological quality of the study:[(a)]

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

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

2.7.2 Undertaking new health economic analysis

As well as reviewing the published health economic literature for each review question, as described above, new health economic analysis was undertaken by the health economist in all areas.

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

- Methods were consistent with the NICE reference case for interventions with health outcomes in NHS settings.[3, 5]
- 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 NICE.

Full methods and results of the cost-effectiveness analysis are described in a separate economic analysis report.

2.7.3 Cost-effectiveness criteria

NICE sets out the principles that committees should consider when judging whether an intervention offers good value for money.[43, 6] In general, an intervention was considered to be cost effective (given that the estimate was considered plausible) if either of the following criteria applied:

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

If the committee recommended an intervention that was estimated to cost more than £20,000 per QALY gained, or did not recommend one that was estimated to cost less than £20,000 per QALY gained, the reasons for this decision are discussed explicitly in 'The committee's discussion of the evidence' section of the relevant evidence report, with reference to issues regarding the plausibility of the estimate or to factors set out in NICE methods manuals.[3]

2.7.4 In the absence of health economic evidence

When making recommendations in areas not in the scope of the health economic analysis and where no relevant published evidence was identified 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.

2.8 Developing recommendations

Over the course of the guideline development process, the committee was presented with:

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

Decisions on whether a recommendation could be made, and if so in which direction, were made on the basis of the committee's interpretation of the available evidence, taking into account the balance of benefits, harms and costs between different courses of action. This was either done formally in an economic model, or informally. The net clinical benefit over harm (clinical effectiveness) was considered, focusing on the magnitude of the effect (or clinical importance), quality of evidence (including the uncertainty) and amount of evidence available. When this was done informally, the committee took into account the clinical benefits and harms when one intervention was compared with another. The assessment of net clinical benefit was moderated by the importance placed on the outcomes (the committee's values and preferences), and the confidence the committee had in the evidence (evidence quality). Secondly, the committee assessed whether the net clinical benefit justified any differences in costs between the alternative interventions. When the clinical harms were judged by the committee to outweigh any clinical benefits, they considered making a recommendation not to offer an intervention. This was dependant on whether the intervention had any reasonable prospect of providing cost-effective benefits to people using services and whether stopping the intervention was likely to cause harm for people already receiving it.

When clinical and health economic evidence was of poor quality, conflicting or absent, the committee decided on whether a recommendation could be made based on its expert opinion. The considerations for making consensus-based recommendations include the balance between potential harms and benefits, the economic costs compared to the economic benefits, current practices, recommendations made in other relevant guidelines, patient preferences and equality issues. The consensus recommendations were agreed through discussions in the committee. The committee also considered whether the uncertainty was sufficient to justify delaying making a recommendation to await further research, taking into account the potential harm of failing to make a clear recommendation (see section 2.8.1 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)[3].

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

2.8.1 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.

2.8.2 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.

2.8.3 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.

2.9 General terms

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

Term	Definition
	neither the patients, clinicians or the people carrying out the statistical analysis know which treatment patients received.
Carer (caregiver)	Someone who looks after family, partners or friends in need of help because they are ill, frail or have a disability.
Case–control study	A study to find out the cause(s) of a disease or condition. This is done by comparing a group of patients who have the disease or condition (cases) with a group of people who do not have it (controls) but who are otherwise as similar as possible (in characteristics thought to be unrelated to the causes of the disease or condition). This means the researcher can look for aspects of their lives that differ to see if they may cause the condition. For example, a group of people with lung cancer might be compared with a group of people the same age that do not have lung cancer. The researcher could compare how long both groups had been exposed to tobacco smoke. Such studies are retrospective because they look back in time from the outcome to the possible causes of a disease or condition.
Case series	Report of a number of cases of a given disease, usually covering the course of the disease and the response to treatment. There is no comparison (control) group of patients.
Clinical efficacy	The extent to which an intervention is active when studied under controlled research conditions.
Clinical effectiveness	How well a specific test or treatment works when used in the 'real world' (for example, when used by a doctor with a patient at home), rather than in a carefully controlled clinical trial. Trials that assess clinical effectiveness are sometimes called management trials. Clinical effectiveness is not the same as efficacy.
Clinician	A healthcare professional who provides patient care. For example, a doctor, nurse or physiotherapist.
Cochrane Review	The Cochrane Library consists of a regularly updated collection of evidence-based medicine databases including the Cochrane Database of Systematic Reviews (reviews of randomised controlled trials prepared by the Cochrane Collaboration).
Cohort study	A study with 2 or more groups of people – cohorts – with similar characteristics. One group receives a treatment, is exposed to a risk factor or has a particular symptom and the other group does not. The study follows their progress over time and records what happens. See also observational study.
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)	A range of values for an unknown population parameter with a stated 'confidence' (conventionally 95%) that it contains the true value. The interval is calculated from sample data, and generally straddles the sample estimate. The 'confidence' value means that if the method used to calculate the interval is repeated many times, then that proportion of intervals will actually contain the true value.

Term	Definition
Confounding factor	Something that influences a study and can result in misleading findings if it is not understood or appropriately dealt with. For example, a study of heart disease may look at a group of people that exercises regularly and a group that does not exercise. If the ages of the people in the 2 groups are different, then any difference in heart disease rates between the 2 groups could be because of age rather than exercise. Therefore age is a confounding factor.
Consensus methods	Techniques used to reach agreement on a particular issue. Consensus methods may be used to develop NICE guidance if there is not enough good quality research evidence to give a clear answer to a question. Formal consensus methods include Delphi and nominal group techniques.
Control group	A group of people in a study who do not receive the treatment or test being studied. Instead, they may receive the standard treatment (sometimes called 'usual care') or a dummy treatment (placebo). The results for the control group are compared with those for a group receiving the treatment being tested. The aim is to check for any differences. Ideally, the people in the control group should be as similar as possible to those in the treatment group, to make it as easy as possible to detect any effects due to the treatment.
Cost–benefit analysis (CBA)	Cost–benefit analysis is one of the tools used to carry out an economic evaluation. The costs and benefits are measured using the same monetary units (for example, pounds sterling) to see whether 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

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

Term	Definition
Extrapolation	An assumption that the results of studies of a specific population will also hold true for another population with similar characteristics.
Follow-up	Observation over a period of time of an individual, group or initially defined population whose appropriate characteristics have been assessed in order to observe changes in health status or health-related variables.
Generalisability	The extent to which the results of a study hold true for groups that did not participate in the research. See also external validity.
Gold standard	A method, procedure or measurement that is widely accepted as being the best available to test for or treat a disease.
GRADE, GRADE profile	A system developed by the GRADE Working Group to address the shortcomings of present grading systems in healthcare. The GRADE system uses a common, sensible and transparent approach to grading the quality of evidence. The results of applying the GRADE system to clinical trial data are displayed in a table known as a GRADE profile.
Harms	Adverse effects of an intervention.
Health economics	Study or analysis of the cost of using and distributing healthcare resources.
Health-related quality of life (HRQoL)	A measure of the effects of an illness to see how it affects someone's day-to-day life.
Heterogeneity or Lack of homogeneity	The term is used in meta-analyses and systematic reviews to describe when the results of a test or treatment (or estimates of its effect) differ significantly in different studies. Such differences may occur as a result of differences in the populations studied, the outcome measures used or because of different definitions of the variables involved. It is the opposite of homogeneity.
Imprecision	Results are imprecise when studies include relatively few patients and few events and thus have wide confidence intervals around the estimate of effect.
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.

Term	Definition
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 Logit model	In statistics, logistic regression is a type of analysis used for predicting the outcome of a binary dependent variable based on one or more predictor variables. It can be used to estimate the log of the odds (known as the 'logit').
Loss to follow-up	A patient, or the proportion of patients, actively participating in a clinical trial at the beginning, but whom the researchers were unable to trace or contact by the point of follow-up in the trial
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.

Term	Definition
Observational study	Individuals or groups are observed or certain factors are measured. No attempt is made to affect the outcome. For example, an observational study of a disease or treatment would allow 'nature' or usual medical care to take its course. Changes or differences in one characteristic (for example, whether or not people received a specific treatment or intervention) are studied without intervening. There is a greater risk of selection bias than in experimental studies.
Odds ratio	A measure of treatment effectiveness. The odds of an event happening in the treatment group, expressed as a proportion of the odds of it happening in the control group. The 'odds' is the ratio of events to non-events.
Opportunity cost	The loss of other healthcare programmes displaced by investment in or introduction of another intervention. This may be best measured by the health benefits that could have been achieved had the money been spent on the next best alternative healthcare intervention.
Outcome	The impact that a test, treatment, policy, programme or other intervention has on a person, group or population. Outcomes from interventions to improve the public's health could include changes in knowledge and behaviour related to health, societal changes (for example, a reduction in crime rates) and a change in people's health and wellbeing or health status. In clinical terms, outcomes could include the number of patients who fully recover from an illness or the number of hospital admissions, and an improvement or deterioration in someone's health, functional ability, symptoms or situation. Researchers should decide what outcomes to measure before a study begins.
P value	The p value is a statistical measure that indicates whether or not an effect is statistically significant. For example, if a study comparing 2 treatments found that one seems more effective than the other, the p value is the probability of obtaining these, or more extreme results by chance. By convention, if the p value is below 0.05 (that is, there is less than a 5% probability that the results occurred by chance) it is considered that there probably is a real difference between treatments. If the p value is 0.001 or less (less than a 1% probability that the results occurred by chance), the result is seen as highly significant. If the p value shows that there is likely to be a difference between treatments, the confidence interval describes how big the difference in effect might be.
Placebo	A fake (or dummy) treatment given to participants in the control group of a clinical trial. It is indistinguishable from the actual treatment (which is given to participants in the experimental group). The aim is to determine what effect the experimental treatment has had – over and above any placebo effect caused because someone has received (or thinks they have received) care or attention.
Polypharmacy	The use or prescription of multiple medications.
Posterior distribution	In Bayesian statistics this is the probability distribution for a statistic based after combining established information or belief (the prior) with new evidence (the likelihood).
Power (statistical)	The ability to demonstrate an association when one exists. Power is related to sample size; the larger the sample size, the greater the power and the lower the risk that a possible association could be missed.
Preoperative	The period before surgery commences.
Prevalence	See Pre-test probability.

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

Term	Definition
	condition. Unlike prospective studies, it does not cover events that occur after the study group is selected.
Review question	In guideline development, this term refers to the questions about treatment and care that are formulated to guide the development of evidence-based recommendations.
Risk ratio (RR)	The ratio of the risk of disease or death among those exposed to certain conditions compared with the risk for those who are not exposed to the same conditions (for example, the risk of people who smoke getting lung cancer compared with the risk for people who do not smoke). If both groups face the same level of risk, the risk ratio is 1. If the first group had a risk ratio of 2, subjects in that group would be twice as likely to have the event happen. A risk ratio of less than 1 means the outcome is less likely in the first group. The risk ratio is sometimes referred to as relative risk.
Secondary outcome	An outcome used to evaluate additional effects of the intervention deemed a priori as being less important than the primary outcomes.
Selection bias	Selection bias occurs if: a) The characteristics of the people selected for a study differ from the wider population from which they have been drawn, or b) There are differences between groups of participants in a study in terms of how likely they are to get better.
Sensitivity analysis	A means of representing uncertainty in the results of economic evaluations. Uncertainty may arise from missing data, imprecise estimates or methodological controversy. Sensitivity analysis also allows for exploring the generalisability of results to other settings. The analysis is repeated using different assumptions to examine the effect on the results. One-way simple sensitivity analysis (univariate analysis): each parameter is varied individually in order to isolate the consequences of each parameter on the results of the study. Multi-way simple sensitivity analysis (scenario analysis): 2 or more parameters are varied at the same time and the overall effect on the results is evaluated. Threshold sensitivity analysis: the critical value of parameters above or below which the conclusions of the study will change are identified. Probabilistic sensitivity analysis: probability distributions are assigned to the uncertain parameters and are incorporated into evaluation models based on decision analytical techniques (for example, Monte Carlo simulation).
Significance (statistical)	A result is deemed statistically significant if the probability of the result occurring by chance is less than 1 in 20 ($p<0.05$).
Stakeholder	An organisation with an interest in a topic that NICE is developing a guideline or piece of public health guidance on. Organisations that register as stakeholders can comment on the draft scope and the draft guidance. Stakeholders may be: • manufacturers of drugs or equipment • national patient and carer organisations • NHS organisations • organisations representing healthcare professionals.
State transition model	See Markov model
Stratification	When a different estimate effect is thought to underlie two or more groups based on the PICO characteristics. The groups are therefore kept separate from the outset and are not combined in a meta-

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

References