

Hypertension in adults: diagnosis and management

NICE guideline: methods

NICE guideline NG136

Methods

August 2019

Final

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

Disclaimer

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

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

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 All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

© NICE 2019. All rights reserved. Subject to Notice of rights

Contents

1	Development of the guideline	5
1.1	What is a NICE guideline?.....	5
1.2	Remit.....	5
1.3	Who developed this guideline?.....	6
1.3.1	What this guideline covers	6
1.3.2	What this guideline does not cover.....	6
1.3.3	Relationships between the guideline and other NICE guidance	7
2	Methods	8
2.1	Developing the review questions and outcomes	8
2.2	Searching for evidence.....	13
2.2.1	Clinical and health economics literature searches.....	13
2.3	Identifying and analysing evidence of effectiveness	14
2.3.1	Inclusion and exclusion criteria	14
2.3.2	Type of studies.....	15
2.3.3	Methods of combining clinical studies	16
2.3.4	Appraising the quality of evidence by outcomes.....	19
2.3.5	Assessing clinical importance	26
2.3.6	Clinical evidence statements.....	26
2.4	Identifying and analysing evidence of cost effectiveness	27
2.4.1	Literature review	27
2.4.2	Undertaking new health economic analysis.....	29
2.4.3	Cost-effectiveness criteria	30
2.4.4	In the absence of health economic evidence.....	30
2.5	Developing recommendations	30
2.5.1	Research recommendations	31
2.5.2	Validation process.....	31
2.5.3	Updating the guideline	32
2.5.4	Disclaimer	32
2.5.5	Funding.....	32
3	Acronyms and abbreviations	33
4	Glossary	36
4.1	Guideline-specific terms	36
4.2	General terms	37

1 Development of the guideline

1.1 What is a NICE guideline?

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

NICE guidelines can:

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

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

We produce our guidelines using the following steps:

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

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

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

NICE Pathways brings together all connected NICE guidance.

1.2 Remit

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

The remit for this guideline is:

To update the NICE guideline on hypertension in adults: diagnosis and management (CG127) and to incorporate the identification of malignant hypertension and update the

recommendations on blood pressure management for adults with type 2 diabetes from NG28 type 2 diabetes in adults.

1.3 Who developed this guideline?

A multidisciplinary guideline committee comprising health professionals and researchers as well as lay members developed this guideline (see the list of guideline committee members and the acknowledgements).

The National Institute for Health and Care Excellence (NICE) funds the National Guideline Centre (NGC) and thus supported the development of this guideline. The committee was convened by the NGC and chaired by Anthony Wierzbicki in accordance with guidance from NICE.

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

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

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

1.3.1 What this guideline covers

This guideline is a partial update of NICE guideline Hypertension in adults: diagnosis and management. It updates a number of recommendations while also investigating clinical areas not addressed by the previous guideline. The population covered is adults (people who are 18 years and older) with hypertension, with or without type 2 diabetes. The clinical areas are blood pressure measurement for diagnosis and ongoing monitoring, treatment initiation and targets, treatment with antihypertension medication, relaxation therapies, and the identification of who to refer for same-day specialist review.

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

1.3.2 What this guideline does not cover

This guideline does not cover preventing hypertension, screening for hypertension, specialist management of secondary hypertension (that is, hypertension arising from other medical conditions), or secondary prevention of hypertension for people with established cardiovascular disease.

Areas from Hypertension in adults: diagnosis and management that have not been updated are:

- Measuring blood pressure
- Assessing cardiovascular risk and target organ damage

- Lifestyle interventions (other than relaxation therapies)
- Patient education and adherence to treatment

1.3.3 Relationships between the guideline and other NICE guidance

Related NICE guidelines:

- Multimorbidity: clinical assessment and management (2016) NICE guideline NG56
- Type 2 diabetes in adults: management (2015) NICE guideline NG28
- Type 1 diabetes in adults: diagnosis and management (2015) NICE guideline NG17
- Chronic kidney disease in adults: assessment and management (2014) NICE guideline CG182
- Chronic heart failure: diagnosis and management (2018) NICE guideline NG106
- Cardiovascular disease: risk assessment and reduction, including lipid modification (2014) NICE guideline CG181
- Atrial fibrillation: management (2014) NICE guideline CG180
- Hypertension in pregnancy: diagnosis and management (2011) NICE guideline CG107

Related NICE guidance currently in development:

- Acute coronary syndromes NICE guideline 07 May 2020
- Acute kidney injury: prevention, detection and management (update) NICE guideline 17 June 2020
- Atrial fibrillation: management NICE guideline 17 September 2020
- Chronic kidney disease: assessment and management (update) NICE guideline 17 June 2020
- Hypertension in pregnancy: diagnosis and management (update) NICE guideline 25 June 2019
- Stroke and transient ischaemic attack in over 16s: diagnosis and initial management (update) NICE guideline 01 May 2019

2 Methods

This report sets out in detail the methods used to review the evidence and to develop the recommendations that are presented in each of the evidence reviews for this guideline. This guidance was developed in accordance with the methods outlined in the NICE guidelines manual, 2014 version.³

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

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



2.1 Developing the review questions and outcomes

Review questions were developed using a PICO framework (population, intervention, comparison and outcome) for intervention reviews using a framework of population, index tests; reference standard and target condition for reviews of diagnostic test accuracy; using population, presence or absence of factors under investigation (for example, prognostic factors) and outcomes for prognostic reviews.

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

A total of 11 review questions were identified. Some of these were separated into 2 strata for people with and without type 2 diabetes, resulting in 16 review questions in total.

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

Table 1: Review questions

Evidence report	Type of review	Review questions	Outcomes
A	Diagnostic (RCT)	In adults with suspected primary hypertension, what is the best method of measuring blood pressure (home, ambulatory or clinic measurement) to establish the diagnosis and prevent cardiovascular events?	<p>Critical</p> <ul style="list-style-type: none"> All-cause mortality Health-related quality of life Stroke (ischaemic or haemorrhagic) Myocardial infarction <p>Important</p> <ul style="list-style-type: none"> Heart failure needing hospitalisation Vascular procedures (including both coronary and carotid artery procedures) Angina needing hospitalisation Side effect 1: intolerance to device
A	Diagnostic (accuracy)	In adults with suspected primary hypertension, what is the best method of measuring blood pressure (home, ambulatory or clinic measurement) to establish the diagnosis and prevent cardiovascular events?	<p>Critical</p> <ul style="list-style-type: none"> Sensitivity Specificity Raw data to calculate 2x2 tables to calculate sensitivity and specificity <p>Important</p> <ul style="list-style-type: none"> Area under the curve Likelihood ratios Predictive values
B	Intervention	In adults with treated primary hypertension, what is the best method of measuring blood pressure (home, ambulatory or clinic measurement) to assess the response to treatment and prevent cardiovascular events?	<p>Critical</p> <ul style="list-style-type: none"> All-cause mortality Health-related quality of life Stroke (ischaemic or haemorrhagic) Myocardial infarction <p>Important</p> <ul style="list-style-type: none"> Reduction in clinic blood pressure (BP) Proportion of people controlled to a target Average daily dose of antihypertensive medication Average number of visits Side effect 1: Intolerance to device Side effect 2: Hypotension (dizziness)

Evidence report	Type of review	Review questions	Outcomes
C	Intervention	At what blood pressure and/or cardiovascular disease risk threshold should antihypertensive drug treatment be initiated for adults with hypertension?	<p>Critical</p> <ul style="list-style-type: none"> • All-cause mortality • Health-related quality of life • Stroke (ischaemic or haemorrhagic) • Myocardial infarction <p>Important</p> <ul style="list-style-type: none"> • Heart failure needing hospitalisation • Vascular procedures (including lower limb, coronary and carotid artery procedures) • Angina needing hospitalisation • Side effect 1: Acute kidney injury • Side effect 2: New onset diabetes • Side effect 3: Treatment related admission • Side effect 4: Hypotension (dizziness)
D	Intervention	Should targets used for antihypertensive therapy be based on blood pressure, cardiovascular risk or a combination of both?	<p>Critical</p> <ul style="list-style-type: none"> • All-cause mortality • Health-related quality of life • Stroke (ischaemic or haemorrhagic) • Myocardial infarction <p>Important</p> <ul style="list-style-type: none"> • Heart failure needing hospitalisation • Vascular procedures (including lower limb, coronary and carotid artery procedures) • Angina needing hospitalisation • Discontinuation or dose reduction due to side effects • Resource use • Side effect 1: Acute kidney injury • Side effect 2: New onset diabetes • Side effect 3: Change in creatinine or eGFR • Side effect 4: Hypotension (dizziness)
D	Intervention	What is the optimum blood pressure or cardiovascular risk	<p>Critical</p> <ul style="list-style-type: none"> • All-cause mortality • Health-related quality of life

Evidence report	Type of review	Review questions	Outcomes
		target for adults with treated primary hypertension?	<ul style="list-style-type: none"> • Stroke (ischaemic or haemorrhagic) • Myocardial infarction Important • Heart failure needing hospitalisation • Vascular procedures (including lower limb, coronary and carotid artery procedures) • Angina needing hospitalisation • Discontinuation or dose reduction due to side effects • Resource use • Side effect 1: Acute kidney injury • Side effect 2: New onset diabetes • Side effect 3: Change in creatinine or eGFR • Side effect 4: Hypotension (dizziness)
E	Intervention	Is monotherapy or combination antihypertensive therapy more clinically and cost effective for step 1 treatment for hypertension in adults?	<p>Critical</p> <ul style="list-style-type: none"> • All-cause mortality • Health-related quality of life • Stroke (ischaemic or haemorrhagic) • Myocardial infarction <p>Important</p> <ul style="list-style-type: none"> • Heart failure needing hospitalisation (including lower limb, coronary and carotid artery procedures) • Angina needing hospitalisation • Discontinuation or dose reduction due to side effects • Side effect 1: Acute kidney injury • Side effect 2: New onset diabetes • Side effect 3: Change in creatinine or eGFR • Side effect 4: Hypotension (dizziness)
F	Intervention	What is the most clinically and cost effective sequence for step 2 and step 3 treatment for hypertension in adults?	<p>Critical</p> <ul style="list-style-type: none"> • All-cause mortality • Health-related quality of life • Stroke (ischaemic or haemorrhagic) • Myocardial infarction

Evidence report	Type of review	Review questions	Outcomes
			<p>Important</p> <ul style="list-style-type: none"> • Heart failure needing hospitalisation • Vascular procedures (including lower limb, coronary and carotid artery procedures) • Angina needing hospitalisation • Discontinuation or dose reduction due to side effects • Side effect 1: Acute kidney injury • Side effect 2: New onset diabetes • Side effect 3: Change in creatinine or eGFR • Side effect 4: Hypotension (dizziness)
G	Intervention	What is the most clinically and cost-effective step 4 antihypertensive drug treatment for hypertension in adults?	<p>Critical</p> <ul style="list-style-type: none"> • All-cause mortality • Health-related quality of life • Stroke (ischaemic or haemorrhagic) • Myocardial infarction <p>Important</p> <ul style="list-style-type: none"> • Heart failure needing hospitalisation • Vascular procedures (including lower limb, coronary and carotid artery procedures) • Angina needing hospitalisation • Discontinuation or dose reduction due to side effects • Side effect 1: Acute kidney injury • Side effect 2: New onset diabetes • Side effect 3: Change in creatinine or eGFR • Side effect 4: Hypotension (dizziness)
H	Intervention	What is the clinical and cost-effectiveness of relaxation therapies for the management of primary hypertension in adults?	<p>Critical</p> <ul style="list-style-type: none"> • All-cause mortality • Health-related quality of life • Stroke (ischaemic or haemorrhagic) • Myocardial infarction

Evidence report	Type of review	Review questions	Outcomes
			Important <ul style="list-style-type: none"> Heart failure needing hospitalisation Vascular procedures (including both coronary and carotid artery procedures) Angina needing hospitalisation Cessation or reduction of medication
I	Prognostic	What factors indicate the need for same-day specialist review (including the possible presence of malignant or accelerated hypertension)?	Critical <ul style="list-style-type: none"> All-cause mortality Stroke Diagnosis of malignant or accelerated hypertension Hospitalisation Renal dialysis

2.2 Searching for evidence

2.2.1 Clinical and health economics literature searches

Systematic literature searches were undertaken to identify all published clinical and health economic evidence relevant to the review questions. Searches were undertaken according to the parameters stipulated within the NICE guidelines manual 2017 (see <https://www.nice.org.uk/process/pmg20/>). Databases were searched using relevant medical subject headings, free-text terms and study-type filters where appropriate. Where possible, searches were restricted to papers published in English. Studies published in languages other than English were not reviewed. All searches were updated on 4 October 2018. Papers published or added to databases after this date were not considered. If new evidence, falling outside of the time frame for the guideline searches, is identified, for example in consultation comments received from stakeholders, the impact on the guideline will be considered, and any further action agreed between NGC and NICE staff with a quality assurance role. One study published after the last cut-off date was considered because its impact on recommendations was judged to be potentially substantial and required amending the full evidence review to assess the impact on recommendations.

Prior to running, search strategies were quality assured using a variety of approaches. A second information specialist checked Medline search strategies before being run. Searches were crosschecked with reference lists of highly relevant papers, searches in other systematic reviews were analysed, and committee members were requested to highlight additional studies.

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

- Guidelines International Network database (www.g-i-n.net)
- National Guideline Clearing House (www.guideline.gov)
- National Institute for Health and Care Excellence (NICE; www.nice.org.uk)
- National Institutes of Health Consensus Development Program (consensus.nih.gov)
- NHS Evidence Search (www.evidence.nhs.uk).

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

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

2.3 Identifying and analysing evidence of effectiveness

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

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

2.3.1 Inclusion and exclusion criteria

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

The key population inclusion criteria was:

- Adults with hypertension, with or without type 2 diabetes

The key population exclusion criteria was:

- Pregnant women and women planning pregnancy
- Children and young people (aged under 18 years)
- People with chronic kidney disease A3 or above (heavy proteinuria), or A2 or above for people with type 2 diabetes
- People with type 1 diabetes
- People with secondary causes of hypertension such as tumours or structural vascular defects (including Conn's adenoma, pheochromocytoma and renovascular hypertension).
- People with established cardiovascular disease

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

Studies published prior to 2000 were excluded from evidence reviews that focused on the measurement of blood pressure, such as the diagnosis, monitoring and initiating treatment reviews. This was due to differences in the devices used to measure blood pressure before this date as compared to the present day. Devices used prior to 2000 have been phased out, and due to differences in blood pressure measurement compared to currently used devices, studies published before this date were not included within these evidence reviews.

2.3.2 Type of studies

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

For most intervention reviews in this guideline, parallel randomised controlled trials (RCT) were included because they are considered the most robust type of study design that can produce an unbiased estimate of the intervention effects. RCTs that had a follow up of more than 12 months were included, because a difference in hard outcomes between interventions may not be detected within a shorter follow-up period. Crossover RCTs were included if a sufficient washout period of 4 weeks between treatments was undertaken, otherwise they were excluded. The committee agreed that in the majority of intervention reviews, lower quality evidence (non-randomised studies) would not be adequate enough to inform changes in current practice, and therefore would not be included. One exception was made (initiation of treatment) where non-randomised evidence could provide a direct comparison between treatment thresholds that is less feasible within randomised trials. The committee stated a priori in the protocol that either certain identified variables must be equivalent at baseline or else the analysis had to adjust for any baseline differences. If the study did not fulfil either criterion, it was excluded.

For diagnostic review questions, diagnostic RCTs, cross-sectional studies and cohort studies were included. For prognostic review questions, cohort studies and case-controls were included.

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. Where multiple systematic reviews were available, a decision was made to include the

review most appropriate to the review question based on multiple criteria such as the date of publication, availability of risk of bias assessments and the extent to which the systematic review answered the review question.

2.3.3 Methods of combining clinical studies

2.3.3.1 Data synthesis for intervention reviews

Where possible, meta-analyses were conducted using Cochrane Review Manager (RevMan5)⁷ software to combine the data given in all studies for each of the outcomes of interest for the review question.

Evidence reviews relating to the management of hypertension, other than relaxation therapies, were stratified by the presence or absence of type 2 diabetes.

2.3.3.1.1 Analysis of different types of data

Dichotomous outcomes

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

- All-cause mortality
- Stroke (ischaemic or haemorrhagic)
- Myocardial infarction
- Heart failure needing hospitalisation
- Vascular procedures (including both coronary and carotid artery procedures)
- Angina needing hospitalisation
- Proportion of people controlled to a target
- Discontinuation due to side effects
- Acute kidney injury
- New onset diabetes
- Hypotension (dizziness)
- Resource use

The absolute risk difference was also calculated using GRADEpro¹ software using the median event rate in the control arm of the pooled results.

For binary variables where there were 0 events in either arm or a less than 1% event rate, Peto odds ratios, rather than risk ratios, were calculated. Peto odds ratios are more appropriate for data with a low number of events. Where Peto odds ratios have been used, the risk difference was utilised to calculate an absolute effect. For binary variables where there are 0 events in both arms, the risk difference was utilised to calculate an absolute effect.

Where sufficient information was provided, hazard ratios (HR) were included in preference to RRs or odds ratios for outcomes such as mortality where the time to the event occurring was important for decision-making. In these cases the event rates in the intervention and control arms were manually entered into GRADEpro¹ in order to calculate absolute risk difference.

Continuous outcomes

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

- Health-related quality of life (HRQoL)

- Change in creatinine or estimated glomerular filtration rate (eGFR)
- Average daily dose of antihypertensive medication
- Reduction in clinic blood pressure

The means and standard deviations of continuous outcomes are required for meta-analysis. However, in cases where standard deviations were not reported, this was calculated if the p values, 95% confidence intervals (CI; 95% CI) or standard error were reported. Where p values were reported as 'less than', a conservative approach was undertaken. For example, if a p value was reported as ' $p \leq 0.001$ ', the calculations for standard deviations were based on a p value of 0.001. If these statistical measures were not available then other forms of standard deviation imputation could be utilised if appropriate, as described in section 16.1.3 of the Cochrane Handbook (version 5.1.0, updated March 2011).

2.3.3.1.2 Generic inverse variance

If a study reported only the summary statistic and 95% CI the generic-inverse variance method was used to enter data into RevMan5.⁷ If the control event rate was reported, this was used to generate the absolute risk difference in GRADEpro.¹ Where control event rates were not reported, these were extrapolated from literature that identified the number of events in the non-treated hypertension population if this population matched the control group arm within the evidence. These values were entered manually into GRADEpro in order to calculate absolute risk difference.¹ If multivariate analysis was used to derive the summary statistic but no adjusted control event rate was reported, no absolute risk difference was calculated.

2.3.3.1.3 Heterogeneity

Statistical heterogeneity was assessed for each meta-analysis estimate by visually inspecting the forest plots, considering the chi-squared test for significance at $p < 0.1$ or an I-squared (I^2) inconsistency statistic (with an I-squared value of more than 50% indicating significant heterogeneity) as well as confidence intervals and estimates regarding the minimally important differences (MID). Where significant heterogeneity was present, predefined subgrouping of studies was carried out for:

- Age
- Presence or absence of type 2 diabetes (in absence of diabetes strata)
- Family origin
- Severity of hypertension

If the subgroup analysis resolved heterogeneity within all of the derived subgroups, then each of the derived subgroups for a given outcome were presented separately (providing at least 1 study remained in each subgroup). Assessments of potential differences in effect between subgroups were based on the chi-squared tests for heterogeneity statistics between subgroups. Within study subgroup analyses were interpreted with caution as separating the groups breaks the study randomisation and as such is subject to uncontrolled confounding.

If all predefined strategies of subgrouping were unable to explain statistical heterogeneity within each derived subgroup, then a random effects (DerSimonian and Laird) model was employed to the entire group of studies in the meta-analysis. A random-effects model assumes a distribution of populations, rather than a single population. This leads to a widening of the confidence interval around the overall estimate, thus providing a more realistic interpretation of the true distribution of effects across more than 1 population. If, however, the committee considered the heterogeneity was so large that meta-analysis was inappropriate, then the results were not pooled and were described narratively.

2.3.3.2 Data synthesis for diagnostic reviews

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

2.3.3.2.1 Diagnostic RCTs

Diagnostic RCTs (sometimes referred to as test-and-treat trials) are a randomised comparison of 2 diagnostic tests, with study outcomes being clinically important consequences of the diagnosis (patient-related outcome measures similar to those in intervention trials, such as mortality). Participants are randomised to receive test A or test B, followed by identical therapeutic interventions based on the results of the test (so someone with a positive result would receive the same treatment regardless of whether they were diagnosed by test A or test B). Downstream patient outcomes are then compared between the 2 groups. As treatment is the same in both arms of the trial, any differences in patient outcomes will reflect the accuracy of the tests in correctly establishing who does and does not have the condition. Data were synthesised using the same methods for intervention reviews (see section 2.3.3.1.1 above).

2.3.3.2.2 Diagnostic accuracy studies

For diagnostic test accuracy studies, a positive result on the index test was found if the patient had values of the measured quantity above or below a threshold value and different thresholds could be used. The committee prespecified the types of measurements and thresholds for diagnosis including whether or not data could be pooled across a range of thresholds. Diagnostic test accuracy measures used in the analysis were area under the receiver operating characteristics (ROC) curve (AUC), and, for different thresholds (if appropriate), sensitivity and specificity. The threshold of a diagnostic test is defined as the value at which the test can best differentiate between those with and without the target condition. In practice this varies amongst studies. If a test has a high sensitivity then very few people with the condition will be missed (few false negatives). For example, a test with a sensitivity of 97% will only miss 3% of people with the condition. Conversely, if a test has a high specificity then few people without the condition would be incorrectly diagnosed (few false positives). For example, a test with a specificity of 97% will only incorrectly diagnose 3% of people who do not have the condition as positive. For this guideline, specificity was considered more important than sensitivity due to the consequences and costs of inappropriate treatment (false positive result). Coupled forest plots of sensitivity and specificity with their 95% CIs across studies (at various thresholds) were produced for each test, using RevMan5.⁷ In order to do this, 2x2 tables (the number of true positives, false positives, true negatives and false negatives) were directly taken from the study if given, or else were derived from raw data or calculated from the set of test accuracy statistics.

Diagnostic meta-analysis was conducted where appropriate, that is, when 3 or more studies were available per test, and where 2x2 data were available. Test accuracy for the studies was pooled using the bivariate method for the direct estimation of summary sensitivity and specificity using a random-effects approach in WinBUGS software.⁹ The bivariate method uses logistic regression on the true positives, true negatives, false positives and false negatives reported in the studies. Overall sensitivity and specificity and confidence regions were plotted (using methods outlined by Novielli 2010⁶). Pooled sensitivity and specificity and their 95% CIs were reported in the clinical evidence summary tables.

Area under the ROC curve (AUC) data for each study were also plotted on a graph where possible for each diagnostic test, although AUC data was not available for most of the evidence within the diagnostic accuracy review. The AUC describes the overall diagnostic accuracy across the full range of thresholds. The following rule of thumb was used for evaluating AUCs:

- ≤ 0.50 : worse than chance

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

2.3.3.3 Data synthesis for prognostic reviews

No evidence was identified within the prognostic review (evidence report I)

2.3.4 Appraising the quality of evidence by outcomes

2.3.4.1 Intervention reviews

The evidence for outcomes from the included RCTs and, where appropriate, non-randomised intervention studies, were evaluated and presented using an adaptation of the 'Grading of Recommendations Assessment, Development and Evaluation (GRADE) toolbox' that the international GRADE working group (<http://www.gradeworkinggroup.org/>) developed. The software (GRADEpro¹) that the GRADE working group developed was used to assess the quality of each outcome taking into account individual study quality and the meta-analysis results.

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

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

Quality element	Description
Risk of bias	Limitations in the study design and implementation may bias the estimates of the treatment effect. Major limitations in studies decrease the confidence in the estimate of the effect. Examples of such limitations are selection bias (often due to poor allocation concealment), performance and detection bias (often due to a lack of blinding of the patient, healthcare professional or assessor) and attrition bias (due to missing data causing systematic bias in the analysis).
Indirectness	Indirectness refers to differences in study population, intervention, comparator and outcomes between the available evidence and the review question.
Inconsistency	Inconsistency refers to an unexplained heterogeneity of effect estimates between studies in the same meta-analysis.
Imprecision	Results are imprecise when studies include relatively few patients and few events (or highly variable measures) and thus have wide confidence intervals around the estimate of the effect relative to clinically important thresholds. 95% confidence intervals denote the possible range of locations of the true population effect at a 95% probability, and so wide confidence intervals may denote a result that is consistent with conflicting interpretations (for example, a result may be consistent with both clinical benefit and clinical harm) and thus be imprecise.
Publication bias	Publication bias is a systematic underestimate or overestimate of the underlying beneficial or harmful effect due to the selective publication of studies. A closely related phenomenon is where some papers fail to report an outcome that is inconclusive, thus leading to an overestimate of the effectiveness of that outcome.
Other issues	Sometimes randomisation may not adequately lead to group equivalence of confounders, and if so this may lead to bias, which should be taken into account. Potential conflicts of interest, often caused by excessive pharmaceutical company involvement in the publication of a study, should also be noted.

Details of how the 4 main quality elements (risk of bias, indirectness, inconsistency and imprecision) were appraised for each outcome are given below. Publication or other bias was only taken into consideration in the quality assessment if it was apparent.

2.3.4.1.1 Risk of bias

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

Table 3: Principle domains of bias in randomised controlled trials

Limitation	Explanation
Selection bias (sequence generation and allocation concealment)	<p>If those enrolling patients are aware of the group to which the next enrolled patient will be allocated, either because of a non-random sequence that is predictable or because a truly random sequence was not concealed from the researcher, this may translate into systematic selection bias. This may occur if the researcher chooses not to recruit a participant into that specific group because of:</p> <ul style="list-style-type: none"> • knowledge of that participant's likely prognostic characteristics, and • a desire for 1 group to do better than the other.
Performance and detection bias (lack of blinding)	<p>Patients, caregivers, those adjudicating or recording outcomes, and data analysts should not be aware of the arm to which patients are allocated. Knowledge of the group can influence:</p> <ul style="list-style-type: none"> • the experience of the placebo effect • performance in outcome measures • the level of care and attention received, and • the methods of measurement or analysis <p>all of which can contribute to systematic bias.</p>
Attrition bias	<p>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.</p>
Selective outcome reporting	<p>Reporting of some outcomes and not others based on the results can also lead to bias, as this may distort the overall impression of efficacy.</p>
Other limitations	<p>For example:</p> <ul style="list-style-type: none"> • 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.

Where IPDs or systematic reviews were included, their original risk of bias assessments were used if these were in line with the guideline methods, for example using the Cochrane risk for bias tool. Where risk of bias assessments were available for some, but not all, studies included within the IPDs or systematic reviews, additional risk of bias assessments were conducted and incorporated with the existing assessments per outcome. All systematic reviews that were included were also critically appraised as a whole using the appropriate study design checklist as specified in the NICE guidelines manual.³ Where risk of bias assessments for the individual studies included within one of the systematic reviews were not available, the critical appraisal of the systematic review itself was incorporated into the GRADE assessments for overall quality assessment per outcome.

2.3.4.1.2 Indirectness

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

Indirectness could also be given a 'very serious' rating of -2 from just 1 source of indirectness based on the extent of the indirectness.

An example of population indirectness is studies including mixed populations including people beyond the scope of this guideline, such as people with chronic kidney disease (CKD), established cardiovascular disease (CVD) and type 1 diabetes. As a general rule in this guideline, studies where 10–15% of the population were not the population of interest were given a 'serious' rating of -1, and studies with 16–20% were given a 'very serious rating of -2'. Studies were excluded where more than 20% of the population were not relevant to the review question (in other words, an indirect population).

An example of intervention indirectness is the use of drugs that are not available in the UK, which could have different treatment effects to those within the same class that are available in the UK; in these cases, the same 20% threshold for inclusion due to indirectness was applied.

Outcomes were also considered for indirectness. If no data were available on each cardiovascular event in isolation, then data for combined major cardiovascular events (as a composite outcome), rather than separate reporting of each event type, were considered for inclusion but given a 'serious indirectness' rating of -1. Similarly, if no data were available for the outcome of myocardial infarction, coronary heart disease was considered but also given a 'serious indirectness' rating of -1.

2.3.4.1.3 Inconsistency

Inconsistency refers to an unexplained heterogeneity of results for an outcome across different studies. When estimates of the treatment effect across studies differ widely, this suggests true differences in the underlying treatment effect, which may be due to differences in populations, settings or doses. Statistical tests for heterogeneity, as detailed below, are heuristic, and reviewers took the I^2 and chi-squared results into account alongside other

information including the relative positions of study confidence intervals and point estimates regarding the MIDs.

When heterogeneity existed within an outcome (chi-squared $p < 0.1$, or $I^2 > 50\%$) but no plausible explanation could be found, the quality of evidence for that outcome was downgraded. Inconsistency for that outcome was given a 'serious' score of -1 if the I^2 was 50–74%, and a 'very serious' score of -2 if the I^2 was 75% or more.

If inconsistency could be explained based on prespecified subgroup analysis (that is, each subgroup had an $I^2 < 50\%$), the committee took this into account and considered whether to make separate recommendations based on the variation in effect across subgroups within the same outcome. In such a situation, the quality of evidence was not downgraded for those emergent outcomes.

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

2.3.4.1.4 Imprecision

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

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

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

- For categorical outcomes, the MIDs were taken to be odds ratio (OR) or RRs of 0.8 and 1.25. For 'positive' outcomes such as 'patient satisfaction', the OR or RR of 0.8 is taken as the line denoting the boundary between no clinically important effect and a clinically significant harm, while the OR or RR of 1.25 is taken as the line denoting the boundary between no clinically important effect and a clinically significant benefit. For 'negative' outcomes such as bleeding, the opposite occurs, so the OR or RR of 0.8 is taken as the line denoting the boundary between no clinically important effect and a clinically significant

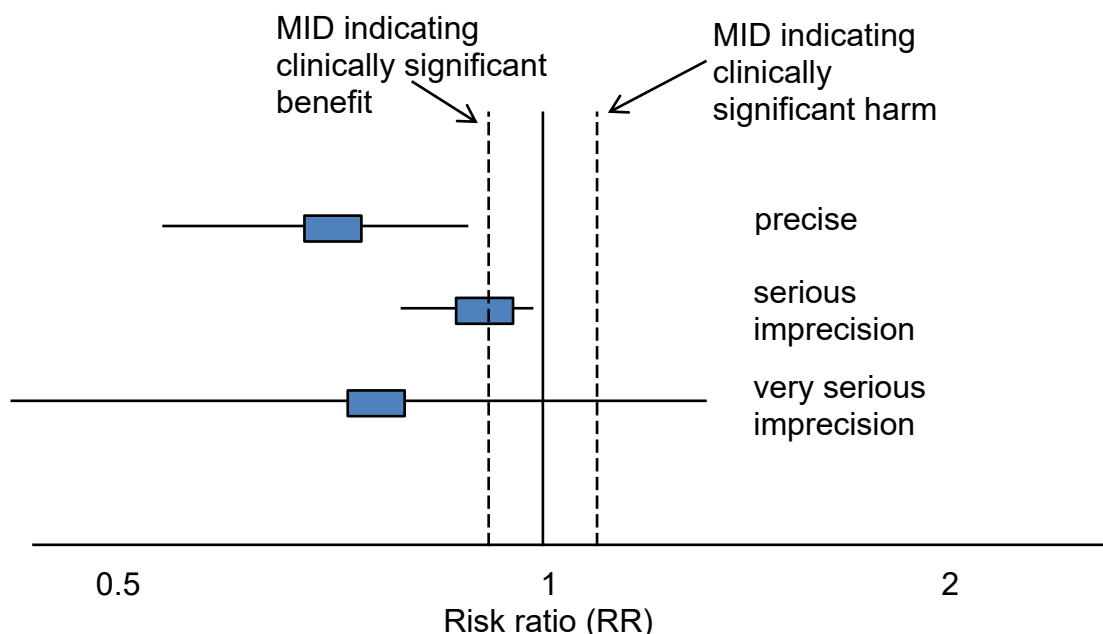
benefit, while the RR of 1.25 is taken as the line denoting the boundary between no clinically important effect and a clinically significant harm.

- For mortality, any change was considered to be clinically important and the imprecision was assessed based on whether the confidence intervals crossed the line of no effect, that is whether the result was consistent with both benefit and harm.
- For continuous outcome variables the MID was taken as half the median baseline standard deviation of that variable, across all studies in the meta-analysis. Hence the MID denoting the minimum clinically significant benefit was positive for a 'positive' outcome (for example, a quality of life measure where a higher score denotes better health), and negative for a 'negative' outcome (for example, a visual analogue scale [VAS] pain score). Clinically significant harms were the converse of these. If baseline values are unavailable, then half the median comparator group standard deviation of that variable was taken as the MID.

The clinical importance rating was subject to amendment after discussion with the committee. If the committee decided that the clinical importance should be altered, after consideration of absolute as well as relative effects, this was allowed provided that any such decision was consistent across the evidence reviews and not influenced by any bias towards making stronger or weaker recommendations for specific outcomes.

For this guideline, established MIDs were used for SF-36⁸ quality of life outcomes. For all other outcomes, no appropriate MIDs for continuous or dichotomous outcomes were found in the literature, and the default method of assessing imprecision was adopted.

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



2.3.4.1.5 Overall grading of the quality of clinical evidence

Once an outcome had been appraised for the main quality elements, as above, an overall quality grade was calculated for that outcome. The scores (0, -1 or -2) from each of the main quality elements were summed to give a score that could be anything from 0 (the best possible) to -8 (the worst possible). However, scores were capped at -3. This final score

was then applied to the starting grade that had originally been applied to the outcome by default, based on study design. All RCTs started as high and the overall quality became moderate, low or very low if the overall score was -1, -2 or -3 points respectively. The significance of these overall ratings is explained in Table 4. The reasons for downgrading in each case were specified in the footnotes of the GRADE tables.

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

Table 4: Overall quality of outcome evidence in GRADE

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

2.3.4.2 Diagnostic studies

Risk of bias and indirectness of evidence for diagnostic data were evaluated by study using the Quality Assessment of Diagnostic Accuracy Studies version 2 (QUADAS-2) checklists (see appendix H in the NICE guidelines manual 2014³). Risk of bias and applicability in primary diagnostic accuracy studies in QUADAS-2 consists of 4 domains (see Figure 3):

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

Figure 3: Summary of QUADAS-2 with list of signalling, risk of bias and applicability questions.

Domain	Patient selection	Index test	Reference standard	Flow and timing
Description	Describe methods of patient selection. Describe included patients (prior testing, presentation, intended use of index test and setting)	Describe the index test and how it was conducted and interpreted	Describe the reference standard and how it was conducted and interpreted	Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram). Describe the time interval and any interventions between index test(s) and reference standard
Signalling questions (yes/no/unclear)	Was a consecutive or random sample of patients enrolled?	Were the index test results interpreted without knowledge of the results of the reference standard?	Is the reference standard likely to correctly classify the target condition?	Was there an appropriate interval between index test(s) and reference standard?

Domain	Patient selection	Index test	Reference standard	Flow and timing
	Was a case–control design avoided?	If a threshold was used, was it pre-specified?	Were the reference standard results interpreted without knowledge of the results of the index test?	Did all patients receive a reference standard?
	Did the study avoid inappropriate exclusions?			Did all patients receive the same reference standard?
				Were all patients included in the analysis?
Risk of bias (high/low/unclear)	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Concerns regarding applicability (high/low/unclear)	Are there concerns that the included patients do not match the review question?	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?	

2.3.4.2.1 *Inconsistency*

Inconsistency refers to an unexplained heterogeneity of results for an outcome across different studies. Inconsistency was assessed by visual inspection of the sensitivity and specificity plots and summary area under the curve (AUC) plots. Particular attention was placed on the specificity threshold the committee set as an acceptable level to recommend a test (80%) and on AUC values above or below 50% (where diagnosis is based on chance alone). The evidence was downgraded by 1 increment if the individual study values varied across 2 areas: where specificity or AUC values of individual studies are both above and below 50%, or downgraded by 2 increments if the individual study values varied across 3 areas, where AUC values of individual studies are above and below 50%.

2.3.4.2.2 *Imprecision*

The judgement of precision was based on visual inspection of the confidence region around the summary specificity point from the diagnostic meta-analysis, if a diagnostic meta-analysis was conducted. This was because specificity was agreed to be the primary outcome for decision making for this review. Where a diagnostic meta-analysis was not conducted, imprecision was assessed according to the range of point estimates or, if only 1 study contributed to the evidence, the 95% CI around the single study. As a general rule (after discussion with the committee) a variation of 0–20% was considered precise, 20–40% serious imprecision, and more than 40% very serious imprecision. Imprecision was assessed on the primary outcome measure for decision-making (specificity).

2.3.4.2.3 *Overall grading*

Quality rating started at high for prospective and retrospective cross-sectional studies, and each major limitation (risk of bias, indirectness, inconsistency and imprecision) brought the rating down by 1 increment to a minimum grade of very low, as explained for intervention reviews.

2.3.5 Assessing clinical importance

The committee assessed the evidence by outcome in order to determine if there was, or potentially was, a clinically important benefit, a clinically important harm or no clinically important difference between interventions. To facilitate this, binary outcomes were converted into absolute risk differences (ARD) using GRADEpro¹ software: the median control group risk across studies was used to calculate the ARD and its 95% CI from the pooled risk ratio.

The assessment of clinical benefit, harm, or no benefit or harm was based on the point estimate of absolute effect for intervention studies, which was standardised across the reviews. The committee considered for most of the dichotomous outcomes in the intervention reviews that if at least 50 more participants per 1,000 (5%) achieved the outcome of interest in the intervention group compared to the comparison group for a positive outcome, then this intervention was considered beneficial. The same point estimate but in the opposite direction applied for a negative outcome. The actual threshold differed slightly per outcome, where the severity, and the resource implications, of cardiovascular events or adverse events were taken into account.

For continuous outcomes if the mean difference was greater than the minimally important difference (MID) then this represented a clinical benefit or harm. Established MIDs were used where possible, such as for the SF-36⁸ quality of life scale. For other continuous outcomes, including change in eGFR or creatinine, default MIDs were used. In these cases, the MID was taken as half the median baseline standard deviation of that variable, across all studies in the meta-analysis. Hence the MID denoting the minimum clinically significant benefit was positive for a 'positive' outcome (for example, a quality of life measure where a higher score denotes better health), and negative for a 'negative' outcome (for example, a visual analogue scale [VAS] pain score). Clinically significant harms will be the converse of these. If baseline values are unavailable, then half the median comparator group standard deviation of that variable will be taken as the MID.

For diagnostic accuracy reviews, clinical importance was determined using a threshold set by the committee as an acceptable threshold to recommend a test. These clinical decision thresholds were discussed and agreed with the committee when setting the protocols. The committee agreed that a test with a specificity above 80% would be acceptable. The committee noted that specificity was of most importance and was the primary outcome in this review, and so clinical importance and thus decision making was based mainly on this whilst also taking into account sensitivity and AUC data.

The committee carried out this assessment for each outcome, and GRADE evidence summary tables were produced to compile the committee's assessments of clinical importance per outcome, alongside the evidence quality and the uncertainty in the effect estimate (imprecision).

2.3.6 Clinical evidence statements

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

- The number of studies and the number of participants for a particular outcome.
- An indication of the direction of clinical importance (if 1 treatment is beneficial or harmful compared to the other or whether there is no difference between the 2 tested treatments).
- A description of the overall quality of the evidence (GRADE overall quality).

2.4 Identifying and analysing evidence of cost effectiveness

The committee is required to make decisions based on the best available evidence of both clinical effectiveness and cost effectiveness. Guideline recommendations should be based on the expected costs of the different options in relation to their expected health benefits (that is, their 'cost effectiveness') rather than the total implementation cost. However, the committee will also need to be increasingly confident in the cost effectiveness of a recommendation as the cost of implementation increases. Therefore, the committee may require more robust evidence on the effectiveness and cost effectiveness of any recommendations that are expected to have a substantial impact on resources; any uncertainties must be offset by a compelling argument in favour of the recommendation. The cost impact or savings potential of a recommendation should not be the sole reason for the committee's decision.³

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

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

2.4.1 Literature review

The health economists:

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

2.4.1.1 Inclusion and exclusion criteria

Full economic evaluations (studies comparing costs and health consequences of alternative courses of action: cost–utility, cost-effectiveness, cost–benefit and cost–consequences analyses) and comparative costing studies that addressed the review question in the relevant population were considered potentially includable as health economic evidence.

Studies that only reported cost per hospital (not per patient), or only reported average cost effectiveness without disaggregated costs and effects were excluded. Literature reviews, abstracts, posters, letters, editorials, comment articles, unpublished studies and studies not in English were excluded. Studies published before 2002 and studies from non-OECD countries or the US 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. When exclusions occurred on this basis, they are noted in the relevant evidence report.

For more details about the assessment of applicability and methodological quality, see Table 5 below and the economic evaluation checklist (appendix H of the NICE guidelines manual³)

and the health economics review protocol, which can be found in each of the evidence reports.

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

2.4.1.2 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. The health economist made these assessments 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 5 for more details.

Table 5: 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) <ul style="list-style-type: none"> • 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) <ul style="list-style-type: none"> • 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 1 strategy minus the mean cost of a comparator strategy.
Incremental effects	The mean QALYs (or other selected measure of health outcome) associated with 1 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 manual³

2.4.2 Undertaking new health economic analysis

As well as reviewing the published health economic literature for each review question, as described above, the health economist undertook a new health economic analysis in selected areas. The committee agreed the priority areas for new analysis after forming the review questions and considering the existing health economic evidence.

The committee identified initiating treatment and diagnosis as the highest priority areas for original health economic modelling.

The committee felt that the initiation of antihypertensive drug treatment in specific populations was an important area for modelling. Initiating treatment involves the cost of treatment, but this has to be traded off against the benefit from treatment in terms of cardiovascular events avoided. The clinical evidence review identified evidence in different blood pressure thresholds, but no evidence was identified relating to cardiovascular risk. There was evidence to suggest relative treatment benefit in people with stage 1 hypertension (systolic blood pressure [BP] 140–159 mmHg), in terms of reducing cardiovascular events. However, there was uncertainty about the cost effectiveness of initiating treatment in this population because the same relative treatment benefit would lead to different absolute benefits in people with lower cardiovascular risk compared to people with higher cardiovascular risk. The trade-offs involved in treatment in terms of the costs versus benefits, as well as the benefits versus harms, were captured in a cost–utility analysis comparing antihypertensive treatment with no treatment in people with stage 1 hypertension and different cardiovascular risk levels.

The 2011 hypertension guideline² conducted original modelling comparing the diagnostic measurement methods (clinic blood pressure measurement [CBPM], home blood pressure measurement [HBPM], and ambulatory blood pressure measurement [ABPM]). New diagnostic accuracy data from the clinical review of this hypertension update was identified, showing that HBPM had become more accurate. The committee felt it was important to determine if the gold standard of ABPM was still the most cost effective option. To test this, a new analysis was added to the previous model. The new diagnostic accuracy data was the only input added to the model, and none of the previous analyses (the previous base case or sensitivity analyses) were updated using the new data,

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 (note that for the minor diagnosis model update the new accuracy data was added as a new analysis, run probabilistically and with no further sensitivity analysis).
- Another health economist at the NGC peer-reviewed the model.

Full methods and results of the cost-effectiveness analysis for the treatment initiation threshold for people with stage 1 hypertension, are described in a separate economic analysis report. The methods and results of the minor update to the cost-effectiveness analysis for blood pressure monitoring for confirming a diagnosis of hypertension are described in the economic evidence section of the diagnosis chapter.

2.4.3 Cost-effectiveness criteria

NICE's report 'Social value judgements: principles for the development of NICE guidance' sets out the principles that committees should consider when judging whether an intervention offers good value for money.⁴ In general, an intervention was considered to be cost effective (given that the estimate was considered plausible) if either of the following criteria applied:

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

If the committee recommended an intervention that was estimated to cost more than £20,000 per QALY gained, or did not recommend one that was estimated to cost less than £20,000 per QALY gained, the reasons for this decision are discussed explicitly in 'The committee's discussion of the evidence' section of the relevant evidence report with reference to issues regarding the plausibility of the estimate or to the factors set out in 'Social value judgements: principles for the development of NICE guidance'.⁴

2.4.4 In the absence of health economic evidence

When no relevant published health economic studies were found and a new analysis was not prioritised, the committee made a qualitative judgement about cost effectiveness by considering expected differences in resource use between options and relevant UK NHS unit costs, alongside the results of the review of clinical effectiveness evidence.

The UK NHS costs reported in the guideline are those that were presented to the committee and were correct at the time recommendations were drafted. They may have changed subsequently before the time of publication. However, we have no reason to believe they have changed substantially.

2.5 Developing recommendations

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

- Summaries of clinical and health economic evidence and quality (as presented in evidence reports [A–I]).
- 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 labelled appendix 1).

Recommendations were drafted based on the committee's interpretation of the available evidence taking into account the balance of benefits and the harms and costs between different courses of action. This was either done formally in an economic model or informally. Firstly, the net clinical benefit over harm (clinical effectiveness) was considered focusing on the critical outcomes. When this was done informally, the committee took into account the clinical benefits and harms when 1 intervention was compared with another. The assessment of net clinical benefit was moderated by the importance placed on the outcomes (the committee's values and preferences) and the confidence the committee had in the evidence (evidence quality). Secondly, the committee assessed whether the net clinical benefit justified any differences in costs between the alternative interventions.

When clinical and health economic evidence was of poor quality, conflicting or absent, the committee drafted recommendations based on its expert opinion. The considerations for

making consensus-based recommendations include the balance between potential harms and benefits, the economic costs compared to the economic benefits, current practices, recommendations made in other relevant guidelines, patient preferences and equality issues. The consensus recommendations were agreed through committee discussions.

All recommendations, those supported by clinical evidence, those agreed through committee consensus and combinations of these approaches, involve uncertainty. The guideline developers assessed the quality of the evidence and the committee discussed their personal uncertainty around their consensus. In doing so, the committee also considered whether the uncertainty was sufficient and research lacking to justify making a recommendation for further research (see section 2.5.1 below).

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

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

- The actions health professionals need to take.
- The information readers need to know.
- The strength of the recommendation (for example, the word 'offer' was used for strong recommendations and 'consider' for weaker recommendations).
- The involvement of patients (and their carers if needed) in decisions on treatment and care.
- Consistency with NICE's standard advice on recommendations about drugs, waiting times and ineffective interventions (see section 9.2 in the NICE guidelines manual³).

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

2.5.1 Research recommendations

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

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

2.5.2 Validation process

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

2.5.3 Updating the guideline

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

2.5.4 Disclaimer

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

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

2.5.5 Funding

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

3 Acronyms and abbreviations

Acronym or abbreviation	Description
ABPM	Ambulatory blood pressure monitoring
ACE inhibitor	Angiotensin-converting-enzyme inhibitor
ACR	Albumin:creatinine ratio
AF	Atrial fibrillation
AKI	Acute kidney injury
ARB	Angiotensin II receptor blockers
ARD	Absolute risk differences
AUC	Area under the curve
BMI	Body mass index
BNF	British National Formulary
BP	Blood pressure
CABG	Coronary artery bypass graft
CAD	Coronary artery disease
CBP	Clinic blood pressure
CBPM	Clinic blood pressure monitoring
CCB	Calcium channel blocker
CEA	Cost-effectiveness analysis
CHD	Coronary heart disease
CI	Confidence interval
CKD	Chronic kidney disease
CPRD	Clinical Practice Research Datalink
CUA	Cost–utility analysis
CV	Cardiovascular
CVD	Cardiovascular disease
DA	Deterministic analysis
DBP	Diastolic blood pressure
ECG	Electrocardiogram
eGFR	Estimated glomerular filtration rate
ELSA	English Longitudinal Study of Ageing
EQ-5D	Euroqol 5 dimensions (scale: 0.0 [death] to 1.0 [full health], negative values mean worse than death)
ESH	European Society of Hypertension
FN	False negative
FP	False positive
FU	Follow up
GC	Guideline Committee
GRADE	Grading of recommendations assessment, development and evaluation
HBPM	Home blood pressure monitoring
HCHS	Hospital & Community Health Services
HDL	High-density lipoproteins
HES	Hospital Episodes Statistics

Acronym or abbreviation	Description
HF	Heart failure
HF-REF	Reduced ejection fraction heart failure
HF-PEF	Preserved ejection fraction heart failure
HM	Home monitoring
HRG	Healthcare resource group
HSE	Health survey for England
HTA	Health technology assessment
ICER	Incremental cost-effectiveness ratio
IPD	Individual patient data
IU	International unit
IV	Intravenous
LVH	Left ventricular hypertrophy
MAP	Mean arterial pressure
MD	Mean difference
MI	Myocardial infarction
NGC	National Guideline Centre
NHS EED	NHS Economic Evaluation Database
NICE	National Institute of Health and Care Excellence
NLR	Negative likelihood ratio
NLR	Negative likelihood ratio
NNT	Numbers needed to treat
NPV	Negative predictive value
N/A	Not applicable
NR	Not reported
NRS	Non-randomised study
OBPM	Office blood pressure measurement
OECD	Organisation for Economic Co-operation and Development
ONS	Office of National Statistics
OR	Odds ratio
PA	Probabilistic analysis
PAD	Peripheral arterial disease
PCI	Percutaneous coronary intervention
PES	Post-event state
PHE	Public Health England
PICO	Population intervention comparison outcomes
PLR	Positive likelihood ratio
PO	Per os (by mouth, orally)
PPV	Positive predictive value
QALY	Quality-adjusted life year
RCT	Randomised control trial
RD	Risk difference
ROC	Receiver operating characteristic
RR	Relative risk or risk ratio
RRT	Renal replacement therapy
SA	Stable angina

Acronym or abbreviation	Description
SBP	Systolic blood pressure
SD	Standard deviation
SF-12, SF-36	Short form 12-point or 36-point questionnaire, respectively
SMD	Standardised mean difference
SMR	Standardised mortality ratio
SR	Systematic review
Str	Stroke
TIA	Transient ischaemic attack
TM	Telemonitoring
TN	True negative
TP	True positive
TX/Tx	Treatment
UA	Unstable angina
VAS	Visual analogue scale

4 Glossary

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

4.1 Guideline-specific terms

Term	Definition
Accelerated hypertension	A severe increase in blood pressure to levels over 180/120 mmHg (and often over 220/120 mmHg), associated with new or progressive target organ damage, usually with signs of papilloedema (swelling of the optic nerve) or retinal haemorrhage. Also known as malignant hypertension.
Excessive coffee consumption	5 or more cups of coffee per day.
Emergency symptoms	These are effects of extremely high blood pressure that indicate the likely presence of significant pathology in need of urgent investigation and treatment. These include: <ul style="list-style-type: none"> • symptoms or signs of accelerated hypertension (clinic blood pressure higher than 180/120 mmHg) with signs of retinal haemorrhage or papilloedema • symptoms of suspected pheochromocytoma (for example, labile or postural hypotension, headache, palpitations, pallor, abdominal pain or and diaphoresis)
End organ damage	See target organ damage.
Established cardiovascular disease	Past medical history of stroke/transient ischemic attack (TIA), heart attack, narrowed peripheral arteries or an interventional procedure. Cardiovascular disease (CVD) is a general term for conditions affecting the heart or blood vessels. It's usually associated with a build-up of fatty deposits inside the arteries (atherosclerosis) and an increased risk of blood clots. It can also be associated with damage to arteries in organs such as the brain, heart, kidneys and eyes through deposition of glassy material within the artery walls (arteriosclerosis). CVD is 1 of the main causes of death and disability in the UK, but it can often largely be prevented by leading a healthy lifestyle.
Life-threatening symptoms	These are effects of extremely high blood pressure likely to indicate a risk of death within days. These include new onset confusion, new onset chest pain, signs of new heart failure or acute renal impairment.
Masked hypertension	Clinic blood pressure measurements are normal (less than 140/90 mmHg), but blood pressure measurements are higher when taken outside the clinic using average daytime ABPM or average HBPM blood pressure measurements.
Persistent hypertension	High blood pressure at repeated clinical encounters.
Stage 1 hypertension	Clinic blood pressure ranging from 140/90 mmHg to 159/99 mmHg and subsequent ambulatory blood pressure monitoring (ABPM) daytime average or home blood pressure monitoring (HBPM) average blood pressure ranging from 135/85 mmHg to 149/94 mmHg .
Stage 2 hypertension	Clinic blood pressure is of 160/100 mmHg or higher but less than 179/109 mmHg and subsequent ABPM daytime average or HBPM average blood pressure is of 150/95 mmHg or higher.
Target organ damage	This refers to damage to target organs such as the heart, brain, kidneys and eyes. Examples are left ventricular hypertrophy, chronic kidney disease, hypertensive retinopathy, or increased urine albumin: creatinine ratio.

Term	Definition
White-coat effect	A discrepancy of more than 20/10 mmHg between clinic and average daytime ABPM or average HBPM blood pressure measurements at the time of diagnosis.

4.2 General terms

Term	Definition
Abstract	Summary of a study, which may be published alone or as an introduction to a full scientific paper.
Algorithm (in guidelines)	A flow chart of the clinical decision pathway described in the guideline, where decision points are represented with boxes, linked with arrows.
Allocation concealment	The process used to prevent advance knowledge of group assignment in an RCT. The allocation process should be impervious to any influence by the individual making the allocation, by being administered by someone who is not responsible for recruiting participants.
Applicability	How well the results of a study or NICE evidence review can answer a clinical question or be applied to the population being considered.
Arm (of a clinical study)	Subsection of individuals within a study who receive 1 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 1 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 1 in which neither patients nor the researchers and doctors know which study group the patients are in. A triple blind study is 1 in which neither the

Term	Definition
	patients, clinicians nor the people carrying out the statistical analysis know which treatment patients received.
Carer (caregiver)	Someone who looks after family, partners or friends in need of help because they are ill, frail or have a disability.
Case–control study	<p>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.</p> <p>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.</p>
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	<p>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.</p> <p>Clinical effectiveness is not the same as efficacy.</p>
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)	<p>There is always some uncertainty in research. This is because a small group of patients is studied to predict the effects of a treatment on the wider population. The confidence interval is a way of expressing how certain we are about the findings from a study, using statistics. It gives a range of results that is likely to include the ‘true’ value for the population.</p> <p>The CI is usually stated as ‘95% CI’, which means that the range of values has a 95 in a 100 chance of including the ‘true’ value. For</p>

Term	Definition
	<p>example, a study may state that “based on our sample findings, we are 95% certain that the ‘true’ population blood pressure is not higher than 150 and not lower than 110”. In such a case the 95% CI would be 110 to 150.</p> <p>A wide confidence interval indicates a lack of certainty about the true effect of the test or treatment – often because a small group of patients has been studied. A narrow confidence interval indicates a more precise estimate (for example, if a large number of patients have been studied).</p>
Confounding factor	<p>Something that influences a study and can result in misleading findings if it is not understood or appropriately dealt with.</p> <p>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.</p>
Consensus methods	<p>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.</p>
Control group	<p>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.</p> <p>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.</p>
Cost–benefit analysis (CBA)	<p>Cost–benefit analysis is 1 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.</p>
Cost–consequences analysis (CCA)	<p>Cost–consequences analysis is 1 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.</p>
Cost-effectiveness analysis (CEA)	<p>Cost-effectiveness analysis is 1 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).</p>
Cost-effectiveness model	<p>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.</p>
Cost–utility analysis (CUA)	<p>Cost–utility analysis is 1 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.</p>
Credible interval (CrI)	<p>The Bayesian equivalent of a confidence interval.</p>

Term	Definition
Decision analysis	An explicit quantitative approach to decision-making under uncertainty, based on evidence from research. This evidence is translated into probabilities, and then into diagrams or decision trees which direct the clinician through a succession of possible scenarios, actions and outcomes.
Deterministic analysis	In economic evaluation, this is an analysis that uses a point estimate for each input. In contrast, see Probabilistic analysis
Diagnostic odds ratio	The diagnostic odds ratio is a measure of the effectiveness of a diagnostic test. It is defined as the ratio of the odds of the test being positive if the subject has a disease relative to the odds of the test being positive if the subject does not have the disease.
Discounting	Costs and perhaps benefits incurred today have a higher value than costs and benefits occurring in the future. Discounting health benefits reflects individual preference for benefits to be experienced in the present rather than the future. Discounting costs reflects individual preference for costs to be experienced in the future rather than the present.
Disutility	The loss of quality of life associated with having a disease or condition. See Utility
Dominance	A health economics term. When comparing tests or treatments, an option that is both less effective and costs more is said to be 'dominated' by the alternative.
Drop-out	A participant who withdraws from a trial before the end.
Economic evaluation	An economic evaluation is used to assess the cost effectiveness of healthcare interventions (that is, to compare the costs and benefits of a healthcare intervention to assess whether it is worth doing). The aim of an economic evaluation is to maximise the level of benefits – health effects – relative to the resources available. It should be used to inform and support the decision-making process; it is not supposed 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 1 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

Term	Definition
	trials, observational studies, expert opinion (of clinical professionals or patients).
Exclusion criteria (literature review)	Explicit standards used to decide which studies should be excluded from consideration as potential sources of evidence.
Exclusion criteria (clinical study)	Criteria that define who is not eligible to participate in a clinical study.
Extended dominance	If Option A is both more clinically effective than Option B and has a lower cost per unit of effect, when both are compared with a do-nothing alternative then Option A is said to have extended dominance over Option B. Option A is therefore cost effective and should be preferred, other things remaining equal.
Extrapolation	An assumption that the results of studies of a specific population will also hold true for another population with similar characteristics.
Follow-up	Observation over a period of time of an individual, group or initially defined population whose appropriate characteristics have been assessed in order to observe changes in health status or health-related variables.
Generalisability	The extent to which the results of a study hold true for groups that did not participate in the research. See also external validity.
Gold standard	A method, procedure or measurement that is widely accepted as being the best available to test for or treat a disease.
GRADE, GRADE profile	A system developed by the GRADE Working Group to address the shortcomings of present grading systems in healthcare. The GRADE system uses a common, sensible and transparent approach to grading the quality of evidence. The results of applying the GRADE system to clinical trial data are displayed in a table known as a GRADE profile.
Harms	Adverse effects of an intervention.
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 1 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 1 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)

Term	Definition
	threshold. If the threshold is £20,000 per QALY gained then the INB is calculated as $(£20,000 \times \text{QALYs gained}) - \text{Incremental cost}$.
Indirectness	The available evidence is different to the review question being addressed, in terms of PICO (population, intervention, comparison and outcome).
Intention-to-treat analysis (ITT)	An assessment of the people taking part in a clinical trial, based on the group they were initially (and randomly) allocated to. This is regardless of whether or not they dropped out, fully complied with the treatment or switched to an alternative treatment. Intention-to-treat analyses are often used to assess clinical effectiveness because they mirror actual practice: that is, not everyone complies with treatment and the treatment people receive may be changed according to how they respond to it.
Intervention	In medical terms this could be a drug treatment, surgical procedure, diagnostic or psychological therapy. Examples of public health interventions could include action to help someone to be physically active or to eat a more healthy diet.
Intraoperative	The period of time during a surgical procedure.
Kappa statistic	A statistical measure of inter-rater agreement that takes into account the agreement occurring by chance.
Length of stay	The total number of days a participant stays in hospital.
Licence	See 'Product licence'.
Life years gained	Mean average years of life gained per person as a result of the intervention compared with an alternative intervention.
Likelihood ratio	The likelihood ratio combines information about the sensitivity and specificity. It tells you how much a positive or negative result changes the likelihood that a patient would have the disease. The likelihood ratio of a positive test result (LR+) is sensitivity divided by (1 minus specificity).
Long-term care	Residential care in a home that may include skilled nursing care and help with everyday activities. This includes nursing homes and residential homes.
Logistic regression or Logit model	In statistics, logistic regression is a type of analysis used for predicting the outcome of a binary dependent variable based on 1 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.
Negative predictive value (NPV)	In screening or diagnostic tests: A measure of the usefulness of a screening or diagnostic test. It is the proportion of those with a negative test result who do not have the disease, and can be interpreted as the probability that a negative test result is correct. It is calculated as follows: $\text{TN}/(\text{TN}+\text{FN})$
Net monetary benefit (NMB)	The value in monetary terms of an intervention net of its cost. The NMB can be calculated for a given cost-effectiveness threshold. If the

Term	Definition
	<p>threshold is £20,000 per QALY gained then the NMB for an intervention is calculated as $(£20,000 \times \text{mean QALYs}) - \text{mean cost}$. 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.</p>
Non-randomised intervention study	<p>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.</p> <p>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.</p>
Number needed to treat (NNT)	<p>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.</p> <p>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.</p>
Observational study	<p>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 1 characteristic (for example, whether or not people received a specific treatment or intervention) are studied without intervening.</p> <p>There is a greater risk of selection bias than in experimental studies.</p>
Odds ratio	<p>Odds are a way to represent how likely it is that something will happen (the probability). An odds ratio compares the probability of something in 1 group with the probability of the same thing in another.</p> <p>An odds ratio of 1 between 2 groups would show that the probability of the event (for example a person developing a disease, or a treatment working) is the same for both. An odds ratio greater than 1 means the event is more likely in the first group. An odds ratio less than 1 means that the event is less likely in the first group.</p> <p>Sometimes probability can be compared across more than 2 groups – in this case, 1 of the groups is chosen as the 'reference category', and the odds ratio is calculated for each group compared with the reference category. For example, to compare the risk of dying from lung cancer for non-smokers, occasional smokers and regular smokers, non-smokers could be used as the reference category. Odds ratios would be worked out for occasional smokers compared with non-smokers and for regular smokers compared with non-smokers. See also confidence interval, risk ratio.</p>
Opportunity cost	<p>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.</p>
Outcome	<p>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</p>

Term	Definition
	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 1 seems more effective than the other, the p value is the probability of obtaining these results by chance. By convention, if the p value is below 0.05 (that is, there is less than a 5% probability that the results occurred by chance) it is considered that there probably is a real difference between treatments. If the p value is 0.001 or less (less than a 1% probability that the results occurred by chance), the result is seen as highly significant. If the p value shows that there is likely to be a difference between treatments, the confidence interval describes how big the difference in effect might be.
PICO	A PICO framework is where population (P) terms were combined with Intervention (I) and in some cases Comparison (C) terms. Outcomes (O) are rarely used in search strategies for interventions as these concepts may not be well described in title, abstract or indexes and therefore difficult to retrieve.
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).
Positive predictive value (PPV)	In screening or diagnostic tests: A measure of the usefulness of a screening or diagnostic test. It is the proportion of those with a positive test result who have the disease, and can be interpreted as the probability that a positive test result is correct. It is calculated as follows: $TP/(TP+FP)$
Power (statistical)	The ability to demonstrate an association when 1 exists. Power is related to sample size; the larger the sample size, the greater the power and the lower the risk that a possible association could be missed.
Preoperative	The period before surgery commences.
Pre-test probability	In diagnostic tests: The proportion of people with the target disorder in the population at risk at a specific time point or time interval. Prevalence may depend on how a disorder is diagnosed.
Prevalence	See Pre-test probability.
Primary care	Healthcare delivered outside hospitals. It includes a range of services provided by GPs, nurses, health visitors, midwives and other healthcare professionals and allied health professionals such as dentists, pharmacists and opticians.
Primary outcome	The outcome of greatest importance, usually the 1 in a study that the power calculation is based on.

Term	Definition
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'.
Receiver operated characteristic (ROC) curve	A graphical method of assessing the accuracy of a diagnostic test. Sensitivity is plotted against 1 minus specificity. A perfect test will have a positive, vertical linear slope starting at the origin. A good test will be somewhere close to this ideal.
Reference standard	The test that is considered to be the best available method to establish the presence or absence of the outcome – this may not be the 1 that is routinely used in practice.
Reporting bias	See 'Publication bias'.
Resource implication	The likely impact in terms of finance, workforce or other NHS resources.
Retrospective study	A research study that focuses on the past and present. The study examines past exposure to suspected risk factors for the disease or condition. Unlike prospective studies, it does not cover events that occur after the study group is selected.

Term	Definition
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)	<p>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).</p> <p>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.</p>
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	<p>Selection bias occurs if:</p> <ul style="list-style-type: none"> 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	<p>How well a test detects the thing it is testing for.</p> <p>If a diagnostic test for a disease has high sensitivity, it is likely to pick up all cases of the disease in people who have it (that is, give a 'true positive' result). But if a test is too sensitive it will sometimes also give a positive result in people who don't have the disease (that is, give a 'false positive').</p> <p>For example, if a test were developed to detect if a woman is 6 months pregnant, a very sensitive test would detect everyone who was 6 months pregnant, but would probably also include those who are 5 and 7 months pregnant.</p> <p>If the same test were more specific (sometimes referred to as having higher specificity), it would detect only those who are 6 months pregnant, and someone who was 5 months pregnant would get a negative result (a 'true negative'). But it would probably also miss some people who were 6 months pregnant (that is, give a 'false negative').</p> <p>Breast screening is a 'real-life' example. The number of women who are recalled for a second breast screening test is relatively high because the test is very sensitive. If it were made more specific, people who don't have the disease would be less likely to be called back for a second test but more women who have the disease would be missed.</p>
Sensitivity analysis	<p>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.</p> <p>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.</p> <p>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.</p> <p>Threshold sensitivity analysis: the critical value of parameters above or below which the conclusions of the study will change are identified.</p>

Term	Definition
	Probabilistic sensitivity analysis: probability distributions are assigned to the uncertain parameters and are incorporated into evaluation models based on decision analytical techniques (for example, Monte Carlo simulation).
Significance (statistical)	A result is deemed statistically significant if the probability of the result occurring by chance is less than 1 in 20 ($p < 0.05$).
Specificity	The proportion of true negatives that are correctly identified as such. For example in diagnostic testing the specificity is the proportion of non-cases correctly diagnosed as non-cases. See related term 'Sensitivity'. In terms of literature searching a highly specific search is generally narrow and aimed at picking up the key papers in a field and avoiding a wide range of papers.
Stakeholder	An organisation with an interest in a topic that NICE is developing a guideline or piece of public health guidance on. Organisations that register as stakeholders can comment on the draft scope and the draft guidance. Stakeholders may be: <ul style="list-style-type: none"> • manufacturers of drugs or equipment • national patient and carer organisations • NHS organisations • organisations representing healthcare professionals.
State transition model	See Markov model
Systematic review	A review in which evidence from scientific studies has been identified, appraised and synthesised in a methodical way according to predetermined criteria. It may include a meta-analysis.
Time horizon	The time span over which costs and health outcomes are considered in a decision analysis or economic evaluation.
Transition probability	In a state transition model (Markov model), this is the probability of moving from 1 health state to another over a specific period of time.
Treatment allocation	Assigning a participant to a particular arm of a trial.
Univariate	Analysis that 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

1. GRADE Working Group. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group website. 2011. Available from: <http://www.gradeworkinggroup.org/> Last accessed: 03/01/2019
2. National Clinical Guideline Centre. Hypertension: the clinical management of primary hypertension in adults: update of clinical guidelines 18 and 34. NICE clinical guideline 127. London. National Clinical Guideline Centre, 2011. Available from: <http://guidance.nice.org.uk/CG127>
3. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual. London. National Institute for Health and Care Excellence, 2014. Available from: <http://www.nice.org.uk/article/PMG20/chapter/1%20Introduction%20and%20overview>
4. National Institute for Health and Clinical Excellence. Social value judgements: principles for the development of NICE guidance. London. National Institute for Health and Clinical Excellence, 2008. Available from: <https://www.nice.org.uk/media/default/about/what-we-do/research-and-development/social-value-judgements-principles-for-the-development-of-nice-guidance.pdf>
5. National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal 2013. London. National Institute for Health and Clinical Excellence, 2013. Available from: <http://publications.nice.org.uk/pmg9>
6. Novielli N, Cooper NJ, Abrams KR, Sutton AJ. How is evidence on test performance synthesized for economic decision models of diagnostic tests? A systematic appraisal of Health Technology Assessments in the UK since 1997. *Value in Health*. 2010; 13(8):952-957
7. Review Manager (RevMan) [Computer program]. Version 5. Copenhagen. The Nordic Cochrane Centre, The Cochrane Collaboration, 2015. Available from: <http://tech.cochrane.org/Revman>
8. Ware J, Kosinski M, Gandek B. SF-36 Health Survey: Manual & Interpretation Guide. Boston, Massachusetts. 1993.
9. WinBUGS [Computer programme] version 1.4. Cambridge. MRC Biostatistics Unit University of Cambridge, 2015. Available from: <http://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/>