National Institute for Health and Care Excellence

Final

Atrial fibrillation: diagnosis and management

NICE guideline: methods

NICE guideline NG196

Methods

April 2021

Final

Developed by the National Guideline Centre, Royal College of Physicians

Disclaimer

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

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

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 All NICE guidance is subject to regular review and may be updated or withdrawn.

ISBN: 978-1-4731-4043-1

Contents

1 Development of the guideline

1.1 What is a NICE guideline?

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

NICE guidelines can:

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

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

We produce our guidelines using the following steps:

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

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

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

NICE Pathways brings together all connected NICE guidance.

1.2 Remit

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

The remit for this guideline is a partial update of CG180 with a full scoping process. The topic areas to be updated are:

- Diagnosis and assessment - Identification and assessment: presenting symptoms/pulse palpitation
- Assessment of stroke and bleeding risks
- Interventions to prevent stroke
- Rate and rhythm control
- Prevention and management of postoperative atrial fibrillation

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 Dr Simon Mackenzie in accordance with guidance from NICE.

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

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

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

1.3.1 What this guideline covers

This guideline covers the diagnosis and treatment of adults with atrial fibrillation. This includes identification and assessment of presenting symptoms/pulse palpitation; assessment of stroke and bleeding risks; interventions to prevent stroke; rate and rhythm control; prevention and management of postoperative atrial fibrillation; and prevention of recurrence of AF. 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 populations with congenital heart disease precipitating AF. This guideline update does not include some areas of the previous AF guideline, because these were identified by the NICE surveillance report as not requiring an update. These areas include personalised package of care and information; referral for specialised management; management of people presenting acutely with atrial fibrillation; rate versus rhythm control; and initial management of stroke and atrial fibrillation. Optimal combination of antiplatelet and anticoagulant therapies for people who have had an acute coronary syndrome (ACS) and have an indication for anticoagulation was not covered as this is

covered by other guidance. Finally, percutaneous atrial appendage occlusion was not included as it is no longer in routine use in the NHS

1.3.3 Relationships between the guideline and other NICE guidance

NICE technology appraisals to be incorporated in this guidance:

- Edoxaban for preventing stroke and systemic embolism in people with non-valvular atrial fibrillation (2015) NICE technology appraisal TA355
- Rivaroxaban for the prevention of stroke and systemic embolism in people with atrial fibrillation (2012) NICE technology appraisal guidance TA256
- Apixaban for preventing stroke and systemic embolism in people with nonvalvular atrial fibrillation (2012) NICE technology appraisal TA275
- Dabigatran etexilate for the prevention of stroke and systemic embolism in atrial fibrillation (2012) NICE technology appraisal TA249

Related NICE technology appraisals:

- Percutaneous radiofrequency ablation for atrial fibrillation (2006) NICE interventional procedure IP168
- WatchBP Home A for opportunistically detecting atrial fibrillation during diagnosis and monitoring of hypertension (2013) NICE medical technologies guidance 13

Related NICE interventional procedures guidance:

- Percutaneous balloon cryoablation for pulmonary vein isolation in atrial fibrillation (2012) NICE interventional procedure guidance IPG427
- Thoracoscopic exclusion of the left atrial appendage in atrial fibrillation (with or without other cardiac surgery) for the prevention of thromboembolism (2011) NICE interventional procedure guidance IPG400
- Percutaneous endoscopic catheter laser balloon pulmonary vein isolation for atrial fibrillation (2011) NICE interventional procedure guidance IPG399
- Percutaneous occlusion of the left atrial appendage in non-valvular atrial fibrillation for the prevention of thromboembolism (2010) NICE interventional procedure guidance IPG349
- Percutaneous (non-thorascopic) epicardial catheter radiofrequency ablation for atrial fibrillation (2009) NICE interventional procedure guidance IPG286

Related NICE guidelines:

- Venous thromboembolism in over 16s: reducing the risk of hospital-acquired deep vein thrombosis or pulmonary embolism (2018) NICE guideline NG89
- Alcohol-use disorders: diagnosis and management of physical complications (2017). NICE guideline CG100
- Multimorbidity: clinical assessment and management (2016) NICE guideline NG56
- Medicines adherence: involving patients in decisions about prescribed medicines and supporting adherence (2009) NICE guideline CG76
- Chronic heart failure (2018). NICE guideline NG106
- Lead-I electrocardiogram (ECG) devices for detecting atrial fibrillation using single-time point testing in primary care (2019). Diagnostic guidance DG35
- Hypertension (2019). NICE guideline NG 136.

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.[4]

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; and using a framework of population, setting and context for qualitative reviews.

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

A total of 13 review questions were identified.

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

Table 1: Review questions

Evidence report	Type of review	Review questions	Outcomes
A	Diagnostic RCT	What is the most clinically and cost-effective method for detecting atrial fibrillation in people with cardiovascular risk factors for AF and/or symptoms suggestive of AF?	Critical outcomes • Quality of life • Mortality • Stroke and thromboembolism • Major bleeding • All cause hospitalisation • Confirmed diagnosis of AF • Initiated anticoagulants for AF
B	Diagnostic accuracy	What are the most accurate methods for detecting atrial fibrillation in people with cardiovascular risk factors for AF and/or symptoms suggestive of AF?	Diagnostic accuracy outcomes (sensitivity and specificity)
C	Prognostic RCT	What is the most clinically and cost-effective risk stratification tool for predicting stroke or thromboembolic events in people with atrial fibrillation?	Critical outcomes • health-related quality of life • mortality • stroke or thromboembolic complications • major bleeding
D	Prognostic accuracy	What is the most accurate risk stratification tool for predicting stroke or thromboembolic events in people with atrial fibrillation?	• Simple diagnostic (prognostic) accuracy outcomes, such as sensitivity and specificity • C statistic (based on sensitivity and specificity but useful if >1 threshold used). • Calibration outcomes • Reclassification
E	Prognostic RCT	What is the most clinically and cost-effective risk stratification tool for predicting bleeding in people with atrial fibrillation?	Critical outcomes • health-related quality of life • mortality • stroke or thromboembolic complications

Evidence report	Type of review	Review questions	Outcomes
			• major bleeding
F	Prognostic accuracy	What is the most accurate risk stratification tool for predicting bleeding events in people with atrial fibrillation?	• Simple diagnostic (prognostic) accuracy outcomes, such as sensitivity and specificity • C statistic (based on sensitivity and specificity but useful if >1 threshold used). • Calibration outcomes • Reclassification
G	Intervention	What is the most clinically and cost-effective anticoagulant therapy for stroke prevention in people with atrial fibrillation?	Critical outcomes: • Quality of life • All stroke or systemic embolism • All-cause mortality • Myocardial infarction • Clinically relevant non-major bleeding • Minor bleeding • Major bleeding • Intracranial bleeding (ICH) • GI bleeding
H	Intervention	What is the clinical and cost-effectiveness of discontinuing anticoagulation in people whose atrial fibrillation has resolved?	Critical outcomes • health-related quality of life • mortality • stroke or thromboembolic complications • major bleeding • recurrent atrial fibrillation • Exacerbation of heart failure.
I	Intervention	What is the clinical and cost effectiveness of different non-ablative rate control therapies in people with atrial fibrillation?	Critical outcomes • health-related quality of life • mortality • hospitalisation • HF/exacerbation of heart failure. • • Failure of non-ablative rate control
J1	Intervention	What is the clinical and cost effectiveness of different ablative therapies in people with atrial fibrillation?	Critical outcomes • health-related quality of life • stroke or systemic embolism • mortality • Recurrent symptomatic AF (post-blanking period) • hospitalisation with a primary diagnosis of atrial fibrillation

Evidence report	Type of review	Review questions	Outcomes
			• Redo of procedure (catheter/surgical) • HF/exacerbation of heart failure. • Serious AEs Important outcomes • Hospital length of stay
J2	NMA	What is the clinical and cost effectiveness of different ablative therapies in people with paroxysmal atrial fibrillation?	Critical outcomes • stroke or systemic embolism • mortality • Recurrent symptomatic or asymptomatic AF (post-blanking period) • Serious AEs
K	Intervention	What is the clinical and cost-effectiveness of short-term (<6 months) antiarrhythmic drugs following ablation for preventing recurrence of atrial fibrillation?	Critical outcomes • Health related quality of life • Mortality • Stroke or thromboembolic complications • Hospitalisation with a primary diagnosis of atrial arrhythmia • Cardioversion for AF Important outcomes • All cause hospitalisation • Study drug discontinuation • Repeat ablation procedure within 1 year • Any documented atrial arrhythmia
L	Intervention	What is the most clinical and cost effective treatment strategy (rate or rhythm control, or no treatment) for people with atrial fibrillation after cardiothoracic surgery?	Critical outcomes • health-related quality of life • mortality • stroke or thromboembolic complications • Need for rescue DC cardioversion • Rehospitalisation (all cause) • Rehospitalisation for AF • Achievement of sinus rhythm • Adverse events Important outcomes • freedom from anticoagulation • freedom from AAD use • Hospital length of stay

Evidence report	Type of review	Review questions	Outcomes
			• ICU length of stay
M	Intervention	What is the clinical and cost effectiveness of statins in the prevention of atrial fibrillation following cardiothoracic surgery?	Critical outcomes • AF post-surgery • health-related quality of life • mortality • stroke or thromboembolic complications • Hospital readmission Important outcomes • Hospital length of stay • ICU length of stay

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[4] (https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview). 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 10th September 2020. Papers published or added to databases after this date were not considered. If new evidence, falling outside of the timeframe for the guideline searches, is identified, for example in consultation comments received from stakeholders, the impact on the guideline will be considered, and any further action agreed between NGC and NICE staff with a quality assurance role.

Prior to running, search strategies were quality assured using a variety of approaches. Medline search strategies were checked by a second information specialist, using the PRESS checklist.[3] Searches were cross-checked with reference lists of highly relevant papers, searches in other systematic reviews were analysed, and committee members were requested to highlight additional studies.

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

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

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 pre-specified 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.[4] Interventional studies were critically appraised using Cochrane methods and GRADE, Prognostic studies were critically appraised using PROBAST, and diagnostic studies were critically appraised using QUADAS2.
- Extracted key information about interventional study methods and results using 'Evibase', NGC's purpose-built software. Evibase produces summary evidence tables, including critical appraisal ratings. Key information about non-interventional study methods and results was manually extracted onto standard evidence tables and critically appraised separately (evidence tables are included in an appendix to each of the evidence reports).
- Generated summaries of the evidence by outcome. Outcome data were combined, analysed and reported according to study design:
  - Randomised data were meta-analysed where appropriate and reported in GRADE profile tables.
  - Data from non-randomised studies were presented as a range of values in GRADE profile tables or meta-analysed if appropriate.
  - Prognostic data were meta-analysed where appropriate and reported in GRADE profile tables.
  - Diagnostic data studies were meta-analysed where appropriate or presented as a range of values in adapted GRADE profile tables
  - Qualitative data were synthesised across studies and presented as summary statements with accompanying GRADE CERQual ratings for each review finding.
- A sample of a minimum of 10% of the abstract lists of the first 3 sifts by new reviewers and those for complex review questions (for example, prognostic reviews) were double-sifted by a senior research fellow and any discrepancies were rectified. All of the evidence reviews were quality assured by a senior research fellow. This included checking:
  - papers were included or excluded appropriately
  - a sample of the data extractions
  - correct methods were used to synthesise data
  - a sample of the risk of bias assessments.

2.3.1 Inclusion and exclusion criteria

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

The key population inclusion criterion was:

- Adults with atrial fibrillation.

The key population exclusion criterion was:

- Adults with congenital heart disease precipitating AF

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.

2.3.2 Type of studies

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

For diagnostic review questions, diagnostic RCTs, cross-sectional studies and retrospective studies were included. For prognostic review questions, prospective and retrospective cohort studies were included. Case–control studies were not included.

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

2.3.3 Methods of combining clinical studies

2.3.3.1 Data synthesis for intervention reviews

Where possible, meta-analyses were conducted using Cochrane Review Manager (RevMan5)[10] software to combine the data given in all studies for each of the outcomes of interest for the review question.

For some questions initial stratification of the analysis was used prior to any meta-analysis, and this is documented in the individual review question protocols in each evidence report. When additional strata were used this led to substrata (for example, using 2 stratification criteria leads to 4 substrata, using 3 stratification criteria leads to 9 substrata) which were analysed separately.

Such initial stratification of analysis was performed unconditionally (it was not dependent on statistical heterogeneity) because the committee felt *a priori* that the stratification variable(s) would be highly likely to affect the outcome effect estimate. This stratification process is distinct to the 'sub-grouping' process, which was a further stratification (within each initial stratum or overall analysis) that was conditional on statistical heterogeneity in meta-analyses. The 'sub-grouping' process is explained in section 2.3.3.1.3.

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 (non-exhaustive list):

- Stroke or thromboembolic events
- Major bleeding
- Clinically relevant non major bleeding
- Intracranial bleeding
- Gastrointestinal bleeding

- Mortality
- All cause hospitalisation
- Recurrence of atrial fibrillation
- Exacerbation of heart failure
- Serious adverse events.

The absolute risk difference was also calculated using GRADEpro[2] software, using the median event rate in the control arm of the pooled results.

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

Where a meta-analysis contained one or more studies with zero events in *both* arms, risk differences were calculated to allow these studies to be included.

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)
- length of stay in hospital

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

The means and standard deviations of continuous outcomes are required for meta-analysis. However, in cases where standard deviations were not reported, the standard error was calculated if the p values or 95% confidence intervals (95% CI) were reported, and meta-analysis was undertaken with the mean and standard error using the generic inverse variance method in Cochrane Review Manager (RevMan5[10] software. Where p values were reported as 'less than', a conservative approach was undertaken. For example, if a p value was reported as 'p≤0.001', the calculations for standard deviations were based on a p value of 0.001. If these statistical measures were not available then the methods described in section 16.1.3 of the Cochrane Handbook (version 5.1.0, updated March 2011) were applied.

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. [10] If the control event rate was reported this was used to generate the absolute risk difference in GRADEpro.[2] 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 and sub-grouping

For each meta-analysis estimate, statistical heterogeneity was assessed by considering the I-squared (I^2) inconsistency statistic. This was carried out within each stratum, or, if no strata were used, within the overall un-stratified analysis. An I^2>50% was taken to indicate significant heterogeneity. Where significant heterogeneity was present, predefined subgrouping of studies was carried out, for variables that were distinct from those previously used for stratification (if any). These sub-grouping variables were variables that the committee felt might affect the outcome, but where confidence in these effects was not

unequivocal. This is why they were not used as unconditional stratification variables but instead variables that would only be used conditionally if heterogeneity were observed.

The sub-grouping strategies varied between questions, but an example would be the type of oral anticoagulant used in the 'prediction of bleeding' prognostic accuracy question. In this example, if heterogeneity between study 'C statistics' was observed in the overall meta-analysis, studies were sub-grouped into the oral anticoagulant the majority of participants in each study were using: vitamin K antagonists, NOACs, 'mixed (no clear majority)' or 'unclear'. Details of the sub-grouping strategies used are outlined in the protocols for each question.

If the subgroup analysis resolved heterogeneity within *all* of the derived subgroups (I^2 < 50%), then each of the derived subgroups were subsumed into separate outcomes (providing at least 1 study remained in each subgroup). For example, consider an initial meta-analysis of 6 studies for the outcome of '*serious adverse events'*. Initially the I^2 value is 65%, but after separating out the 3 studies where participants had CHADSVASC <2 and the 3 studies where participants had CHADSVASC of 2 or more, I^2 was reduced to <50% in each of these sub-groups. In such a case, instead of the single outcome of '*serious adverse events'*, this would be separated into 2 outcomes '*serious adverse events in people with CHADSVASC <2'* and '*serious adverse events in people without CHADSVASC of 2 or more*.

Such resolution of heterogeneity through sub-grouping was interpreted with caution, as although an association between reductions of variability of CHADSVASC score (through sub-grouping) and reductions in the variability of effect estimates do suggest a causal effect this cannot be definitely assumed. This is because of the possibility that confounding variables other than CHADSVASC score are the true driver of the variability in effect estimates, and that the associations with CHADSVASC are simply correlative.

For some questions more than one subgrouping strategy was applied (for example, as well as the CHADSVASC categories, sub-grouping by 'heart failure/no heart failure' might also be used) and this is documented in the individual review question protocols. These additional subgrouping strategies were applied independently, so subunits of subgroups were not created, unlike the situation with strata.

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 point estimate, reflecting the greater uncertainty inherent in estimating the mean of a distribution of population means rather than a single population mean. If, however, the committee considered the heterogeneity was so large that meta-analysis was inappropriate, then the results were described narratively.

2.3.3.1.4 Complex analysis

Where studies had used a crossover design, paired continuous data were extracted where possible, and forest plots were generated in RevMan5[10] with the generic inverse variance function. When a crossover study had categorical data and the number of subjects with an event in both interventions was known, the standard error (of the log of the risk ratio) was calculated using the simplified Mantel–Haenszel method for paired outcomes. Forest plots were also generated in RevMan5[10] with the generic inverse variance function. If paired continuous or categorical data were not available from the crossover studies, the separate group data were analysed in the same way as data from parallel groups, on the basis that this approach would overestimate the confidence intervals and thus artificially reduce study weighting resulting in a conservative effect. Where a meta-analysis included a mixture of studies using both paired and parallel group approaches, all data were entered into RevMan5[10] using the generic inverse variance function.

2.3.3.2 Network meta-analysis

A network meta-analysis (NMA) was conducted for the question '*What is the clinical and cost effectiveness of different non-ablative rate control therapies in people with atrial fibrillation?'*

The methods for the NMA are presented in Chapter J2, as these methods are very specific to that chapter.

2.3.3.3 Data synthesis for diagnostic test reviews

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

2.3.3.3.1 Diagnostic RCTs

Diagnostic RCTs (sometimes referred to as test and treat trials) are a randomised comparison of 2 or more diagnostic tests or diagnostic strategies, with study outcomes being clinically important consequences of the diagnosis (patient-related outcome measures similar to those in intervention trials, such as mortality). Patients are randomised (for example) 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).

2.3.3.3.2 Diagnostic accuracy studies

Diagnostic accuracy studies measure how well a test can detect those people who truly have the condition, and also how well the test can detect those people who truly do not have the condition. The true existence of the condition is determined by a gold standard test, which is regarded as infallible. A two by two table (Figure 1) contains all the information required to calculate diagnostic accuracy, with the data being counts of people, and all cells being mutually exclusive and exhaustive. The two columns carry information about the gold standard results and the two rows contain information about the test under investigation (the index test).

	Gold standard positive = truly have the condition	**Gold standard negative = truly do NOT have the condition**
Index test positive	98	22
Index test negative	2	178
Total	100	200

Figure 2: A two by two table for diagnostic accuracy

In the example above there are 100 people defined by the gold standard as truly having the condition. Of these, 98 are correctly identified as having the condition by the index test (positive index test), so the sensitivity of the index test is 98/100 = 98%. There are also 200 people defined by the gold standard as truly NOT having the condition. Of these, 178 are correctly identified as not having the condition by the index test (negative index test), so the specificity of the index test is 178/200 = 89%.

In many diagnostic tests, the index test is based on a continuous measurement, and the test is designated positive if the test result is beyond a specific threshold on that continuous scale. The position of this threshold can be varied, and as the threshold changes there is a trade-off between sensitivity and specificity. Assuming that higher values of the measurement are associated with the condition, a low threshold will tend to lead to more people testing positive because detection is triggered by all values *above* that threshold. A low threshold will thus have greater sensitivity, but because it may also tend to pick up people who don't have the condition it will lead to a lower specificity. In contrast, a high threshold may miss people who truly have the condition, because it won't detect people with the condition who have a value below that high threshold. A higher threshold will therefore have lower sensitivity, but will tend to pick out those who don't have the condition and so will have a high specificity. Plotting the sensitivities and specificities across these different thresholds yields the receiver operated characteristics (ROC) curve if specificity is plotted as 1-specificity, and the area under this curve provides an overall measure of accuracy over all thresholds. For this guideline, where the diagnostic accuracy study concerned the detection of atrial fibrillation, the thresholds (such as a particular level of inter-beat variability) tended to be fixed and multiple thresholds were not used. Therefore only sensitivity and specificity at the fixed threshold were used, rather than ROC curves. If a test did use different thresholds these were treated as separate tests.

For this guideline, where the diagnostic accuracy review concerned the detection of atrial fibrillation, sensitivity was considered more important than specificity. This was due to the consequences of failing to detect AF (in someone who truly has it) being considered worse than misdiagnosing someone as having AF (when in reality they don't). The most important consequences of failing to diagnose someone with AF is stroke. This is of greater probability and likely to be of greater severity than the consequences of misdiagnosing someone as having AF, which may include bleeding resulting from unnecessary use of anticoagulants.

Coupled forest plots of sensitivity and specificity with their 95% CIs across studies were produced for each test, using RevMan5. [10] In order to do this, 2×2 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. 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.[15] The advantage of this approach is that it produces summary estimates of sensitivity and specificity that account for the correlation between the 2 statistics. Other advantages of this method have been described elsewhere.[9, 12, 13] 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 2010.[7]) Pooled sensitivity and specificity and their 95% CIs were reported in the clinical evidence summary tables.

For scores with fewer than 3 studies, median sensitivity and the paired specificity were reported where possible. If an even number of studies were reported the results of the study with the lower sensitivity value of the 2 middle studies was reported.

2.3.3.4 Data synthesis for prognostic test accuracy reviews (also called 'risk prediction tools' or 'risk prediction rules').

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

2.3.3.4.1 Prognostic RCTs

Prognostic RCTs are a randomised comparison of 2 or more prognostic tests, and are identical in principle to diagnostic RCTs. In such studies, health-related outcomes are consequences of the accuracy of prediction. For example, if the predictive test can accurately predict that someone is likely to have a stroke without anticoagulation, then that patient will benefit from the initiation of anticoagulation treatment informed by the test result. Likewise, if the prognostic test can accurately predict that someone is very unlikely to have a stroke then that patient will probably benefit from not being unnecessarily treated with anticoagulants, which could cause harm. Patients are randomised (for example) to receive test A or test B, followed by identical therapeutic interventions based on the results of the prognostic 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 for the same indication is identical in both arms of the trial, any differences in patient outcomes will reflect the accuracy of the tests in correctly establishing who will get and who will not get the condition. Data were synthesised using the same methods for intervention reviews (see section 2.3.3.1.1 above).

2.3.3.4.2 Prognostic accuracy studies

A prognostic test aims to accurately determine who will, and who will not, attain a particular prognostic outcome (for example, stroke) in the future. This is analogous to a diagnostic test, which aims to accurately determine who has, and who has not, a particular disease. The difference between them is that whilst the diagnostic test measures the accuracy of detecting a current condition, the prognostic test measures the accuracy of predicting a later event (determining who actually gets the outcome or not). Therefore, while the gold standard for diagnostic tests is the best available method of diagnosis, the gold standard for prognostic tests is always the later measurement of the outcome. In the review for detection of stroke risk, the later outcome was stroke or thromboembolic events. For the review for detection of bleeding risk, the later outcome was major bleeding, or other definitions of bleeding.

C statistics

In this guideline, the accuracy of different prediction tools were analysed at a variety of test thresholds within each study, and so areas under the ROC curve (AUC or ‘C statistic’) were useful measures of overall accuracy (see section 2.3.3.3.2). The C statistic describes the overall diagnostic accuracy across the full range of thresholds. The following criteria were used for evaluating C statistics:

- ≤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.

C statistics across different studies were meta-analysed using the generic inverse variance option (for continuous variables) on RevMan. The derived forest plots were amended using the ‘paint’ program so that the null line was removed. Unlike the measures of effect in most meta-analyses, C statistics are not measures representing the differences or ratios between two groups, and are instead a single group value (although the ultimate frame of reference is the gold standard). A null line indicating that there is ‘no difference between groups’ therefore has little meaning in this context.

Sensitivity and specificity

Sensitivity and specificity data were also collected for specific thresholds where available in the papers. This was necessary as prediction tools will be used clinically with specific thresholds, and so knowledge of accuracy at these specific thresholds is vital.

Coupled forest plots of sensitivity and specificity with their 95% CIs across studies were produced for each test, using RevMan5.[10] In order to do this, 2×2 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. 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.[15] The advantage of this approach is that it produces summary estimates of sensitivity and specificity that account for the correlation between the 2 statistics. Other advantages of this method have been described elsewhere.[9, 12, 13] 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 2010.[7]) Pooled sensitivity and specificity and their 95% CIs were reported in the clinical evidence summary tables.

For scores with fewer than 3 studies, median sensitivity and the paired specificity were reported where possible. If an even number of studies were reported the results of the study with the lower sensitivity value of the 2 middle studies was reported.

Net reclassification improvement (NRI)

Net reclassification improvement (NRI) was also used to evaluate prognostic accuracy data. This is expressed in terms of one (index) risk tool to another (comparator) risk tool, and gives a score between -2 and +2 (with +2 representing the best possible performance of the index tool relative to the comparator, and -2 the worst). The score represents the net improvement of the index test relative to the comparator in terms of the proportion of true cases (judged by later development of stroke/TE) that are correctly up-classified by the tool (relative to any false negative classifications yielded by the comparator), and the proportion of false cases (judged by the lack of later stroke/TE) that are correctly down-classified by the tool (relative to any false positive classifications yielded by the comparator). Meanwhile, incorrect up-classification or incorrect down-classification of the index relative to the comparator convey negative scores to the NRI, and so if a score is negative overall this indicates the index is less accurate than the comparator.

NRI data for the prognostic reviews were meta-analysed where possible, using the standard continuous outcomes facility on RevMan.

For NRI data, any NRI value with 95% confidence intervals not crossing the null line were regarded as of clinical importance.

D statistics

The D statistic is a measure of discrimination, where better discrimination is shown by higher D values. The value can be interpreted as the average log hazard ratio between an individual in the upper half of the risk distribution and an individual in the lower half[14]. No sources were found that provide clinically important thresholds. However, based on the fact that a hazard ratio of 3 could be regarded as a clinically useful ratio of the hazard of the outcome event across the two risk categories, and the natural logarithm of this is 1.1, a D value of 1.1 was taken as representing a clinically important level of discrimination.

Calibration

Measures of calibration assess the ability of a risk prediction model to predict accurately the absolute level of risk that is subsequently observed. Calibration concerns how well the predicted risks compare to observed risks. A model is well calibrated if, for every 100 patients given a prediction of p%, the observed number of events is close to p. Calibration is evaluated either by calculating the Hosmer-Lemeshow test statistic, or preferably by plotting predicted risks against observed risks (calibration plot). This involves predicted outcome probabilities (on the x-axis) plotted against observed outcome frequencies (on the y-axis). A well-calibrated model shows predictions lying on or around the 45° line of the calibration plot; perfect calibration shows a slope of 1 and intercept of 0, although some caveats have recently been identified. Other informative measures of model performance include the R^2 and the Brier score. R^2 characterizes the degree of variation in risk explained by the model. The R2 threshold taken as representing clinical importance was 0.5. Although no sources were available to confirm the appropriateness of this figure, the choice of 0.5 was based on the fact that if a tool could not explain at least half of the variability in outcome it would have limited clinical utility.

The adjusted R2 has been proposed as a better measure, as it accounts for the number of predictors and helps to prevent overfitting. Brier scores are a similar measure of performance, which are used when the outcome of interest is categorical instead of continuous.

Calibration measures the accuracy of absolute risk prediction better than discrimination methods (such as C statistics or sensitivity/specificity). The absolute level of bleeding risk is what will be used clinically to allow the clinician and patient to make a shared decision on risk reduction through attention to modifiable risk factors for bleeding. Therefore calibration was regarded as a particularly important measure of effect for prediction of bleeding risk.

Calibration data were mostly synthesised using narrative methods, because data were often presented graphically. However where appropriate, data were meta-analysed.

2.3.4 Appraising the quality of evidence by outcomes

2.3.4.1 Intervention reviews (including diagnostic and prognostic RCT reviews)

The evidence for outcomes from the included RCTs and, where appropriate, non-randomised intervention studies, were evaluated and presented using an adaptation of the 'Grading of Recommendations Assessment, Development and Evaluation (GRADE) toolbox' developed by the international GRADE working group (http://www.gradeworkinggroup.org/). The software (GRADEpro[2]) developed by the GRADE working group was used to assess the quality of each outcome, taking into account individual study quality and the meta-analysis results.

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

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

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

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

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

2.3.4.1.1 Risk of bias

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

Table 3: Principle domains of bias in randomised controlled trials

Limitation	Explanation
Selection bias (sequence generation and allocation concealment)	If those enrolling patients are aware of the group to which the next enrolled patient will be allocated, either because of a non-random sequence that is predictable, or because a truly random sequence was not concealed from the researcher, this may translate into systematic selection bias. This may occur if the researcher chooses not to recruit a participant into that specific group because of: • knowledge of that participant's likely prognostic characteristics, and • a desire for one group to do better than the other.
Performance bias (lack of blinding of patients and healthcare professionals)	Patients and healthcare professionals or caregivers who are caring for the patient should not be aware of the arm to which patients are allocated. Knowledge of the group can influence: • the experience of the placebo effect • the level of care and attention received, Both of these can influence outcomes, which can contribute to systematic bias. This is equally true whether the outcome is subjective or objective. Even a highly objective outcome such as mortality may be affected by the prior levels of care and attention given at the treatment stage.

Limitation	Explanation
Attrition bias	Attrition bias results from an unaccounted for loss of data beyond a certain level (a differential of 10% between groups). Loss of data can occur when participants are compulsorily withdrawn from a group by the researchers (for example, when a per-protocol approach is used) or when participants do not attend assessment sessions. If the missing data are likely to be different from the data of those remaining in the groups, and there is a differential rate of such missing data from groups, systematic attrition bias may result.
Detection bias	Those adjudicating or recording outcomes should not be aware of the arm to which patients are allocated. Knowledge of the group can influence the methods of measurement or analysis, which can contribute to systematic bias. This is dependent on the subjectivity of an outcome. For highly objective outcomes, such as death or laboratory results, it is usually not possible for the adjudicator's knowledge of group allocation to affect the recorded outcome. However for subjective outcomes it is often quite easy for this knowledge to influence results.
Selective outcome reporting	Reporting of some outcomes and not others on the basis of the results can also lead to bias, as this may distort the overall impression of efficacy.
Other limitations	For example: • Stopping early for benefit observed in randomised trials, in particular in the absence of adequate stopping rules. • Use of unvalidated patient-reported outcome measures. • Lack of washout periods to avoid carry-over effects in crossover trials. • Recruitment bias in cluster-randomised trials.

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

2.3.4.1.2 Indirectness

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

2.3.4.1.3 Inconsistency

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

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

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

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

2.3.4.1.4 Imprecision

The criteria applied for imprecision were based on the 95% CIs for the pooled estimate of effect, and the minimal important differences (MID) for the outcome. The MIDs are the threshold for appreciable benefits and harms, 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 3. As for inconsistency, since the imprecision score was based on the meta-analysis results, the score represented the whole outcome and so weighted averaging across studies was not necessary.

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

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

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

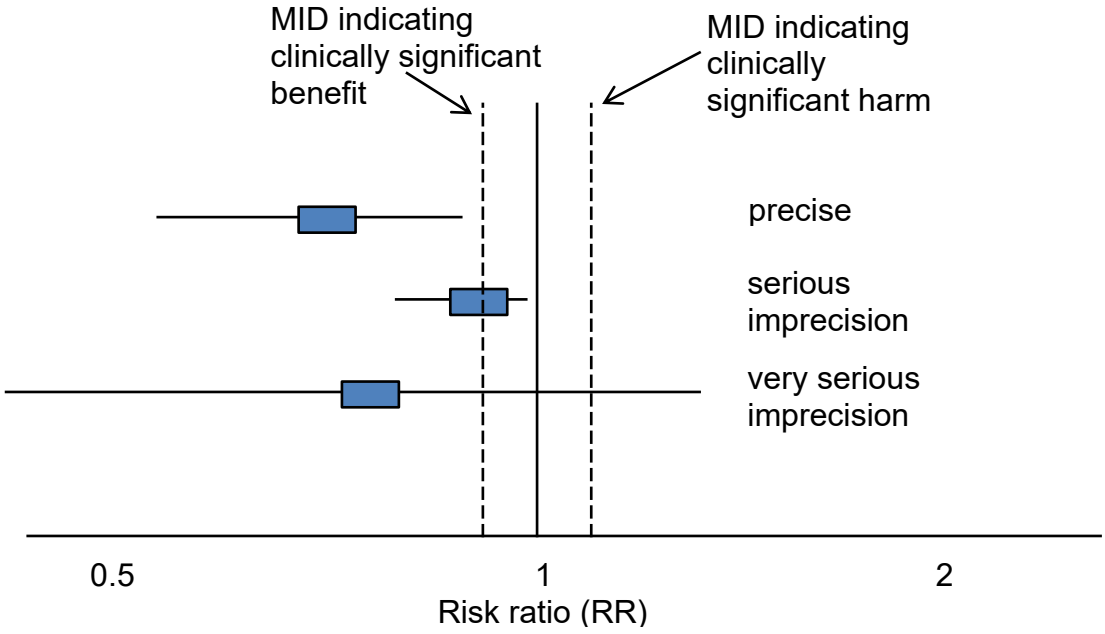
- For continuous outcome variables the MID was taken as half the median baseline standard deviation of that variable, across all studies in the meta-analysis. Hence the MID denoting the minimum clinically significant benefit was positive for a 'positive' outcome (for example, a quality of life measure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

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

For this guideline, no appropriate MIDs for continuous or dichotomous outcomes were found in the literature, and so the default method was adopted.

When risk differences were used as the measure of effect for meta-analysis (which was necessary if any studies had zero events in both arms) then imprecision was based on the optimum information size. This was calculated by a 'power analysis', assessing if the statistical power of the meta-analysis was <80% (very serious), 80-90% (serious), or >90% (not serious) using the following purpose-built calculator:
https://www.stat.ubc.ca/~rollin/stats/ssize/b2.html

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



2.3.4.1.5 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 accuracy studies

2.3.4.2.1 Risk of bias and indirectness

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[4]). Risk of bias and applicability in primary diagnostic accuracy studies in QUADAS-2 consists of 4 domains (see Figure 4):

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

Figure 4: 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 2×2 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?
	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	Are there concerns that the included patients	Are there concerns that the index test, its conduct, or	Are there concerns that the target condition as defined by the	

Domain	Patient selection	Index test	Reference standard	Flow and timing
(high/low/ unclear)	do not match the review question?	interpretation differ from the review question?	reference standard does not match the review question?	

2.3.4.2.2 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. If there were any studies with 95% CIs that did not overlap with any other, then a rating of serious inconsistency was given. For tools with only single studies no inconsistency rating was given.

2.3.4.2.3 Imprecision

The judgement of precision was based on the position of the 95% confidence intervals for sensitivity and specificity relative to two clinical thresholds at 0.60 and 0.90. The 0.60 threshold represented the threshold accuracy below which the tool would not be clinically useful, and the 0.90 threshold represented the threshold above which the tool might be recommended. Serious imprecision was recorded if the 95% CIs crossed one of these clinical thresholds, and very serious imprecision was recorded if the 95% CIs crossed both clinical thresholds.

If a meta-analysis was undertaken the 95% CIs of the summary sensitivity/specificity was used. If only 2 studies were available then the 95% CIs of the median sensitivity value and paired specificity value were used. If only 1 study was available the 95% CI of the single sensitivity and specificity values were used.

2.3.4.2.4 Overall grading

Quality rating started at High for prospective and retrospective cohort 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.

2.3.4.3 Prognostic accuracy studies

2.3.4.3.1 Risk of bias

Risk of bias was initially assessed per study using the PROBAST tool.

PROBAST criteria were as follows:

- Appropriateness of data sources?
- Appropriateness of inclusion and exclusion criteria?
- Appropriate similarity of health across participants?
- Were predictors defined or assessed in the same way for all?
- Predictor assessments made without knowledge of outcome data?
- Predictors all available at time model meant to be used?
- All relevant predictors analysed?
- Pre-specified outcome used?
- Predictors excluded from outcome definition?
- Outcome defined in same way for all?
- Outcome determined without knowledge of predictor information?
- Reasonable number of outcome events? (100)
- Time interval between baseline and outcome appropriate? (5 years)
- All enrolled included in analysis?

- Missing data handled appropriately?
- Non-binary predictors handled appropriately?
- Complexities in data accounted for?
- Relevant performance measures?
- Model recalibrated or likely that calibration not needed?

Possible responses were not applicable, unclear, yes or no.

For each study risk of bias was downgraded by 1 (serious risk of bias) if blinding of assessors was not reported, and/or attrition bias (>10% loss) was suspected. Risk of bias was downgraded by 2 (very serious risk of bias) if the studies with the aforementioned limitations also had insufficient numbers of events (<100) and/or inappropriately short follow up times (<5 years) to be able to accurately predict risk.

An overall risk of bias rating was then pooled across studies covering the same outcome, using the meta-analysis weighting.

2.3.4.3.2 Indirectness

Indirectness was assessed by the extent to which the population, index test or outcome differed from the protocol definition. Indirectness was planned to be downgraded by 1 (serious risk of indirectness) if there was one departure from protocol, or by 2 (very serious risk of indirectness) if there were two or more departures from protocol. However no studies were downgraded for indirectness.

2.3.4.3.3 Inconsistency

Where data were pooled, an I^2 of 50-74% was deemed serious inconsistency and an I^2 of 75% or above was deemed very serious inconsistency. If no pooling were possible, inconsistency was assessed by inspection of the degree of overlap of confidence intervals between studies: if one of more Cis did not overlap then a rating of serious inconsistency was given.

2.3.4.3.4 Imprecision

The judgement of precision was based on the spread of confidence intervals. For C statistic data, two clinical thresholds were used: AUCs of 0.5 and 0.7. The threshold of 0.5 marked the boundary between no predictive value better than chance and a predictive value better than chance. The threshold of 0.7 marked the boundary above which the committee might consider recommendations. If the 95% Cis crossed one of these thresholds a rating of serious imprecision was given and if they crossed both of these thresholds a rating of very serious imprecision as given.

The judgement of precision for sensitivity was based on the position of the 95% confidence intervals relative to two clinical thresholds at 0.60 and 0.90. The 0.60 threshold represented the threshold accuracy below which the tool would not be clinically useful, and the 0.90 threshold represented the threshold above which the tool might be recommended. Serious imprecision was recorded if the 95% CIs crossed one of these clinical thresholds, and very serious imprecision was recorded if the 95% CIs crossed both clinical thresholds. The judgement of precision for specificity was based on the position of the 95% confidence intervals relative to two clinical thresholds at 0.10 and 0.50. The 0.10 threshold represented the threshold accuracy below which the tool would not be clinically useful, and the 0.50 threshold represented the threshold above which the tool might be recommended. Serious imprecision was recorded if the 95% CIs crossed one of these clinical thresholds, and very serious imprecision was recorded if the 95% CIs crossed both clinical thresholds.

If a meta-analysis was undertaken the 95% CIs of the summary sensitivity/specificity was used. If only 2 studies were available then the 95% CIs of the median sensitivity value and

paired specificity value were used. If only 1 study was available the 95% CI of the single sensitivity and specificity values were used.

For the measure of 'D', precision was based on a clinically important threshold of 1.1. Therefore, if the CIs crossed 1.1 a rating of serious imprecision was made.

For the NRI data if either of the 95% CIs passed across 0 then this was graded as seriously imprecise.

For R^2 calibration data precision was based on a clinically important threshold of 0.5. Therefore, if the CIs crossed 0.5 a rating of serious imprecision was made.

2.3.4.3.5 Overall rating

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 (ARDs) using GRADEpro[2] 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 outcomes in the intervention reviews that if at least 100 more participants per 1000 (10%) achieved the outcome of interest in the intervention group compared to the comparison group for a positive outcome then this intervention was considered beneficial. The same point estimate but in the opposite direction applied for a negative outcome. For the critical outcome of mortality any reduction represented a clinical benefit. For adverse events 50 events or more per 1000 (5%) represented clinical harm. For continuous outcomes if the mean difference was greater than the minimally important difference (MID) then this represented a clinical benefit or harm. For outcomes such as mortality any reduction or increase was considered to be clinically important.

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

2.4 Identifying and analysing evidence of cost effectiveness

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

uncertainties must be offset by a compelling argument in favour of the recommendation. The cost impact or savings potential of a recommendation should not be the sole reason for the committee's decision.[4]

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

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

2.4.1 Literature review

The health economists:

- Identified potentially relevant studies for each review question from the health economic search results by reviewing titles and abstracts. Full papers were then obtained.
- Reviewed full papers against pre-specified inclusion and exclusion criteria to identify relevant studies (see below for details).
- Critically appraised relevant studies using economic evaluations checklists as specified in the NICE guidelines manual.[4]
- Extracted key information about the studies' methods and results into health economic evidence tables (which can be found in appendices to the relevant evidence reports).
- Generated summaries of the evidence in NICE health economic evidence profile tables (included in the relevant evidence report for each review question) – see below for details.

2.4.1.1 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 2003 and studies from non-OECD countries or the USA were also excluded, on the basis that the applicability of such studies to the present UK NHS context is likely to be too low for them to be helpful for decision-making.

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

For more details about the assessment of applicability and methodological quality see Table 5 below and the economic evaluation checklist (appendix H of the NICE guidelines manual[4]) and the health economics review protocol, which can be found in each of the evidence reports.

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

2.4.1.2 NICE health economic evidence profiles

NICE health economic evidence profile tables were used to summarise cost and cost-effectiveness estimates for the included health economic studies in each evidence review

report. The health economic evidence profile shows an assessment of applicability and methodological quality for each economic study, with footnotes indicating the reasons for the assessment. These assessments were made by the health economist using the economic evaluation checklist from the NICE guidelines manual.[4] It also shows the incremental costs, incremental effects (for example, quality-adjusted life years [QALYs]) and incremental cost-effectiveness ratio (ICER) for the base case analysis in the study, as well as information about the assessment of uncertainty in the analysis. See Table 5 for more details.

When a non-UK study was included in the profile, the results were converted into pounds sterling using the appropriate purchasing power parity.[8]

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

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

2.4.2 Undertaking new health economic analysis

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

The committee identified anticoagulant therapy and ablation techniques as the highest priority areas for original health economic modelling. The committee decided that identifying which anticoagulant was the most cost-effective was the highest priority for economic modelling on the account of the large number of patients affected by potential recommendations, the current variation in uptake of DOACs nationally and the likelihood there will be sufficient clinical effectiveness data to inform model parameters. The second priority was identifying whether ablation was cost effective compared to antiarrhythmic drugs and which type pf ablation was the most cost-effective. This question has a potentially significant resource impact due to the cost of ablation. Furthermore, there is a lack of health economic evidence comparing all interventions and on the long term cost effectiveness of these interventions.

Of note, the anticoagulation model was conducted by the NICE Technical Support Unit, as they were the authors of an existing economic model on anticoagulants (Sterne 2017[11]) and it was deemed appropriate for them to update this model for inclusion in the guideline.

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.[4, 5]
- The committee was involved in the design of the model, selection of inputs and interpretation of the results.
- Model inputs were based on the systematic review of the clinical literature supplemented with other published data sources where possible.
- When published data were not available committee expert opinion was used to populate the model.
- Model inputs and assumptions were reported fully and transparently.
- The results were subject to sensitivity analysis and limitations were discussed.
- The model was peer-reviewed: for the anticoagulation model the peer review was undertaken by the BMJ group and for the ablation model by another health economist at the NGC.

Full methods and results of the cost-effectiveness analyses for anticoagulation and ablation are described in the separate economic analysis reports.

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.[6] In general, an intervention was considered to be cost effective (given that the estimate was considered plausible) if either of the following criteria applied:

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

If the committee recommended an intervention that was estimated to cost more than £20,000 per QALY gained, or did not recommend one that was estimated to cost less than £20,000 per QALY gained, the reasons for this decision are discussed explicitly in 'The committee's discussion of the evidence' section of the relevant evidence report, with reference to issues regarding the plausibility of the estimate or to the factors set out in 'Social value judgements: principles for the development of NICE guidance'.[6]

When QALYs or life years gained are not used in the analysis, results are difficult to interpret unless one strategy dominates the others with respect to every relevant health outcome and cost.

2.4.4 In the absence of health economic evidence

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

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

2.5 Developing recommendations

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

- Summaries of clinical and health economic evidence and quality (as presented in evidence reports [A–M]).
- 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 and summary ROC curves (in appendices to the relevant evidence reports).
- A description of the methods and results of the cost-effectiveness analyses undertaken for the guideline (in a separate economic analysis report).

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

When clinical and health economic evidence was of poor quality, conflicting or absent, the committee drafted recommendations based on its expert opinion. The considerations for making consensus-based recommendations include the balance between potential harms and benefits, the economic costs compared to the economic benefits, current practices, recommendations made in other relevant guidelines, patient preferences and equality issues. The consensus recommendations were agreed through discussions in the committee. The committee also considered whether the uncertainty was sufficient to justify delaying making a recommendation to await further research, taking into account the potential harm of failing to make a clear recommendation (see section 2.5.1 below).

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

some patients are particularly averse to some side effect and others are not. In these circumstances the recommendation is generally weaker, although it may be possible to make stronger recommendations about specific groups of patients.

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

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

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

2.5.1 Research recommendations

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

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

2.5.2 Validation process

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

2.5.3 Updating the guideline

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

2.5.4 Disclaimer

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

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

2.5.5 Funding

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

3 Acronyms and abbreviations

Acronym or abbreviation	Description
AAD	Antiarrhythmic drugs
AF	Atrial fibrillation
CABG	Coronary artery bypass grafting
CI	Confidence intervals
CRNMB	Clinically relevant non-major bleeding
CVA	Cerebrovascular accident (stroke)
DC cardioversion	Direct current cardioversion
DOAC	Direct-acting oral anticoagulant
ECG	Electrocardiography
GC	Guideline Committee
GRADE	Grading of recommendations, assessment, development and evaluation.
HR	Hazard ratio
INR	International Normalised Ratio
K+	Potassium
LAAO	Left atrial appendage occlusion
LDL	Low density lipids
Na+	Sodium
NGC	National Guideline Centre
NICE	National Institute for Health and Care Excellence
NOAC	Novel oral anticoagulant
NRI	Net reclassification Index
NYHA	New York Heart Association
OR	Odds ratio
QALY	Quality adjusted life year
RR	Risk ratio (also known as relative risk)
TIA	Transient ischaemic Attack
TTR	Time in therapeutic range
VKA	Vitamin K antagonist

4 Glossary

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

4.1 Guideline-specific terms

Term	Definition
3-lead devices	Detection device that utilises 3 leads (see 12 lead ECG)
6-lead devices	Detection device that utilises 6 leads (see 12 lead ECG)
12-lead ECG	An electrocardiogram (ECG) is a graphical record of the direction and magnitude of the electrical activity generated by the heart. The 12-lead ECG is considered to be the gold standard for detecting persistent atrial fibrillation. The 'leads' refer to 12 different directions across the heart that are displayed. A 12 lead ECG therefore gives a very detailed view of the three dimensional flow of electrical charge through the heart. Only 10 cables, each attached to a skin electrode, are needed to create these 12 leads. Leads I, II, III, aVR, aVL, aVF are called the limb leads and look at the heart in a coronal plane from the left mid-arm, left leg, right leg, right shoulder, left shoulder and directly upwards from the feet respectively. Leads V1 to V6 are called the chest leads and look at the heart in the transverse plane, also in slightly different directions.
Ablation	Destruction, scarring or removal of body tissue, usually via a catheter procedure, but occasionally via open surgery. In the context of atrial fibrillation, ablation refers to the targeted destruction of sections of cardiac tissue in order to prevent the conduction abnormalities that contribute to cardiac arrhythmia. In this glossary, all definitions of the specific forms of ablation refer to the specific context of atrial fibrillation.
Acute presentation	Patients presenting to secondary or tertiary medical care on account of new or recurrent symptoms which my either be due to new-onset AF or to deterioration in rate control of existing AF.
Adjusted Dose	The situation where the dosage of a drug is adjusted to attain a particular physiological value, e.g. the dosage of warfarin may be adjusted to attain a particular INR value.
AF burden	A measure of the degree to which the presence of AF has a detrimental effect on the patient's quality of life. It is normally measured either as the proportion of time spent in AF, or the number of AF episodes per unit time.
AF recurrence	The recurrence of an episode of AF following one or more prior episodes of the arrhythmia in either its paroxysmal or persistent form.
Algorithm	A computer program that determines whether an ECG recorded by a device is abnormal. The program calculates multiple parameters of the recorded signal to determine if it is abnormal. For example, if the co-efficient of variation of the inter-beat interval calculated by the algorithm is above a pre-defined threshold the device will sound an alarm or record that an event has occurred.
Ambulatory-ECG	An ECG monitoring tool in which a continuous ECG recording is made while the patient remains able to walk around freely and pursue most normal daily activities

Term	Definition
Antiarrhythmic	A drug or interventional procedure that has a therapeutic effect against cardiac arrhythmias.
Antiarrhythmic drugs	Drugs used to suppress abnormal rhythms of the heart. The five main classes are: Class I agents interfere with the sodium (Na+) channel; Class II agents are anti-sympathetic nervous system agents; Class III agents affect potassium (K+) efflux; Class IV agents affect calcium channels and the AV node; Class V agents work by other or unknown mechanisms.
Anticoagulation	A form of thromboprophylaxis involving the use of anticoagulant drugs such as warfarin that inhibit the coagulation/clotting of blood.
Anticoagulant drug	Anticoagulants are drugs that prevent or reduce coagulation of blood, increasing the clotting time. Common examples are vitamin K antagonists (such as warfarin) and the direct-acting oral anticoagulants.
Antiplatelet therapy	A form of thromboprophylaxis involving the use of antiplatelet drugs (such as aspirin) that prevent platelet aggregation and inhibit the formation of blood clots.
Antithrombotic therapy	See ‘thromboprophylaxis’.
Aortic plaque	The deposits of atherosclerotic plaque within the aorta. The extent of aortic plaque is classified as ‘simple’, ‘moderate’ or ‘complex’
Aortic stenosis	An abnormal narrowing of the aortic valve
Arrhythmia	Abnormal cardiac rhythms, such as atrial fibrillation.
Arrhythmia surgery	Antiarrhythmic surgical interventions to treat the abnormal heart rhythm
Atrial arrhythmias	Cardiac arrhythmias that originate in the atria. AF is an atrial arrhythmia. See also ‘arrhythmia’.
Atrial contractile function	A measurement of the contractile function of the atria. This is normally measured using echocardiography.
Atrial fibrillation (AF)	An atrial arrhythmia characterised by an absence of regular P waves on an electrocardiogram, and normally resulting in a fast ventricular response. See also ‘atrial arrhythmia’.
Atrial filling fraction	A measurement of the contractile function of the atria. This is normally measured using echocardiography
Atrioventricular node ablation	Use of energy (usually radiofrequency) to destroy tissue of the atrioventricular node to alter conduction of electrical signals through this part of the heart.
Atrioventricular-blocking drug	A drug that inhibits the ability of the atrioventricular node to conduct electrical signals to the ventricles.
Beta-blockers	A class of drugs that work by inhibiting the activation of beta-adrenergic receptors. They slow the heart rate by blocking some of the actions of adrenaline on the heart and are used in the heart rate control of atrial fibrillation. They are classified as "Type II" in the Vaughan Williams classification.
Blanking period	Period after surgery (usually 1 to 3 months) during which recurrences of atrial fibrillation are not counted as true recurrences.

Term	Definition
Bleeding risk tool	A risk prediction tool that is able to predict (with variable degrees of accuracy) whether a patient will, or will not, suffer from bleeding.
Blood pressure monitor	These devices measure blood pressure on a beat-by-beat basis, and can therefore be used to collect information about heart rhythm if the inter-beat intervals are analysed. Newer machines may be fitted with an inbuilt algorithm that calculates if the inter-beat variability is within the normal range, or if it indicates arrhythmia.
Blood pressure sphygmanometer	See ‘blood pressure monitor’.
Bradycardia	A slow heartbeat. The occurrence of bradycardia is often recorded as an adverse event to some antiarrhythmic or chronotropic drugs. Such occurrences are referred to as bradycardic events.
Calcium channel blockers	A class of drugs that work by inhibiting the movement of calcium through calcium channels and are used in rate control. They are classified as "Type IV" in the Vaughan Williams classification.
Cardioembolic stroke	An embolic stroke whose aetiology is presumed to be the embolization of an intra-cardiac thrombus.
Cardiomegaly	An abnormal enlargement of the heart. It is normally measured in terms of the cardiothoracic ratio from a chest X-ray or by measurement using echocardiography.
Cardiomemo	An event recorder that records cardiac rhythm when activated by the patient
Cardiothoracic ratio (CTR)	See ‘cardiomegaly’.
Cardiothoracic surgery	Surgery performed on organs within the thorax, including the heart, lungs and oesophagus.
Cardioversion	In the context of AF, cardioversion is the process of restoring normal sinus rhythm. There are two commonly used forms of cardioversion: electrical cardioversion and pharmacological cardioversion. The former involves the administration of a transthoracic electrical shock and may be referred to as direct current cardioversion; the latter involves the administration of antiarrhythmic drugs.
Carer (caregiver)	Someone who looks after family, partners or friends in need of help because they are ill, frail or have a disability.
Catheter ablation	Ablation of heart tissue via a catheter introduced through the chest wall or into the interior of the heart via blood vessels. It usually refers to catheters introduced via blood vessels. See also ‘ablation’.
Cerebral infarction	Damage to the brain following a reduction of blood supply to that area, resulting in a stroke.
Cerebrovascular accident (CVA)	See ‘stroke’.
Cerebrovascular disease	Disease of the blood vessels within the brain. Cerebrovascular disease can be caused by blocked or otherwise damaged blood vessels and is the cause of strokes. See also ‘stroke’.
Chronotropic	In the context of pharmacology, the ability of a therapeutic intervention to control heart rate.

Term	Definition
Chronotropic incompetence	The inability of the body to appropriately alter heart rate during periods of physical exertion. See also ‘chronotropic’.
Class I AADs	See ‘sodium channel blockers’.
Class III AADs	See ‘potassium channel blockers’.
Clinically relevant non-major bleeding (CRNMB)	Any sign or symptom of bleeding that does not fit the criteria for the ISTH definition of major bleeding but does meet at least one of the following criteria: 1) requiring medical intervention by a healthcare professional, 2) leading to hospitalization or increased level of care, and 3) prompting a face to face (i.e., not just a telephone or electronic communication) evaluation.
Congestive heart failure (CHF)	Heart failure characterised by the inability of the heart to adequately support the body’s physiological requirements.
Conventional anticoagulation	The use of oral anticoagulation as a means of thromboprophylaxis, aiming for a target INR (usually 2.5, range 2–3) with monitoring and dose adjustment in an anticoagulation clinic.
Coronary artery bypass grafting	A surgical procedure designed to treat coronary artery disease by diverting blood around narrowed or clogged areas of major arteries to improve blood flow and the delivery of oxygen to the heart.
Coronary artery disease	A disease which affects the arteries of the heart, normally through atherosclerosis of the coronary arteries, reducing the supply of blood to the heart and causing ischaemia and angina. See also ‘ischaemic heart disease’.
Coumarin derivative	An anticoagulant drug that is derived from coumarin. Examples of coumarin derivatives include the anticoagulant warfarin.
Cryoballoon ablation	Catheter ablation using sub-freezing temperatures to destroy tissue. The catheter is positioned within the heart via the blood vessels. See also ‘ablation’.
CT scan	Computed tomography scan, an imaging technique using X-rays.
Day case	In the context of cardioversion, a day case refers to the discharge of patients following elective cardioversion on the same day on which they were admitted.
Defibrillator	In the context of AF, a device used to deliver the electrical shock used in electrical cardioversion.
Diastolic	Relating to the phase of the cardiac cycle where the chambers of the heart fill with blood prior to being pumped out during the subsequent systolic phase. See also ‘systolic’.
Direct current cardioversion	Cardioversion performed using an electric shock. See ‘cardioversion’
Direct-acting oral anticoagulants (DOACs)	A group of anticoagulants that includes drugs such as apixaban, edoxaban, dabigatran and rivaroxaban. These drugs may also be referred to as ‘novel oral anticoagulants (NOACs)’, but have been referred to as DOACs throughout this guideline.
Drug-eluting stents	Special metallic devices which are placed within the coronary artery to reduce the likelihood of coronary stenosis recurring following angioplasty (balloon dilatation of the coronary artery). Drug eluting

Term	Definition
	stents have special drugs within their structure that greatly reduce the recurrence of stenosis.
Dyspnoea	Breathlessness.
Echocardiogram	An examination of the heart using ultrasound-imaging techniques. An echocardiogram may be performed by placing the ultrasound device across the chest (transthoracic echocardiography), or by inserting it down the gullet to view the heart from behind (transoesophageal echocardiography).
Electrical cardioversion (ECV)	See 'cardioversion'.
Electrocardiograph	A device which traces the electrical activity of the heart by recording the electrical potentials at electrodes placed at various locations around the chest. The recording produced by the electrocardiograph is referred to as an electrocardiogram.
Electrocardiography	A graphical recording of the direction and magnitude of the electrical activity of cardiac tissue. As an action potential moves towards an electrode this is recorded as an upward inflection, and as it moves away from the electrode this is recorded as a downward inflection.
Electrolyte abnormalities	Abnormalities or an imbalance in one or more of the body's salts or other chemicals in the blood circulation.
Embolic	The passage within the blood stream of a body (e.g. blood clot), which has formed somewhere and ends up elsewhere within the body (e.g. brain).
Embolism	A blockage of blood flow caused by an embolus blocking a blood vessel. See also 'embolus'.
Embolus	A small piece of material (which may, for example, be a blood clot, an air bubble or a piece of atherosclerotic deposit) that may block blood vessels. See also 'thrombus'.
Event-ECG recorder	An ECG recording device, which only produces an ECG recording when susceptible electrical activity is detected. It may be triggered automatically or by the patient upon the occurrence of symptoms. See 'cardiomemo'.
Exercise tolerance	A measure of a patient's capacity for physical exertion.
Extra cellular fluid volume	A term that refers to the fluid bathing the body's cells.
Focal AF	AF secondary to a focus of abnormal cells (e.g. near the pulmonary veins) that can initiate AF.
Functional heart disease	Abnormalities of cardiac function – either in systole or diastole.
Gastrointestinal bleeding	All forms of bleeding occurring within the gastrointestinal tract, from the mouth to the rectum.
Haemodynamic function	An assessment of cardiac function.
Haemodynamic instability	Where cardiac function is compromised so that the patient becomes clinically unstable.
Haemorrhagic death	Death caused by a haemorrhagic event such as an intracranial haemorrhage.

Term	Definition
Haemorrhagic stroke	A stroke secondary to cerebral haemorrhage, resulting from a rupture of a blood vessel in the brain. About 13% of all strokes are haemorrhagic.
Haemorrhagic transformation	The situation where there is bleeding into a (usually large) cerebral infarction, especially in the early phase of a stroke.
Heart failure	See 'congestive heart failure'.
Heart murmur	An audible sound with or without a stethoscope, which relates to abnormal flow within the heart or an abnormal communication within the circulatory system.
Heart rate	The rate at which the heart performs a complete cycle of coordinated muscular contraction. It is measured in beats per minute (bpm).
Heart rate monitors	Devices that record the pulse. These may detect arrhythmia through variations in inter-beat interval.
Holter monitor	An ambulatory ECG recording device.
Hybrid ablation	Ablation of cardiac tissue occurring as the result of two concurrently applied catheter approaches. One catheter is placed through the chest and heart walls (thoracoscopy) and the other is positioned within the heart via blood vessels. See also 'ablation'.
Hyperadrenergic state	Situations where there is abnormal circulating adrenaline (and similar hormones) and/or activation of the sympathetic nervous system e.g. 'fight or flight' reaction.
Hypertension	Abnormally high blood pressure.
Infarction	An ischaemic lesion. Cerebral infarctions can result in stroke, and myocardial infarctions can result in a heart attack. See also 'myocardial infarction'.
Informed dissent	The situation whereby a patient elects to abstain from receiving the optimal therapeutic intervention in the knowledge that this could cause them harm.
Inotropic	Drugs that can stimulate the contraction of the heart
International Normalised ratio	The international normalised ratio (INR) is a laboratory measurement of clotting time, developed by the World Health Organisation (WHO). The higher the number, the longer it takes for blood to clot. The INR is calculated as a ratio of the patient's prothrombin time (PT) to a control PT, which is standardized for the potency of the thromboplastin reagent. This uses the following formula: INR = Patient PT ÷ Control PT. The INR number should be between 2 and 3 if VKAS (see Vitamin K Antagonists) are taken. INR measurements are required when VKAs are taken but are not generally measured for people using NOACs, on the basis that clotting time is more predictable when NOACs are taken.
Intra-cardiac	Occurring within the chambers of the heart.
Intracranial haemorrhage or bleeding	Bleeding occurring within the skull. A sub-set of intracranial bleeding is intracerebral bleeding, which occurs within the brain. This may result in a haemorrhagic stroke.

Term	Definition
Intraoperative	The period of time during a surgical procedure.
Intubation	Being intubated with a transoesophageal breathing tube connected to a mechanical ventilator.
Ischaemic heart disease	Heart disease characterised by a reduced supply of blood to the heart. See also 'coronary artery disease'.
Ischaemic stroke	A stroke resulting from a blockage of cerebral arteries by a thrombus or other embolus.
K^+ blockers	See 'potassium channel blockers'.
Lacunar infarction	Stroke secondary to blockage of the small vessels especially at the border of zones supplied by different arteries.
Laser ablation	Catheter ablation using laser energy to destroy tissue. The catheter is positioned within the heart via the blood vessels. See also 'ablation'.
Lead, leads	See '12 lead ECG'.
Lead I devices	A lead I device records electrical flow in a single direction only. See also '12 lead ECG'.
Left atrial appendage occlusion	This procedure involves the implantation of a device which closes off the left atrial appendage and aims to prevent blood clots within this appendage from entering the bloodstream and resulting in a stroke. This may be considered as an alternative to anticoagulant therapy in certain circumstances, such as when anticoagulation is contraindicated.
Left atrial appendage velocity	A measurement of the blood flow within the left atrial appendage, usually on TEE
Left ventricular dysfunction (LVD)	Impaired function of the left ventricle.
Left ventricular ejection fraction (LVEF)	The percentage of blood within the left ventricle that is ejected at each contraction.
Left ventricular end diastolic diameter (LVEDD)	A measurement of the size of the heart on echo, referring to the internal dimension of the heart in diastole.
Left ventricular end systolic diameter (LVESD)	A measurement of the size of the heart on echo, referring to the internal dimension of the heart in systole.
Lone AF	AF that occurs in the absence of any comorbid cardiovascular disease or other precipitants of AF.
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.
Magnetic resonance imaging (MRI)	A non-invasive imaging technique allowing detailed examination of the heart.
Major bleeding	Major bleeding is defined [according to International Society on Thrombosis and Haemostasis (ISTH) criteria] as clinically overt bleeding accompanied by a decrease in the haemoglobin level of at least 2 g/dl or transfusion of at least 2 units of packed red cells, occurring at a critical site (intracranial, intraocular, intraspinal, intra-

Term	Definition
	articular, intramuscular with compartment syndrome, pericardial, retroperitoneal), or resulting in death
Management strategy	The overarching plan on how to treat a particular patient. In the context of AF, there are two main management strategies – rate control and rhythm control.
Mapping catheter	A non-ablative catheter that guides the location of the ablative catheter by identifying regions of abnormal conduction.
Maximum workload	A measure of exercise tolerance. See 'exercise tolerance'.
Medically refractory	In the context of AF, a patient is medically refractory if successive trials of different drugs and attempts at cardioversion fail to adequately control the symptoms or pathophysiology of AF.
Minor bleeding	An acute clinically overt event not meeting the criteria for either major or clinically relevant non major bleeding
Mitral annular abnormalities	Echo abnormalities of the mitral valve ring/annulus, such as mitral annular calcification.
Mitral regurgitation	A backwards flow of blood through the mitral valve normally caused by a dysfunctional mitral valve disease. Mitral regurgitation is classified as 'mild', 'moderate' or 'severe'.
Mitral stenosis	An abnormal narrowing of the mitral valve. It can be measured echocardiographically by the mitral valve area.
Mitral valve calcification	Deposition of calcium on the mitral valve.
Mitral valve disease	Common generic term for disease of the mitral valve.
Mitral valve prolapse	Condition where one or more mitral valve leaflets do not oppose correctly and there is backward movement of the valve into the atrium, leading to mitral regurgitation.
Mitral valvuloplasty	Stretching of the mitral valve, at surgery or using a balloon technique.
Mobile ECG devices	Any ECG device that is light enough to be carried in a pocket, in a small waist-bag or on the wrist. These are usually lead I devices, but may also be other lead devices or heart rate monitors.
Monotherapy	In the context of drug therapy, the administration of a single drug for a particular indication.
Myocardial infarction (MI)	Heart attack.
Na^+ channel blockers	See 'sodium channel blockers'.
New-onset atrial fibrillation	A patient presenting to medical care with atrial fibrillation whose new or changing symptoms suggest that the episode of AF commenced less than 48 hours prior to presentation.
New York Heart Association (NYHA)	A score graded between 1 and 4 that measures cardiac function. Those patients with a score of 4 are considered to have severe heart failure; those with a score of 1 are considered to have asymptomatic or mild heart failure.
Novel oral anticoagulants (NOACs)	See 'direct-acting oral anticoagulants'.

Term	Definition
Nurse-led cardioversion	Practice where the cardioversion procedure is organised, performed and patient follow-up undertaken by specialist nurses.
Open surgical ablation	See 'surgical ablation'.
Pacing	The situation where a device (a pacemaker) complements or replaces the natural conducting system of the heart.
Palpitations	The experience of one's own heartbeat as an awareness of the heart beating or a thumping sensation originating in the chest.
Paroxysmal AF	AF which terminates spontaneously within seven days of onset and most often within 48 hours of onset.
Patent foramen ovale	A 'hole in the heart' where there is a congenital connection between the left and right atria.
Percutaneous coronary intervention (PCI)	Any procedure on the heart undertaken by insertion of a device (e.g. stent) through a small hole in an artery (e.g. radial artery, femoral artery).
Perioperative	The period from admission through surgery until discharge, encompassing the pre-operative and post-operative periods.
Peripheral artery disease	Atherosclerotic vascular disease involving the peripheral arteries .
Permanent AF	AF which is accepted without attempted cardioversion or which cannot be terminated by cardioversion.
Persistent AF	AF present continuously for seven days or more or terminated by cardioversion.
Pharmacological cardioversion (PCV)	See 'cardioversion'.
Photoplethysmography	Photoplethysmography works by shining light into the tissues of the finger (or sometimes the tissues of the face) and continuously measuring the amount of light that is absorbed by the tissues. The quantity of light absorbed is proportional to the quantity of blood in the microvasculature, which varies over time due to the pulse wave. Therefore photoplethysmography indirectly allows the pulse waves to be analysed, which provides information on the heart rhythm.
Pill-in-the-pocket	A management strategy for paroxysmal AF involving the patient self-administering antiarrhythmic drugs only upon the onset of an episode of AF.
Platelet-thrombus	Blood clot that is rich in platelets rather than fibrin.
Pneumonectomy	Removal of whole or part of a lung.
Polypharmacy	The use or prescription of multiple medications.
Postoperative	Pertaining to the period after patients leave the operating theatre, following surgery.
Postoperative atrial fibrillation	Atrial fibrillation that is present after patients leave the operating theatre, following surgery.
Potassium channel blockers	Potassium channel blockers reduce the conduction of potassium ions (K+) through potassium channels. They are known as Class III agents in the Vaughan Williams classification. They prolong the action

Term	Definition
	potential duration and refractory period, which prevents re-entrant arrhythmias.
Precipitant	A disease process, toxin, or physiological abnormality which is known to predispose towards development of AF. In many cases, AF precipitants may not be identifiable, in other cases there are identifiable precipitants such as heart failure or alcohol excess.
Preoperative	The period before surgery commences.
Pro-arrhythmic	Pre-disposing to the development of cardiac arrhythmias.
Prophylactic	Having a preventative action against one or more adverse events
Pulmonary vein isolation	Pulmonary vein isolation describes the main form of ablation applied to cardiac tissue. The procedure uses ablation to create a circumferential scar around the entrance of the pulmonary veins where they connect to the left atrium. This isolates the tissue of the pulmonary vein from the rest of the heart – thus, if a focus precipitating AF is from within the pulmonary veins, the abnormal electrical impulses cannot affect the heart rhythm.
Pulse palpation	The act of feeling for, and counting, the pulse.
QT prolongation	The prolongation of the QT interval on an electrocardiogram
Radiofrequency multi-electrode ablation	Catheter ablation using radiofrequency energy to destroy tissue. This is delivered to multiple locations of tissue at the same time via a multi-electrode. The catheter is positioned within the heart via the blood vessels. See also ‘ablation’.
Radiofrequency point by point ablation	Catheter ablation using radiofrequency energy to destroy tissue. This is delivered to single points of tissue in turn (‘point by point’) via a single irrigated catheter tip. This may be applied via the chest and heart walls as an integral part of the thoracoscopy ablation procedure, but more commonly is positioned within the heart via the blood vessels. See also ‘ablation’.
Rapid atrial fibrillation	AF that is associated with a very fast heartbeat.
Rate control	The attempt to treat AF not through the restoration of sinus rhythm, but through the control of the ventricular rate and the management of stroke risk. See also ‘rhythm control’.
Recurrent atrial fibrillation	One or more episodes of atrial fibrillation occurring after the initial episode was resolved by returning to sinus rhythm.
Re-do of procedure (ablation)	The repeat of an ablation procedure after recurrence has occurred.
Resolved atrial fibrillation	A term used to describe when AF that was previously documented but is no longer detectable. The clinical code “atrial fibrillation resolved” is widely used in general practice. The concept is controversial as long term follow up of patients with AF indicate an increasing burden of AF. Evidence has shown that patients coded as having “AF resolved” remain at increased risk of the complications of AF such as stroke. Implying that AF is recurring in a large proportion of “AF resolved” patients. Following successful AF ablation (catheter or surgical) AF could resolve. AF transiently occurring as a consequence of known triggers of AF such as major cardiac surgery may also resolve although this remains unproven. The term is often misused when spontaneously

Term	Definition
	occurring AF is no longer detectable even though no attempts to permanently maintain sinus rhythm have been made. The committee considered that AF should not be considered to be truly resolved unless ablation or surgery has been performed.
Rhythm control	The attempt to treat AF through the restoration and maintenance of sinus rhythm. See also 'rate control'.
Right bundle branch block (RBBB)	A conduction abnormality of the heart due to impaired conduction down the right bundle of His.
Risk prediction tool	A score, based on past or current medical history, and/or biomarkers, that is able to predict (with variable degrees of accuracy) whether a patient will, or will not, develop a clinical outcome (such as stroke) in the future. This allows for appropriate management, such as preventative measures to be put in place, or for greater surveillance to be applied.
Risk stratification tool	See 'risk prediction tool'.
Self-management	In the context of anticoagulation, the process of the patient testing their own blood and making dose-adjustments where necessary.
Self-testing	In the context of anticoagulation, the process of the patient testing their own blood and their treating physician recommending dose-adjustments where necessary.
Side effect	An adverse event that occurs because of a therapeutic intervention.
Sinus rhythm	The normal pattern of electrical activity (and subsequent muscular contraction) of the heart.
Sodium channel blockers	Sodium channel blockers reduce the conduction of sodium ions (Na+) through sodium channels. They are classified as "Type I" in the Vaughan Williams classification. These drugs help to reduce cardiac arrhythmias.
Spontaneous cardioversion	The process of cardioversion that occurs in the absence of any therapeutic interventions.
Spontaneous echo contrast	Smoke-like appearance within the chambers of the heart – usually on TOE – which indicates stasis of blood within the chamber.
Statins	Statins, also known as HMG-CoA reductase inhibitors, are a class of lipid-lowering medications, which may also modify the likelihood of developing atrial fibrillation after cardiothoracic surgery. There are several specific drugs, which vary in their ability to reduce cholesterol, as well as in their side effects and drug interactions. Examples of statins are atorvastatin, pravastatin, rosuvastatin, simvastatin and LDL cholesterol.
Stroke	A stroke is a life-threatening event due to the blood supply to part of the brain being impaired or completely cut off. This may be a result of a thrombus blocking a cerebral artery, arterial narrowing due to atherosclerosis, or through haemorrhage.
Stroke risk tool	A risk prediction tool that is able to predict (with variable degrees of accuracy) whether a patient will, or will not, suffer from a stroke.
Structural heart disease	The presence of abnormalities of the heart valves, muscle, chambers etc.

Term	Definition
Sudden cardiac death syndrome	The condition whereby a patient dies suddenly and unexpectedly with no obvious precipitants.
Supervised management	In the context of anticoagulation management, supervised management refers to the situation where a clinician determines any dose adjustments and takes blood measurements.
Supraventricular	Pertaining to the atria, e.g. supraventricular arrhythmia is an abnormal heart rhythm originating in the atria.
Surgical ablation	Ablation of cardiac tissue performed directly during open heart surgery. See also 'ablation'.
Systemic emboli	Emboli that has reached the systemic circulation, potentially causing a systemic embolism. See 'embolic'.
Systolic	Relating to the phase of contraction of the chambers of the heart during which they eject blood following the diastolic phase. See also 'diastolic'.
Tachycardia-induced cardiomyopathy	A form of cardiomyopathy (damage to the heart muscle cells) caused by an excessive heart rate.
Temporal pattern	The pattern distinguishing between different subtypes of AF.
Thoracoscopic ablation	See 'thoracoscopy ablation'.
Thoracoscopy ablation	Ablation of cardiac tissue via a thoracoscope" placed through the chest and heart walls. See also 'ablation'.
Thromboembolic event	Formation of a clot in a blood vessel (thrombus) that breaks free and blocks another vessel. The clot may obstruct a vessel in the lungs (pulmonary embolism), brain (stroke), gastrointestinal tract, kidneys, or leg.
Thromboembolic stroke	Thrombus that has travelled to the brain circulating leading to blockage of an artery and causing a stroke. See 'embolic', 'stroke'.
Thromboembolism	The embolisation (disloading and transportation in the blood) of a thrombus.
Thromboprophylaxis	The administration of antithrombotic therapy (anticoagulation, antiplatelet therapy) for the prevention of thrombus formation.
Thrombus	A blood clot in a blood vessel. A thrombus is a type of embolus. See embolus
Thyrotoxicosis	A disease caused by the hyperactivity of the thyroid glands.
Time in therapeutic range	The percentage of time that a person is within the target range of INR values, which is usually between 2 and 3.
TOE-guided cardioversion	In the context of cardioversion, the management of pericardioversion thromboembolic risk through the use of transoesophageal echocardiography (TOE) to screen for intra-cardiac thrombi alongside parenteral anticoagulation. See also 'conventional anticoagulation'.
Torsades de pointes	A type of ventricular arrhythmia, which is a polymorphic ventricular tachycardia characterised by 'twisting of points' and commonly associated with a prolonged QT interval on the ECG.

Term	Definition
Transient ischaemic attack	A transient ischaemic attack (TIA) is caused by a temporary disruption of blood supply to part of the brain. It is sometimes called a "mini stroke". This can cause symptoms similar to those in a stroke, such as speech disturbance, visual impairment and sensory and motor deficits. The effects last from a few minutes to several hours, but are fully resolved within 24 hours.
Transoesphageal echocardiography	See ‘echocardiogram’.
Treatment failure	Failure of the prescribed drug regimen to work. Demonstrated by a lack of clinical improvement or reduction in arrhythmia, etc.
Valvular heart disease	Diseases of heart valves, e.g. mitral valve disease.
Vascular death	Death caused by a cardiovascular disease or adverse cardiovascular event such as an acute myocardial infarction.
Vascular disease	Disease of the vascular system, including both coronary and peripheral blood vessels.
Vaughan-Williams	A classification system of antiarrhythmic drugs, depending on whether the drugs activity is as a sodium-channel blocker (Class I), a beta-blocker (Class II), a repolarisation-prolonging agent (Class III), or a calcium-channel blocker (Class IV).
Ventricular arrhythmias	Cardiac arrhythmias that originate in the ventricles. See also ‘arrhythmia’.
Ventricular rate control	See ‘rate control’.
Vitamin K antagonist (VKA)	An anticoagulant that works by inhibiting the clotting effects of vitamin K. An example is warfarin.
Volume loss	A term that usually refers to the amount of blood lost.
Wall motion index (WMI)	An echocardiographic measure of the contractile function of the ventricles.

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 people recruiting patients into an RCT having advance knowledge of the allocation sequence. The allocation sequence is the pre-defined random order of allocation to the groups (such as *intervention, control, intervention, intervention, control*…) according to which participants are allocated as they are recruited. Allocation concealment is important, because if the recruiters know the group that the next patient will be assigned to, they may express any bias by selectively recruiting or not recruiting that patient, depending on that patient’s characteristics. For example, if the next allocation is known to be to the intervention group, which the recruiter

Term	Definition
	subconsciously favours, and the next eligible patient appears frail and unlikely to have good outcomes, then the recruiter might avoid recruitment of that patient to the study. Allocation concealment attempts to reduce the risk of such bias.
Applicability	How well the results of a study or NICE evidence review can answer a clinical question or be applied to the population being considered.
Arm (of a clinical study)	Subsection of individuals within a study who receive one particular intervention, for example placebo arm.
Association	Statistical relationship between 2 or more events, characteristics or other variables. The relationship may or may not be causal.
Base case analysis	In an economic evaluation, this is the main analysis based on the most plausible estimate of each input. In contrast, see Sensitivity analysis.
Baseline	The initial set of measurements at the beginning of a study (after run-in period where applicable), with which subsequent results are compared.
Bayesian analysis	A method of statistics, where a statistic is estimated by combining established information or belief (the 'prior') with new evidence (the 'likelihood') to give a revised estimate (the 'posterior').
Before-and-after study	A study that investigates the effects of an intervention by measuring particular characteristics of a population both before and after taking the intervention, and assessing any change that occurs.
Bias	Influences on a study that can make the results look better or worse than they really are. Bias can even make it look as if a treatment works when it does not. Bias can occur by chance, deliberately or as a result of systematic errors in the design and execution of a study. It can also occur at different stages in the research process, for example, during the collection, analysis, interpretation, publication or review of research data. For examples see selection bias, performance bias, information bias, confounding factor, and publication bias.
Blinding	A way to prevent healthcare professionals, patients and outcome assessors in a clinical trial from knowing which study group each patient is in so they cannot influence the results. The purpose of 'blinding' or 'masking' is to protect against bias. If healthcare professionals know the group allocation of a patient, and tend to favour one group, this may influence the care given, which can then influence results. Similarly, if patients know their group allocation this can affect beliefs about their care and prognosis, which again can affect outcomes. Finally if the outcome assessor knows the group allocation then this can affect the way that outcomes are measured, which can also influence outcome. A single-blinded study is usually one in which patients are blinded (for example whether they are taking the experimental drug or a placebo) but may also refer to solitary blinding of the assessor. A double-blinded study is one in which both patients and healthcare professionals are blinded. A triple blind study is one in which neither the patients, clinicians or the outcome assessors know which treatment patients received. Whilst it is always possible to blind outcome assessors, blinding patients and healthcare professionals may sometimes be impossible (for example when comparing surgery to medical treatments) but this does not mean the study is not at risk of bias.
C statistic	The area under a receiver operated characteristic (ROC) curve, which provides an integrated measure of accuracy (sensitivity and specificity) at the full range of test thresholds. See *Receiver operated characteristic (ROC) curve*
Calibration	In a general sense this refers to the definition of values of a measure using the values derived from a gold standard method applied to the same object of measurement.

Term	Definition
	In the context of this guideline, calibration refers to the plotting of observed risks of an outcome against predicted risks derived from a prediction test. Good calibration will lead to a straight line that is close to the line extending at 45 degrees from the origin. This indicates that the test is able to accurately predict the actual risks.
Carer (caregiver)	Someone who looks after family, partners or friends in need of help because they are ill, frail or have a disability.
Case–control study	A study to find out the cause(s) of a disease or condition. This is done by comparing a group of patients who have the disease or condition (cases) with a group of people who do not have it (controls) but who are otherwise as similar as possible (in characteristics thought to be unrelated to the causes of the disease or condition). The researcher then asks the patient about exposure to risk factors in the past in an attempt to find out if past exposure is associated with the current condition (disease or no disease). 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. If the hypothesis is that tobacco smoke causes cancer, the researcher could analyse the differences between the cases and controls in terms of exposure to tobacco smoke in the past. This is an observational design and so it is vital that confounding by other factors is accounted for. Such studies are limited by the dependence on recall by the patient, both for exposure to the independent variable (i.e. smoking) and potential confounders (i.e. exposure to industrial pollutants, or dietary factors).
Case series	Report of a number of cases of a given disease, usually covering the course of the disease and the response to treatment. There is no comparison (control) group of patients.
Clinical efficacy	The extent to which an intervention is active when studied under controlled research conditions.
Clinical effectiveness	How well a specific test or treatment works when used in the 'real world' (for example, when used by a doctor with a patient at home), rather than in a carefully controlled clinical trial. Trials that assess clinical effectiveness are sometimes called management trials. Clinical effectiveness is not the same as efficacy.
Clinician	A healthcare professional that 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.

Term	Definition
Confidence interval (CI)	There is always some uncertainty in research. This is because a small group of patients is studied to predict the effects of a treatment on the wider population. The confidence interval is a way of expressing how certain we are about the findings from a study, using statistics. It gives a range of results that is likely to include the 'true' value for the population. The CI is usually stated as '95% CI', which means that the range of values has a 95 in a 100 chance of including the 'true' value. For example, a study may state that "based on our sample findings, we are 95% certain that the 'true' population blood pressure is not higher than 150 and not lower than 110". In such a case the 95% CI would be 110 to 150. A wide confidence interval indicates a lack of certainty about the true effect of the test or treatment – often because a small group of patients has been studied. A narrow confidence interval indicates a more precise estimate (for example, if a large number of patients have been studied).
Confounding factor	Something that influences a study and can result in misleading findings if it is not understood or appropriately dealt with. For example, a study of heart disease may look at a group of people that exercises regularly and a group that does not exercise. If the ages of the people in the 2 groups are different, then any difference in heart disease rates between the 2 groups could be because of age rather than (or in addition to) exercise. Therefore age is a confounding factor.
Consensus methods	Techniques used to reach agreement on a particular issue. Consensus methods may be used to develop NICE guidance if there is not enough good quality research evidence to give a clear answer to a question. Formal consensus methods include Delphi and nominal group techniques.
Control group	A group of people in a study who do not receive the treatment or test being studied. Instead, they may receive the standard treatment (sometimes called 'usual care') or a dummy treatment (placebo). The results for the control group are compared with those for a group receiving the treatment being tested. The purpose of a placebo control group is to allow the elimination of non-treatment effects so that the measure of effect is the treatment effect. The purpose of an active control is also to allow the elimination of non-treatment effects, so that the measure of effect is simply the difference in treatment effects. Ideally, the people in the control group should be as similar as possible to those in the treatment group. If the control group is sufficiently similar then non-treatment effects contributing to the outcome, such as the placebo effect or natural recovery, should cancel out when the group means are subtracted (as in a mean difference) or when the group risks are divided (as in a risk ratio).
Cost–benefit analysis (CBA)	Cost–benefit analysis is one of the tools used to carry out an economic evaluation. The costs and benefits are measured using the same monetary units (for example, pounds sterling) to see whether the benefits exceed the costs.
Cost–consequences analysis (CCA)	Cost–consequences analysis is one of the tools used to carry out an economic evaluation. This compares the costs (such as treatment and hospital care) and the consequences (such as health outcomes) of a test or treatment with a suitable alternative. Unlike cost–benefit analysis or cost-effectiveness analysis, it does not attempt to summarise outcomes in a single measure (like the quality-adjusted life year) or in financial terms. Instead, outcomes are shown in their natural units (some of which may be monetary) and it is left to decision-makers to determine whether, overall, the treatment is worth carrying out.

Term	Definition
Cost-effectiveness analysis (CEA)	Cost-effectiveness analysis is one of the tools used to carry out an economic evaluation. The benefits are expressed in non-monetary terms related to health, such as symptom-free days, heart attacks avoided, deaths avoided or life years gained (that is, the number of years by which life is extended as a result of the intervention).
Cost-effectiveness model	An explicit mathematical framework, which is used to represent clinical decision problems and incorporate evidence from a variety of sources in order to estimate the costs and health outcomes.
Cost–utility analysis (CUA)	Cost–utility analysis is one of the tools used to carry out an economic evaluation. The benefits are assessed in terms of both quality and duration of life, and expressed as quality-adjusted life years (QALYs). See also utility.
Credible interval (CrI)	The Bayesian equivalent of a confidence interval.
Decision analysis	An explicit quantitative approach to decision-making under uncertainty, based on evidence from research. This evidence is translated into probabilities, and then into diagrams or decision trees which direct the clinician through a succession of possible scenarios, actions and outcomes.
Deterministic analysis	In economic evaluation, this is an analysis that uses a point estimate for each input. In contrast, see Probabilistic analysis
Diagnostic odds ratio	The diagnostic odds ratio is a measure of the effectiveness of a diagnostic test. It is defined as the ratio of the odds of the test being positive if the subject has a disease relative to the odds of the test being positive if the subject does not have the disease.
Discounting	Costs and perhaps benefits incurred today have a higher value than costs and benefits occurring in the future. Discounting health benefits reflects individual preference for benefits to be experienced in the present rather than the future. Discounting costs reflects individual preference for costs to be experienced in the future rather than the present.
Disutility	The loss of quality of life associated with having a disease or condition. See Utility
Dominance	A health economics term. When comparing tests or treatments, an option that is both less effective and costs more is said to be ‘dominated’ by the alternative.
Drop-out	A participant who withdraws from a trial before the end.
Economic evaluation	An economic evaluation is used to assess the cost effectiveness of healthcare interventions (that is, to compare the costs and benefits of a healthcare intervention to assess whether it is worth doing). The aim of an economic evaluation is to maximise the level of benefits – health effects – relative to the resources available. It should be used to inform and support the decision-making process; it is not supposed to replace the judgement of healthcare professionals. There are several types of economic evaluation: cost–benefit analysis, cost–consequences analysis, cost-effectiveness analysis, cost-minimisation analysis and cost–utility analysis. They use similar methods to define and evaluate costs, but differ in the way they estimate the benefits of a particular drug, programme or intervention.
Effect (as in effect measure, treatment effect, estimate of effect, effect size)	A measure that shows the magnitude of the outcome in one group compared with that in a control group. For example, if the absolute risk reduction is shown to be 5% and it is the outcome of interest, the effect size is 5%. The effect size is usually tested, using statistics, to find out how likely it is that the effect is a result of the treatment and has not just happened by chance (that is, to see if it is statistically significant).

Term	Definition
Effectiveness	How beneficial a test or treatment is under usual or everyday conditions, compared with doing nothing or opting for another type of care.
Efficacy	How beneficial a test, treatment or public health intervention is under ideal conditions (for example, in a laboratory), compared with doing nothing or opting for another type of care.
Epidemiological study	The study of a disease within a population, defining its incidence and prevalence and examining the roles of external influences (for example, infection, diet) and interventions.
EQ-5D (EuroQol 5 dimensions)	A standardised instrument used to measure health-related quality of life. It provides a single index value for health status.
Evidence	Information on which a decision or guidance is based. Evidence is obtained from a range of sources including randomised controlled trials, observational studies, expert opinion (of clinical professionals or patients).
Exclusion criteria (literature review)	Explicit standards used to decide which studies should be excluded from consideration as potential sources of evidence.
Exclusion criteria (clinical study)	Criteria that define who is not eligible to participate in a clinical study.
Extended dominance	If Option A is both more clinically effective than Option B and has a lower cost per unit of effect, when both are compared with a do-nothing alternative then Option A is said to have extended dominance over Option B. Option A is therefore cost effective and should be preferred, other things remaining equal.
Extrapolation	An assumption that the results of studies of a specific population will also hold true for another population with similar characteristics.
Follow-up	Observation over a period of time of an individual, group or initially defined population whose appropriate characteristics have been assessed in order to observe changes in health status or health-related variables.
Generalisability	The extent to which the results of a study hold true for groups that did not participate in the research. See also external validity.
Gold standard	A method, procedure or measurement that is widely accepted as being the best available to test for or treat a disease.
GRADE, GRADE profile	A system developed by the GRADE Working Group to address the shortcomings of present grading systems in healthcare. The GRADE system uses a common, sensible and transparent approach to grading the quality of evidence. The results of applying the GRADE system to clinical trial data are displayed in a table known as a GRADE profile.
Harms	Adverse effects of an intervention.
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	When studies include relatively few patients and have few events (or a high standard deviation in a continuous outcome) this will lead to wide confidence intervals (CI) around the estimate of effect. The CIs represent the range of values where the true effect in the population is

Term	Definition
	likely to lie, so if the CIs cross thresholds of clinical benefit, clinical harm, or no clinical benefit/harm this means that the result is consistent with potentially conflicting effects in the population. Imprecision is the crossing of such thresholds by the CIs.
Inclusion criteria (literature review)	Explicit criteria used to decide which studies should be considered as potential sources of evidence.
Incremental analysis	The analysis of additional costs and additional clinical outcomes with different interventions.
Incremental cost	The extra cost linked to using one test or treatment rather than another. Or the additional cost of doing a test or providing a treatment more frequently.
Incremental cost-effectiveness ratio (ICER)	The difference in the mean costs in the population of interest divided by the differences in the mean outcomes in the population of interest for one treatment compared with another.
Incremental net benefit (INB)	The value (usually in monetary terms) of an intervention net of its cost compared with a comparator intervention. The INB can be calculated for a given cost-effectiveness (willingness to pay) threshold. If the threshold is £20,000 per QALY gained then the INB is calculated as: (£20,000 × QALYs gained) − Incremental cost.
Indirectness	The available evidence is different to the review question being addressed, in terms of PICO (population, intervention, comparison and outcome).
Intention-to-treat analysis (ITT)	An assessment of the people taking part in a clinical trial, based on the group they were initially (and randomly) allocated to. This is regardless of whether or not they dropped out, fully complied with the treatment or switched to an alternative treatment. Intention-to-treat analyses are often used to assess clinical effectiveness because they mirror actual practice: that is, not everyone complies with treatment and the treatment people receive may be changed according to how they respond to it.
Intervention	In medical terms this could be a drug treatment, surgical procedure, diagnostic or psychological therapy. Examples of public health interventions could include action to help someone to be physically active or to eat a more healthy diet.
Intraoperative	The period of time during a surgical procedure.
Kaplan Meier curve	The Kaplan Meier curve charts the cumulative probability of survival over time in a sample of people (where survival denotes not yet having had an event, such as recurrence). Probability is plotted on the y axis and time is plotted on the x axis. The curve begins at 100% probability at time 0, and the probability progressively drops over time as people have first events.
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.

Term	Definition
Logistic regression or Logit model	In statistics, logistic regression is a type of analysis used for predicting the outcome of a binary dependent variable based on one or more predictor variables. It can be used to estimate the log of the odds (known as the 'logit').
Loss to follow-up	A patient, or the proportion of patients, actively participating in a clinical trial at the beginning, but who did not provide outcome data because they withdrew, or because the researchers were unable to trace or contact them 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 with a weighted average 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: TN/(TN+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 threshold is £20,000 per QALY gained then the NMB for an intervention is calculated as: (£20,000 × mean QALYs) − mean cost. The most preferable option (that is, the most clinically effective option to have an ICER below the threshold selected) will be the treatment with the highest NMB.
Net reclassification Index (NRI)	This is expressed in terms of one (index) risk tool to another (comparator) risk tool, and gives a score between -2 and +2 (with +2 representing the best possible performance of the index tool relative to the comparator, and -2 the worst). The score represents the net improvement of the index test relative to the comparator in terms of the proportion of true cases (judged by later development of stroke/TE) that are correctly up-classified by the tool (relative to any false negative classifications yielded by the comparator), and the proportion of false cases (judged by the lack of later stroke/TE) that are correctly down-classified by the tool (relative to any false positive classifications yielded by the comparator). Meanwhile, incorrect up-classification or incorrect down-classification of the index relative to the comparator convey negative scores to the NRI, and so if a score is negative overall this indicates the index is less accurate than the comparator.
Non-randomised intervention study	A quantitative study investigating the effectiveness of an intervention, that does not use randomisation to allocate patients (or units) to treatment groups. Non-randomised studies include observational studies, where allocation to groups occurs through usual treatment decisions or people's preferences. Non-randomised studies can also be experimental, where the investigator has some degree of control over the allocation of treatments. Non-randomised intervention studies can use a number of different study designs, and include cohort studies, case–control studies, controlled before-and-after studies, interrupted-time-series studies and quasi-randomised controlled trials.
Number needed to treat (NNT)	The average number of patients who need to be treated to get a positive outcome. For example, if the NNT is 4, then 4 patients would

Term	Definition
	have to be treated to ensure 1 of them gets better. The closer the NNT is to 1, the better the treatment. For example, if you give a stroke prevention drug to 20 people before 1 stroke is prevented, the number needed to treat is 20. See also number needed to harm, absolute risk reduction.
Observational study	Individuals or groups are observed or certain factors are measured. No attempt is made to affect the outcome. For example, an observational study of a disease or treatment would allow 'nature' or usual medical care to take its course. Changes or differences in one characteristic (for example, whether or not people received a specific treatment or intervention) are studied without intervening. There is a greater risk of selection bias than in experimental studies.
Odds ratio	Odds are different to risks. Odds are defined as the number of people in a group with an event divided by the number of people in that same group without an event. This contrasts with a risk, which is the number of people in a group with an event divided by the sum of the people with and without an event in that same group. Thus, in a group of 10 people, if 2 people have events and 8 people do not have events, the odds are 2 / 8 = 0.25. In contrast the risk would be 2 / (8+2) =0.2. The odds *ratio* (OR) is the ratio of the odds in the intervention group with the odds in the control group. So if group 1 had odds of 0.25 and group 2 had odds of 0.125, the odds ratio would be 2.0. An odds ratio of 1 between 2 groups would show that the odds of the event (for example a person developing a disease, or a treatment working) are the same for both. An odds ratio greater than 1 means the odds of the event is greater in the first group. An odds ratio less than 1 means that the odds of an event are smaller in the first group. Sometimes odds can be compared across more than 2 groups – in this case, one of the groups is chosen as the 'reference category', and the odds ratio is calculated for each group compared with the reference category. For example, to compare the odds of dying from lung cancer for non-smokers, occasional smokers and regular smokers, non-smokers could be used as the reference category. Odds ratios would be worked out for occasional smokers compared with non-smokers and for regular smokers compared with non-smokers. See also confidence interval, risk ratio.
Opportunity cost	The loss of other healthcare programmes displaced by investment in or introduction of another intervention. This may be best measured by the health benefits that could have been achieved had the money been spent on the next best alternative healthcare intervention.
Outcome	The impact that a test, treatment, policy, programme or other intervention has on a person, group or population. Outcomes from interventions to improve the public's health could include changes in knowledge and behaviour related to health, societal changes (for example, a reduction in crime rates) and a change in people's health and wellbeing or health status. In clinical terms, outcomes could include the number of patients who fully recover from an illness or the number of hospital admissions, and an improvement or deterioration in someone's health, functional ability, symptoms or situation. Researchers should decide what outcomes to measure before a study begins.
P value	The p value is a statistical measure that indicates whether or not an effect is statistically significant. For example, if a study comparing 2 treatments found that one seems more effective than the other, the p value is the probability of obtaining these results by chance. By convention, if the p value is below 0.05 (that is, there is less than a 5% probability that the results occurred by

Term	Definition
	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.
Performance bias	Performance bias results from systematic differences in the way that patients from each group are cared for and treated during the intervention phase of a trial, over and above the intrinsic differences relating to the actual treatments themselves. For example one group may experience more contact, more senior nursing care, or more vigilant monitoring. It also refers to differences in patient beliefs across groups about the efficacy of treatment. Both of these differences can influence outcome and thus cause bias. Blinding of healthcare professional and patients can help to reduce such bias.
Perioperative	The period from admission through surgery until discharge, encompassing the preoperative and postoperative periods.
Placebo	A fake (or dummy) treatment given to participants in the control group of a clinical trial. It is indistinguishable from the actual treatment (which is given to participants in the experimental group). The aim is to determine what effect the experimental treatment has had – over and above any placebo effect caused because someone has received (or thinks they have received) care or attention.
Polypharmacy	The use or prescription of multiple medications.
Posterior distribution	In Bayesian statistics this is the probability distribution for a statistic based after combining established information or belief (the prior) with new evidence (the likelihood).
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)
Postoperative	Pertaining to the period after patients leave the operating theatre, following surgery.
Post-test probability	In diagnostic tests: The proportion of patients with that particular test result who have the target disorder (post-test odds/[1 plus post-test odds]).
Power (statistical)	The ability to demonstrate an association when one exists. Power is related to sample size; the larger the sample size, the greater the power and the lower the risk that a possible association could be missed.
Predictive test	A predictive test evaluates the risk that a future event will occur. An example is a stroke prediction tool, which provides a score based on the patients array of characteristics, symptoms and sometimes biomarkers. The score determines the risk of stroke. Often the scores are dichotomised to a binary ‘not at risk/at risk’. For example, the CHADSVASC can score between 0 and 9, but in men a score of 0-1 is generally regarded as denoting low stroke risk and a score of 2 or more is generally regarded as sufficient stroke risk to warrant the initiation of anticoagulants. Predictive tests are evaluated in a similar way to diagnostic tests.
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.

Term	Definition
Prevalence	See Pre-test probability.
Prior distribution	In Bayesian statistics this is the probability distribution for a statistic based on previous evidence or belief.
Primary care	Healthcare delivered outside hospitals. It includes a range of services provided by GPs, nurses, health visitors, midwives and other healthcare professionals and allied health professionals such as dentists, pharmacists and opticians.
Primary outcome	The outcome of greatest importance, usually the one in a study that the power calculation is based on.
Probabilistic analysis	In economic evaluation, this is an analysis that uses a probability distribution for each input. In contrast, see Deterministic analysis.
Product licence	An authorisation from the MHRA to market a medicinal product.
Prognosis	A probable course or outcome of a disease. Prognostic factors are patient or disease characteristics that influence the course. Good prognosis is associated with low rate of undesirable outcomes; poor prognosis is associated with a high rate of undesirable outcomes.
Prognostic test	See predictive test
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.
Prospective cohort study	An observational study where the independent variables (often exposures to a variable that cannot be randomly allocated to groups) and intervening variables (potential confounders) are measured (or, in the case of simple binary independent variables, adjudicated) at baseline. After a follow up period the dependent variables (outcomes) are measured. This allows estimation of the associations between the independent variables at baseline and the later dependent variables. A simple design would have an exposed and unexposed group (i.e. smokers and non-smokers) adjudicated at baseline and outcomes would be measured after the follow up period. As this is an observational design it is vital that potential confounders are measured at baseline and adjusted for in the analysis if necessary.
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 using a random process, such as a random numbers table or a computer-generated random sequence. It means that each individual (or each cluster 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, while the other group (the comparison or control group) receives an alternative

Term	Definition
	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 one that is routinely used in practice.
Reporting bias	See 'Publication bias'.
Resource implication	The likely impact in terms of finance, workforce or other NHS resources.
Retrospective cohort study	This is a form of cohort study that is similar to a prospective cohort study in that all data are collected in real-time. Thus the baseline data such as the exposures and covariates/cofactors is collected before the outcome data as in a prospective study. These data are usually collected by clinicians, as part of normal clinical record keeping, and the data are not normally collected with a specific study question in mind. However these studies differ from prospective cohort studies in that the researcher usually only becomes involved after all the data are collected. The researcher effectively inherits the data and 'looks back' on it: this is the origin of the term 'retrospective'. The researcher uses the retrospective data to estimate associations between variables collected at baseline and outcomes collected after a suitable follow up time. The disadvantage of this design is that very often important covariates/cofactors are absent, as when data were collected there was no study hypothesis driving variable selection. As this is an observational design the consideration of confounding factors is vital.
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)	A risk is simply the probability of an event in a group. So if 4 people experience an event in a group of 10, the risk is 4/10 = 0.4. The risk ratio is the ratio of risks across two groups. If both groups face the same level of risk, the risk ratio is 1. If the first group had a risk ratio of 2, subjects in that group would be twice as likely to have the event happen. A risk ratio of less than 1 means the outcome is less likely in the first group. The risk ratio is sometimes referred to as relative risk.
Secondary outcome	An outcome used to evaluate additional effects of the intervention deemed a priori as being less important than the primary outcomes.
Selection bias	Selection bias occurs if: a) The characteristics of the people selected for a study differ from the wider population from which they have been drawn. This special type of selection bias leads to reductions in the external validity of a study – that is, the degree to which the results can be applied to the target population. b) There are differences between groups in terms of characteristics that may influence outcome. This may arise because of non-random allocation, or because of a lack of allocation concealment in a randomised study. This type of selection bias leads to reductions in internal validity – that is the extent to which the differing outcomes are due to the differing treatments, and not due to other factors. This type

Term	Definition
	of selection bias contributes to the risk of bias assessment performed in GRADE.
Sensitivity	For diagnostic testing, this measures how well a test detects the condition for which it is testing. If a diagnostic test for a disease has high sensitivity, it is likely to pick up all or most cases of the disease in people who truly have it (that is, it will give few or no 'false negatives'). A gold standard method is used to decide who truly has the disease. For example, a gold standard method identifies 100 people who truly have the disease. The index test is then applied to these people, and if the index test is sensitive it will be able to detect all or most of these 100 people. If it identifies 98 people it will have 98% sensitivity. The gold standard is usually the best available method available but may be too impractical or expensive for routine clinical use. For prediction of later conditions, sensitivity measures how well the test predicts a later outcome. If a prediction test for an outcome (which may be the development of a disease) has high sensitivity, it is likely to identify all or most of those people who later go on to develop it (that is, it will give few or no 'false negatives'). The gold standard is simply the development of the disease, defined by a validated method, at a later time-point. For example, out of 100 people who later developed a disease, if 95 tested positive, the sensitivity would be 95%.
Sensitivity analysis	A means of representing uncertainty in the results of economic evaluations. Uncertainty may arise from missing data, imprecise estimates or methodological controversy. Sensitivity analysis also allows for exploring the generalisability of results to other settings. The analysis is repeated using different assumptions to examine the effect on the results. One-way simple sensitivity analysis (univariate analysis): each parameter is varied individually in order to isolate the consequences of each parameter on the results of the study. Multi-way simple sensitivity analysis (scenario analysis): 2 or more parameters are varied at the same time and the overall effect on the results is evaluated. Threshold sensitivity analysis: the critical value of parameters above or below which the conclusions of the study will change are identified. Probabilistic sensitivity analysis: probability distributions are assigned to the uncertain parameters and are incorporated into evaluation models based on decision analytical techniques (for example, Monte Carlo simulation).
Significance (statistical)	A result is deemed statistically significant if the probability of the result occurring by chance is less than 1 in 20 ($p<0.05$).
Specificity	For diagnostic testing, this measures how well a test detects the *absence* of the condition for which it is testing. If a diagnostic test for a disease has high specificity, it is likely to detect *no disease* in people who truly do *not* have it (that is, it will give few or no 'false positives'). A gold standard method is used to decide who truly does not have the disease. For prediction of later conditions, specificity measures how well the test predicts a later outcome. If a prediction test for an outcome (which may be the development of a disease) has high specificity, it is likely to identify all or most of those people who later do not go on to develop it (that is, it will give few or no 'false positives').
Stakeholder	An organisation with an interest in a topic that NICE is developing a guideline or piece of public health guidance on. Organisations that register as stakeholders can comment on the draft scope and the draft guidance. Stakeholders may be: • manufacturers of drugs or equipment

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

References

1. Brockhaus AC, Grouven U, Bender R. Performance of the Peto odds ratio compared to the usual odds ratio estimator in the case of rare events. Biom J. 2016; 58(6):1428-1444

2. GRADE Working Group. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group website. 2011. Available from: http://www.gradeworkinggroup.org/ Last accessed: 31/01/2020.

3. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. Journal of Clinical Epidemiology. 2016; 75:40-46

4. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual [Updated October 2018]. London. National Institute for Health and Care Excellence, 2014. Available from: https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview

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. 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

7. 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

8. Organisation for Economic Co-operation and Development (OECD). Purchasing power parities (PPP). 2012. Available from: https://www.oecd.org/sdd/prices-ppp/ Last accessed: 21/01/2020.

9. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. Journal of Clinial Epidemiolgy. 2005; 58(10):982-990

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

11. Sterne JA, Bodalia PN, Bryden PA, Davies PA, Lopez-Lopez JA, Okoli GN et al. Oral anticoagulants for primary prevention, treatment and secondary prevention of venous thromboembolic disease, and for prevention of stroke in atrial fibrillation: systematic review, network meta-analysis and cost-effectiveness analysis. Health Technology Assessment. 2017; 21(9)

12. Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Statistics in Medicine. 2002; 21(4):589-624

13. Van Houwelingen HC, Zwinderman KH, Stijnen T. A bivariate approach to meta-analysis. Statistics in Medicine. 1993; 12(24):2273-2284

14. White IR, Rapsomaniki E, Emerging Risk Factors C. Covariate-adjusted measures of discrimination for survival data. Biom J. 2015; 57(4):592-613

15. 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/