

# National Institute for Health and Care Excellence

Draft for consultation

## Stroke (update)

### Methodology

*NICE guideline*

*Methods*

*November 2018*

*Draft for consultation*

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



## **Disclaimer**

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

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

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 All NICE guidance is subject to regular review and may be updated or withdrawn.

## **Copyright**

© NICE 2018. All rights reserved. Subject to Notice of rights.

# Contents

<b>1</b>	<b>Development of the guideline</b> .....	<b>5</b>
1.1	What is a NICE guideline?.....	5
1.2	Remit.....	5
1.3	Who developed this guideline?.....	6
1.3.1	What this guideline update covers.....	6
1.3.2	What this guideline update does not cover.....	6
<b>2</b>	<b>Methods</b> .....	<b>8</b>
2.1	Developing the review questions and outcomes.....	8
2.2	Searching for evidence.....	11
2.2.1	Clinical and health economics literature searches.....	11
2.3	Identifying and analysing evidence of effectiveness .....	12
2.3.1	Inclusion and exclusion criteria .....	12
2.3.2	Type of studies.....	13
2.3.3	Methods of combining clinical studies .....	13
2.3.4	Appraising the quality of evidence by outcomes.....	16
2.3.5	Risk prediction tool studies .....	22
2.3.6	Assessing clinical importance .....	23
2.3.7	Clinical evidence statements.....	23
2.4	Identifying and analysing evidence of cost effectiveness .....	23
2.4.1	Literature review .....	24
2.4.2	Undertaking new health economic analysis.....	25
2.4.3	Cost-effectiveness criteria .....	26
2.4.4	In the absence of health economic evidence.....	26
2.5	Developing recommendations .....	26
2.5.1	Research recommendations .....	27
2.5.2	Validation process.....	27
2.5.3	Updating the guideline .....	28
2.5.4	Disclaimer .....	28
2.5.5	Funding.....	28
<b>3</b>	<b>Additional information</b> .....	<b>29</b>
<b>4</b>	<b>Acronyms and abbreviations</b> .....	<b>30</b>
<b>5</b>	<b>Glossary</b> .....	<b>32</b>
5.1	Guideline-specific terms .....	32
5.2	General terms .....	34

# 1 Development of the guideline

## 1.12 What is a NICE guideline?

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

10 NICE guidelines can:

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

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

19 We produce our guidelines using the following steps:

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

29 The guideline is made up of a collection of documents including this Methods report and a  
30 number of evidence reports covering each of the review questions included in the guideline.  
31 These can all be downloaded from NICE at [www.nice.org.uk](http://www.nice.org.uk).

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

34 NICE Pathways brings together all connected NICE guidance.

## 1.25 Remit

36 NICE conducted a surveillance review and determined tht CG68 should be updated in a  
37 number of areas. NICE commissioned the NGC to produce the guideline.

38 The remit for this guideline is:

39 To partially update the clinical guideline on Stroke CG68.

### 1.3.1 Who developed this guideline?

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

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

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

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

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

#### 1.3.24 What this guideline update covers

25 This guideline is a partial update of the NICE guideline Stroke and transient ischaemic attack  
26 in over 16s: diagnosis and initial management. It updates a number of recommendations  
27 while also investigating clinical areas not addressed by the previous guideline. The  
28 population covered is people over 16 with suspected or confirmed transient ischaemic attack  
29 (TIA) or completed strokes – that is, an acute neurological event presumed to be vascular in  
30 origin and causing cerebral ischaemia, cerebral infarction or cerebral haemorrhage. The  
31 clinical areas included in this update are:

- 32 • rapid diagnosis and initial management of TIA,
- 33 • brain imaging after TIA,
- 34 • endovascular treatments for people with acute stroke,
- 35 • blood pressure control for haemorrhagic stroke,
- 36 • decompressive hemicraniectomy,
- 37 • early mobilisation and optimum head positioning of people with acute stroke.

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

#### 1.3.20 What this guideline update does not cover

41 This guideline does not cover management beyond the acute stage of stroke, nor stroke  
42 rehabilitation.

43 Additionally, areas from Stroke and transient ischaemic attack in over 16s: diagnosis and  
44 initial management (CG68) that have not been updated are:

- 45 • The rapid recognition of symptoms and diagnosis

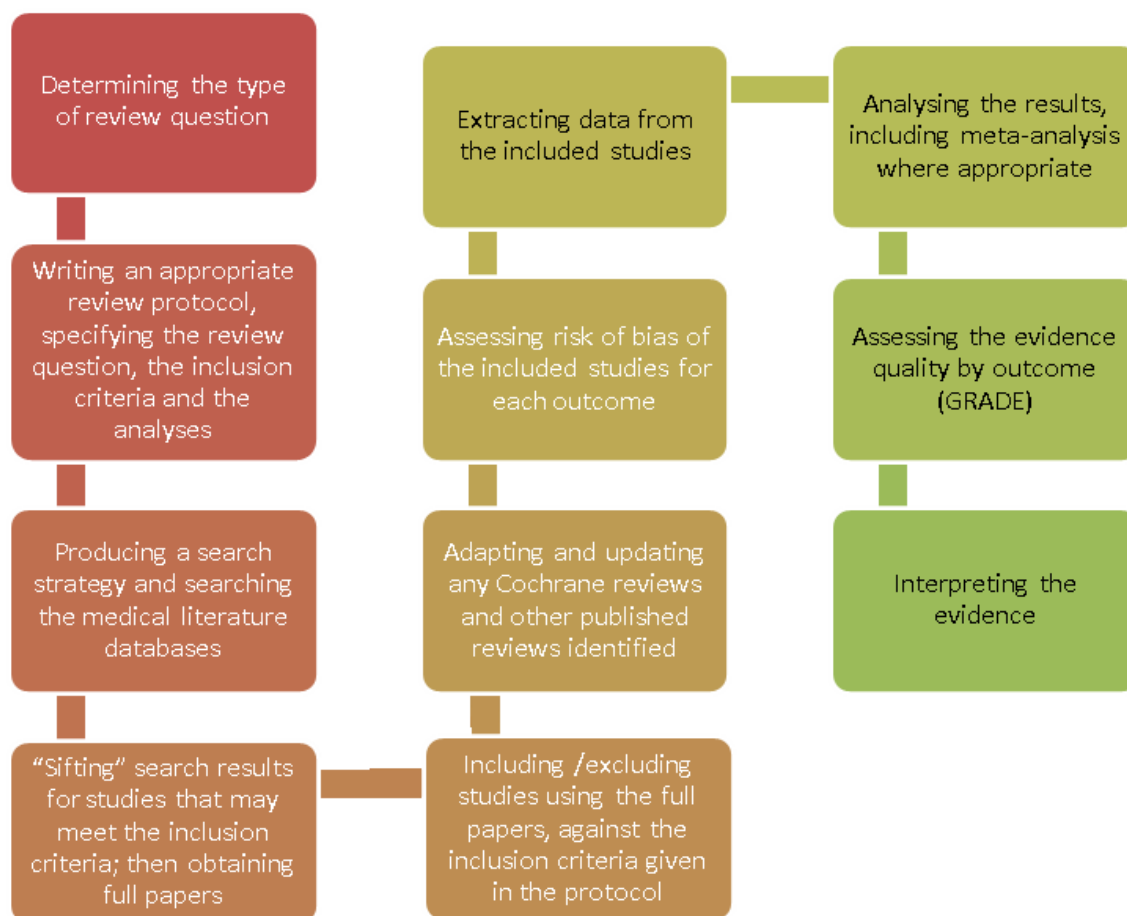
- 1 ○ Pre-hospital prompt recognition of symptoms of TIA and stroke symptoms
- 2 ● Early carotid imaging in people with acute non-disabling stroke or TIA
- 3 ● Urgent carotid endarterectomy and carotid stenting in people with carotid stenosis
- 4 ● Specialist care in acute stroke
- 5 ○ Specialist stroke units
- 6 ○ Brain imaging for the early assessment of people with acute stroke
- 7 ● Pharmacological treatments for people with acute stroke
- 8 ○ Thrombolysis in people with acute ischaemic stroke
- 9 ○ Aspirin and anticoagulant treatment in people with acute ischaemic stroke
- 10 ○ Antiplatelet and anticoagulant treatment in people with acute venous stroke
- 11 ○ Antiplatelet and anticoagulant treatment in people with stroke due to arterial dissection
- 12 ○ Antiplatelet and anticoagulant treatment in people with acute stroke due to antiphospholipid syndrome
- 13
- 14 ○ Reversal of anticoagulation treatment in people with haemorrhagic stroke
- 15 ○ Anticoagulation treatment for other comorbidities in people with acute stroke
- 16 ○ Statin treatment in people with acute stroke
- 17 ● Maintenance or restoration of homeostasis
- 18 ○ Supplemental oxygen therapy
- 19 ○ Blood sugar control
- 20 ○ Blood pressure control in ischaemic stroke
- 21 ● Nutrition and hydration
- 22 ○ Assessment of swallowing function
- 23 ○ Timing of enteral feeding
- 24 ○ Oral nutritional supplementation
- 25 ● Avoidance of aspiration pneumonia
- 26 ● Surgical referral for acute intracerebral haemorrhage

## 2.1 Methods

2 This report sets out in detail the methods used to review the evidence and to develop the  
3 recommendations that are presented in each of the evidence reviews for this guideline. This  
4 guidance was developed in accordance with the methods outlined in the NICE guidelines  
5 manual, 2014 version.<sup>2</sup>

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

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



### 2.1.0 Developing the review questions and outcomes

11 Review questions were developed using a PICO framework (population, intervention,  
12 comparison and outcome) for intervention reviews and diagnostic test-and-treat reviews; and  
13 using population, risk assessment tools, reference standard and outcomes for risk prediction  
14 reviews.

15 This use of a framework guided the literature searching process, critical appraisal and  
16 synthesis of evidence, and facilitated the development of recommendations by the guideline  
17 committee. The review questions were drafted by the NGC technical team and refined and  
18 validated by the committee. The questions were based on the key clinical areas identified in  
19 the scope. The off-label use of intravenous thrombolysis in people aged 80 and over was



1 included in the scope but has not been included in the guideline. This is because the SPC for  
2 alteplase has been changed to include those over 80.

3 A total of 8 review questions were identified.

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

6 **Table 1: Review questions**

Evidence report	Type of review	Review questions	Outcomes
A	Intervention	Should people with a suspected TIA be advised to take aspirin prior to assessment in a TIA clinic?	<p>Critical outcomes:</p> <ul style="list-style-type: none"> <li>• Risk of stroke</li> <li>• Mortality</li> </ul> <p>Important outcomes:</p> <ul style="list-style-type: none"> <li>• Intra-cranial haemorrhage</li> <li>• Major bleeding complications</li> <li>• Functional outcomes <ul style="list-style-type: none"> <li>◦ Modified Rankin Scale (mRS) score</li> </ul> </li> <li>• Quality of life</li> </ul>
B	Risk prediction	How accurately do scoring systems predict the risks of future ischaemic stroke or TIA within the first 7 days in people with suspected TIA?	<p>Statistical outputs may include:</p> <ul style="list-style-type: none"> <li>• Discrimination (area under curve [c statistic])</li> <li>• Calibration (<math>R^2</math>, Brier Score, Hosmer-Lemeshow test statistic; Somers' D statistic),</li> <li>• Calibration plot</li> <li>• Reclassification</li> </ul> <p><u>These will be assessed for the following outcomes:</u></p> <p>Critical clinical effectiveness outcomes:</p> <ul style="list-style-type: none"> <li>• Risk of stroke</li> <li>• Mortality</li> </ul> <p>Important clinical effectiveness outcomes:</p> <ul style="list-style-type: none"> <li>• Functional outcomes <ul style="list-style-type: none"> <li>◦ mRS score</li> </ul> </li> <li>• Quality of life</li> </ul>
C	Diagnostic test and treat	After TIA, what is the optimal brain imaging strategy?	<p>Critical outcomes:</p> <ul style="list-style-type: none"> <li>• Risk of stroke</li> <li>• Mortality</li> </ul> <p>Important outcomes:</p> <ul style="list-style-type: none"> <li>• Functional outcomes <ul style="list-style-type: none"> <li>◦ mRS score</li> </ul> </li> <li>• Quality of life</li> <li>• Change in diagnosis or clinical management</li> </ul>
D	Intervention	What is the clinical and cost	Critical outcomes:

Evidence report	Type of review	Review questions	Outcomes
		effectiveness of endovascular therapy (EVT) with or without intravenous thrombolysis versus intravenous thrombolysis to improve outcomes?	<ul style="list-style-type: none"> <li>• mRS 0–2 or ordinal shift</li> <li>• Mortality</li> </ul> <p>Important outcomes:</p> <ul style="list-style-type: none"> <li>• Intracerebral haemorrhage</li> <li>• Symptomatic intracerebral haemorrhage</li> <li>• Patient reported outcome measures</li> <li>• Quality of life</li> <li>• Length of stay in hospital</li> <li>• Procedural complications</li> </ul>
E	Intervention	What is the safety and efficacy of measures to lower blood pressure versus standard treatment in people with acute intracerebral haemorrhage?	<p>Critical outcomes:</p> <ul style="list-style-type: none"> <li>• mRS 0–2 or ordinal shift</li> <li>• Mortality</li> </ul> <p>Important outcomes:</p> <ul style="list-style-type: none"> <li>• Symptomatic cerebral ischemia</li> <li>• Haemorrhage expansion</li> <li>• Neurological deterioration</li> <li>• Adverse events (renal failure, cord infarction, myocardial infarction)</li> <li>• Quality of life</li> <li>• Achieving blood pressure target</li> </ul>
F	Intervention	Does early mobilisation versus treatment as usual reduce mortality and morbidity in people with acute stroke?	<p>Critical outcomes:</p> <ul style="list-style-type: none"> <li>• mRS 0–2 or ordinal shift</li> <li>• Barthel score if mRS not reported</li> <li>• Mortality</li> </ul> <p>Important outcomes:</p> <ul style="list-style-type: none"> <li>• Recurrent stroke</li> <li>• Adverse events (PE/DVT/pressure sores/pneumonia/falls)</li> <li>• Quality of life</li> <li>• Length of stay</li> <li>• Acute neurological deterioration (worsening of NIHSS)</li> </ul>
G	Intervention	What is the optimal head positioning (sitting up or lying flat) after a stroke to improve outcomes?	<p>Critical outcomes:</p> <ul style="list-style-type: none"> <li>• mRS 0–2 or ordinal shift</li> <li>• Barthel score if mRS not reported</li> <li>• Mortality</li> </ul> <p>Important outcomes:</p>

Evidence report	Type of review	Review questions	Outcomes
			<ul style="list-style-type: none"> <li>• Recurrent stroke</li> <li>• Adverse events (PE/DVT/pressure sores/pneumonia/falls)</li> <li>• Quality of life</li> <li>• Length of stay</li> <li>• Acute neurological deterioration (worsening of NIHSS)</li> </ul>
H	Intervention	Which patients should be referred for decompressive hemicraniectomy?	<p>Critical outcomes:</p> <ul style="list-style-type: none"> <li>• Mortality</li> <li>• mRS 0–3 or ordinal shift</li> </ul> <p>Important outcomes:</p> <ul style="list-style-type: none"> <li>• Quality of life</li> </ul>

## 2.2.1 Searching for evidence

### 2.2.1.2 Clinical and health economics literature searches

3 Systematic literature searches were undertaken to identify all published clinical and health  
4 economic evidence relevant to the review questions. Searches were undertaken according to  
5 the parameters stipulated within the NICE guidelines manual. {National Institute for Health  
6 and Care Excellence, 2014 #23} Databases were searched using relevant medical subject  
7 headings, free-text terms and study-type filters where appropriate. Where possible, searches  
8 were restricted to papers published in English. Studies published in languages other than  
9 English were not reviewed. Dates for each search can be found in appendix B of the  
10 individual reviews.

11 Papers published or added to databases after these dates were not considered. If new  
12 evidence, falling outside of the timeframe for the guideline searches, is identified, for  
13 example in consultation comments received from stakeholders, the impact on the guideline  
14 will be considered, and any further action agreed between NGC and NICE staff with a quality  
15 assurance role.

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

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

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

## 2.3.1 Identifying and analysing evidence of effectiveness

- 2 Research fellows conducted the tasks listed below, which are described in further detail in  
3 the rest of this section:
- 4 • Identified potentially relevant studies for each review question from the relevant search  
5 results by reviewing titles and abstracts. Full papers were then obtained.
  - 6 • Reviewed full papers against prespecified inclusion and exclusion criteria to identify  
7 studies that addressed the review question in the appropriate population, and reported on  
8 outcomes of interest (review protocols are included in an appendix to each of the  
9 evidence reports).
  - 10 • Critically appraised relevant studies using the appropriate study design checklist as  
11 specified in the NICE guidelines manual.<sup>2</sup> Risk prediction studies were critically appraised  
12 using the PROBAST checklist.
  - 13 • Extracted key information about interventional study methods and results using 'Evibase',  
14 NGC's purpose-built software. Evibase produces summary evidence tables, including  
15 critical appraisal ratings. Key information about non-interventional study methods and  
16 results was manually extracted onto standard evidence tables and critically appraised  
17 separately (evidence tables are included in an appendix to each of the evidence reports).
  - 18 • Generated summaries of the evidence by outcome. Outcome data were combined,  
19 analysed and reported according to study design:
    - 20 ○ Randomised data were meta-analysed where appropriate and reported in GRADE  
21 profile tables.
    - 22 ○ Risk stratification data were presented for each study separately and reported in  
23 adapted GRADE profile tables.
  - 24 • A sample of a minimum of 10% of the abstract lists of the first 3 sifts by new reviewers  
25 and those for complex review questions (for example, prognostic reviews) were double-  
26 sifted by a senior research fellow and any discrepancies were rectified. All of the evidence  
27 reviews were quality assured by a senior research fellow. This included checking:
    - 28 ○ papers were included or excluded appropriately
    - 29 ○ a sample of the data extractions
    - 30 ○ correct methods were used to synthesise data
    - 31 ○ a sample of the risk of bias assessments.

### 2.3.1.2 Inclusion and exclusion criteria

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

38 The key population inclusion criterion was:

- 39 • People over 16 with suspected or confirmed TIAs or completed strokes

40 The key population exclusion criterion was:

- 41
- 42 • People with subarachnoid haemorrhage

43 Conference abstracts were not automatically excluded from any review. The abstracts were  
44 initially assessed against the inclusion criteria for the review question and further processed  
45 when a full publication was not available for that review question. If the abstracts were  
46 included the authors were contacted for further information. No relevant conference abstracts

1 were identified for this guideline. Narrative literature reviews, posters, letters, editorials,  
2 comment articles, unpublished studies and studies not in English were excluded.

### 2.3.23 Type of studies

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

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

17 For the diagnostic review question on brain imaging after TIA, only diagnostic RCTs were  
18 considered for inclusion. This was because the committee considered that diagnostic  
19 accuracy outcomes would not answer the review question as they would not provide  
20 information on the downstream effects after imaging in terms of observed stroke, mortality,  
21 functional outcome and changes decision making and clinical management. Thus diagnostic  
22 test and treat RCTs were prioritised in the review to allow comparison of imaging strategies.  
23 For the risk prediction review question, prospective and retrospective validation cohort  
24 studies were included. Case-control studies were not included.

### 2.3.35 Methods of combining clinical studies

#### 2.3.3.26 Data synthesis for intervention reviews

27 Where possible, meta-analyses were conducted using Cochrane Review Manager  
28 (RevMan5)<sup>5</sup> software to combine the data given in all studies for each of the outcomes of  
29 interest for the review question.

30 For some questions additional stratification was used, and this is documented in the  
31 individual review question protocols in each evidence report.

#### 2.3.3.1.32 Analysis of different types of data

##### 33 Dichotomous outcomes

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

- 36 • mortality
- 37 • mRS score 0-2
- 38 • recurrent stroke
- 39 • adverse events.

40 The committee discussed the most appropriate definition of a good functional outcome based  
41 on the mRS scale. It was agreed that for most of the clinical areas that a score of 0, 1 or 2 on  
42 this scale, which corresponds to functional independence, should be used as the definition of  
43 a good functional outcome. However, in the case of people being considered for  
44 decompressive hemicraniectomy, or for those with large vessel occlusion in the proximal

1 posterior circulation being considered for thrombectomy it was agreed that the definition of a  
2 good outcome should be an mRS score of 0, 1, 2 or 3 because of the higher likelihood of  
3 death without intervention.

4 The absolute risk difference was also calculated using GRADEpro<sup>1</sup> software, using the  
5 median event rate in the control arm of the pooled results.

6 For binary variables where there were zero events in either arm or a less than 1% event rate,  
7 Peto odds ratios, rather than risk ratios, were calculated. Peto odds ratios are more  
8 appropriate for data with a low number of events. If there were zero events in both arms a  
9 risk difference was calculated to produce a forest plot and absolute risk estimate.

## 10 **Continuous outcomes**

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

- 13 • health-related quality of life (HRQoL)
- 14 • length of stay in hospital
- 15 • symptom scales (such as visual analogue scale)

16 The means and standard deviations of continuous outcomes are required for meta-analysis.  
17 However, in cases where standard deviations were not reported, the standard error was  
18 calculated if the p values or 95% confidence intervals (95% CI) were reported, and meta-  
19 analysis was undertaken with the mean and standard error using the generic inverse  
20 variance method in Cochrane Review Manager (RevMan5)<sup>5</sup> software. Where p values were  
21 reported as 'less than', a conservative approach was undertaken. For example, if a p value  
22 was reported as 'p≤0.001', the calculations for standard deviations were based on a p value  
23 of 0.001. If these statistical measures were not available then the methods described in  
24 section 16.1.3 of the Cochrane Handbook (version 5.1.0, updated March 2011) were applied.

## 25 **Ordinal shift analysis**

26 For functional outcome measured by the modified Rankin Scale (mRS) both a dichotomised  
27 and an ordinal shift result were reported where available. The use of the common odds ratio  
28 derived from ordinal shift analysis was included in order to analyse changes in the  
29 distribution of participants over the full range of the scale. This is commonly reported in  
30 stroke trials and retains all information captured by this ordinal outcome scale, unlike the  
31 dichotomised outcome. Where possible, the adjusted value of the common odds ratio was  
32 reported in preference to the unadjusted value. When more than one study reported this  
33 outcome for a given comparison the data were meta-analysed.

## 34 **Generic inverse variance**

35 If a study reported only the summary statistic and 95% CI the generic-inverse variance  
36 method was used to enter data into RevMan5.<sup>5</sup> If the control event rate was reported this  
37 was used to generate the absolute risk difference in GRADEpro.<sup>1</sup> If multivariate analysis was  
38 used to derive the summary statistic but no adjusted control event rate was reported no  
39 absolute risk difference was calculated.

### 2.3.3.1.20 **Heterogeneity**

41 Statistical heterogeneity was assessed for each meta-analysis estimate by considering the  
42 chi-squared test for significance at p<0.1 or an I-squared (I<sup>2</sup>) inconsistency statistic (with an I-  
43 squared value of more than 50% indicating significant heterogeneity) as well as the  
44 distribution of effects. Where significant heterogeneity was present, predefined subgrouping  
45 of studies was carried out; this is documented in the individual review question protocols.

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

6 Where heterogeneity was found, all subgrouping strategies were applied, the strategies were  
7 utilised independently, so subunits of subgroups were not created.

8 If all predefined strategies of subgrouping were unable to explain statistical heterogeneity  
9 within each derived subgroup, then a random effects (DerSimonian and Laird) model was  
10 employed to the entire group of studies in the meta-analysis. A random-effects model  
11 assumes a distribution of populations, rather than a single population. This leads to a  
12 widening of the confidence interval around the overall estimate, thus providing a more  
13 realistic interpretation of the true distribution of effects across more than 1 population. If,  
14 however, the committee considered the heterogeneity was so large that meta-analysis was  
15 inappropriate, then the results were described narratively.

### **2.3.3.1.36 Further analysis**

17 Where studies had used a crossover design, paired continuous data were extracted where  
18 possible, and forest plots were generated in RevMan5<sup>5</sup> with the generic inverse variance  
19 function. When a crossover study had categorical data and the number of subjects with an  
20 event in both interventions was known, the standard error (of the log of the risk ratio) was  
21 calculated using the simplified Mantel–Haenszel method for paired outcomes. Forest plots  
22 were also generated in RevMan5<sup>5</sup> with the generic inverse variance function. If paired  
23 continuous or categorical data were not available from the crossover studies, the separate  
24 group data were analysed in the same way as data from parallel groups, on the basis that  
25 this approach would overestimate the confidence intervals and thus artificially reduce study  
26 weighting resulting in a conservative effect. Where a meta-analysis included a mixture of  
27 studies using both paired and parallel group approaches, all data were entered into  
28 RevMan5<sup>5</sup> using the generic inverse variance function.

### **2.3.3.29 Data synthesis for diagnostic test accuracy reviews**

#### **2.3.3.2.10 Diagnostic RCTs**

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

#### **2.3.3.32 Data synthesis for risk prediction tools**

43 For evidence reviews on risk prediction tools, results were presented separately for  
44 discrimination and calibration. The discrimination data were analysed according to the  
45 principles of data synthesis for diagnostic accuracy studies. Discrimination data can indicate  
46 the clinical impact of using a risk prediction tool in clinical practice, and therefore these data  
47 were prioritised for inclusion and decision-making.

- 1 The analysis used the area under the receiver operating characteristics (ROC) curve (AUC)
- 2 or C-statistic value. AUC data for each study were extracted, for each risk prediction tool.
- 3 The AUC describes the overall diagnostic accuracy across the full range of thresholds. The
- 4 following criteria were used for evaluating AUCs:
  - 5 •  $\leq 0.50$ : worse than chance
  - 6 • 0.50–0.60: very poor
  - 7 • 0.61–0.70: poor
  - 8 • 0.71–0.80: moderate
  - 9 • 0.81–0.92: good
  - 10 • 0.91–1.00: excellent or perfect test.
- 11 Heterogeneity or inconsistency amongst studies was visually inspected.
- 12 Calibration data such as r-squared ( $R^2$ ), if reported, were presented separately to the
- 13 discrimination data. Meta-analysis was considered but not performed due to insufficient data
- 14 reported for each of the risk prediction tools. The results were presented for each study
- 15 separately along with the quality rating for the study.

## 2.3.46 Appraising the quality of evidence by outcomes

### 2.3.4.17 Intervention reviews

- 18 The evidence for outcomes from the included RCTs (including diagnostic RCTs) and, where
- 19 appropriate, non-randomised intervention studies, were evaluated and presented using an
- 20 adaptation of the 'Grading of Recommendations Assessment, Development and Evaluation
- 21 (GRADE) toolbox' developed by the international GRADE working group
- 22 (<http://www.gradeworkinggroup.org/>). The software (GRADEpro<sup>1</sup>) developed by the GRADE
- 23 working group was used to assess the quality of each outcome, taking into account individual
- 24 study quality and the meta-analysis results.
- 25 Each outcome was first examined for each of the quality elements listed and defined in Table
- 26 2.

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

Quality element	Description
Risk of bias	Limitations in the study design and implementation may bias the estimates of the treatment effect. Major limitations in studies decrease the confidence in the estimate of the effect. Examples of such limitations are selection bias (often due to poor allocation concealment), performance and detection bias (often due to a lack of blinding of the patient, healthcare professional or assessor) and attrition bias (due to missing data causing systematic bias in the analysis).
Indirectness	Indirectness refers to differences in study population, intervention, comparator and outcomes between the available evidence and the review question.
Inconsistency	Inconsistency refers to an unexplained heterogeneity of effect estimates between studies in the same meta-analysis.
Imprecision	Results are imprecise when studies include relatively few patients and few events (or highly variable measures) and thus have wide confidence intervals around the estimate of the effect relative to clinically important thresholds. 95% confidence intervals denote the possible range of locations of the true population effect at a 95% probability, and so wide confidence intervals may denote a result that is consistent with conflicting interpretations (for example a result may be consistent with both clinical benefit AND clinical harm) and thus be imprecise.
Publication bias	Publication bias is a systematic underestimate or overestimate of the underlying beneficial or harmful effect due to the selective publication of studies. A closely



Quality element	Description
	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.

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

#### 2.3.4.1.14 Risk of bias

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

14 **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: <ul style="list-style-type: none"> <li>• knowledge of that participant's likely prognostic characteristics, and</li> <li>• a desire for one group to do better than the other.</li> </ul>
Performance and detection bias (lack of blinding of patients and healthcare professionals)	Patients, caregivers, those adjudicating or recording outcomes, and data analysts should not be aware of the arm to which patients are allocated. Knowledge of the group can influence: <ul style="list-style-type: none"> <li>• the experience of the placebo effect</li> <li>• performance in outcome measures</li> <li>• the level of care and attention received, and</li> <li>• the methods of measurement or analysis</li> </ul> all of which can contribute to systematic bias.
Attrition bias	Attrition bias results from an unaccounted for loss of data beyond a certain level (a differential of 10% between groups). Loss of data can occur when participants are compulsorily withdrawn from a group by the researchers (for example, when a per-protocol approach is used) or when participants do not attend assessment sessions. If the missing data are likely to be different from the data of those remaining in the groups, and there is a differential rate of such missing data from groups, systematic attrition bias may result.
Selective outcome reporting	Reporting of some outcomes and not others on the basis of the results can also lead to bias, as this may distort the overall impression of efficacy.
Other limitations	For example: <ul style="list-style-type: none"> <li>• Stopping early for benefit observed in randomised trials, in particular in the absence of adequate stopping rules.</li> </ul>

Limitation	Explanation
	<ul style="list-style-type: none"> <li>• Use of unvalidated patient-reported outcome measures.</li> <li>• Lack of washout periods to avoid carry-over effects in crossover trials.</li> <li>• Recruitment bias in cluster-randomised trials.</li> </ul>

#### 2.3.4.1.21 *Indirectness*

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

#### 2.3.4.1.35 *Inconsistency*

16 Inconsistency refers to an unexplained heterogeneity of results for an outcome across  
17 different studies. When estimates of the treatment effect across studies differ widely, this  
18 suggests true differences in the underlying treatment effect, which may be due to differences  
19 in populations, settings or doses for example. When heterogeneity existed within an outcome  
20 (chi-squared  $p < 0.1$ , or  $I^2 > 50\%$ ), but no plausible explanation could be found, the quality of  
21 evidence for that outcome was downgraded. Inconsistency for that outcome was given a  
22 'serious' score of -1 if the  $I^2$  was 50–74%, and a 'very serious' score of -2 if the  $I^2$  was 75%  
23 or more.

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

29 Since the inconsistency score was based on the pooled effect estimate from the meta-  
30 analysis, the score represented the whole outcome and so weighted averaging across  
31 studies was not necessary.

#### 2.3.4.1.42 *Imprecision*

33 The criteria applied for imprecision were based on the 95% CIs for the pooled estimate of  
34 effect, and the minimal important differences (MID) for the outcome. The MIDs are the  
35 threshold for appreciable benefits and harms, separated by a zone either side of the line of  
36 no effect where there is assumed to be no clinically important effect. If either end of the 95%  
37 CI of the overall estimate of effect crossed 1 of the MID lines, imprecision was regarded as  
38 serious and a 'serious' score of -1 was given. This was because the overall result, as  
39 represented by the span of the confidence interval, was consistent with 2 interpretations as  
40 defined by the MID (for example, both no clinically important effect and clinical benefit were  
41 possible interpretations). If both MID lines were crossed by either or both ends of the 95% CI  
42 then imprecision was regarded as very serious and a 'very serious' score of -2 was given.  
43 This was because the overall result was consistent with all 3 interpretations defined by the  
44 MID (no clinically important effect, clinical benefit and clinical harm). This is illustrated in  
45 Figure 2. As for inconsistency, since the imprecision score was based on the thepooled

1 effect estimate from the meta-analysis, the score represented the whole outcome and so  
2 weighted averaging across studies was not necessary.

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

14 In this guideline, MIDs found in the literature were used to assess imprecision for the EQ-5D  
15 and SF-36 measures of health-related quality of life. These values are displayed below:

16 **Table 4: MIDs used to assess imprecision for the EQ-5D and SF-36 measures**

MIDs for assessing between group differences Outcome	MID for imprecision	MID for clinical importance	Source
SF-36 <sup>^</sup>	Physical component summary: 2 Mental component summary: 3 Physical functioning: 3 Role-physical: 3 Bodily pain: 3 General health: 2 Vitality: 2 Social functioning: 3 Role-emotional: 4 Mental health: 3		User’s manual for the SF-36v2 Health Survey, Third Edition
EQ5D*	GRADE defaults	0.03	NICE agreed for use in Low Back Pain & Low back Pain committee opinion

17 <sup>^</sup>Note: the SF-12 manual does not specify MIDs. It does however signpost to the SF-36  
18 manual for guidance on interpretation, therefore in this guideline we used the same MIDs for  
19 the SF-12.

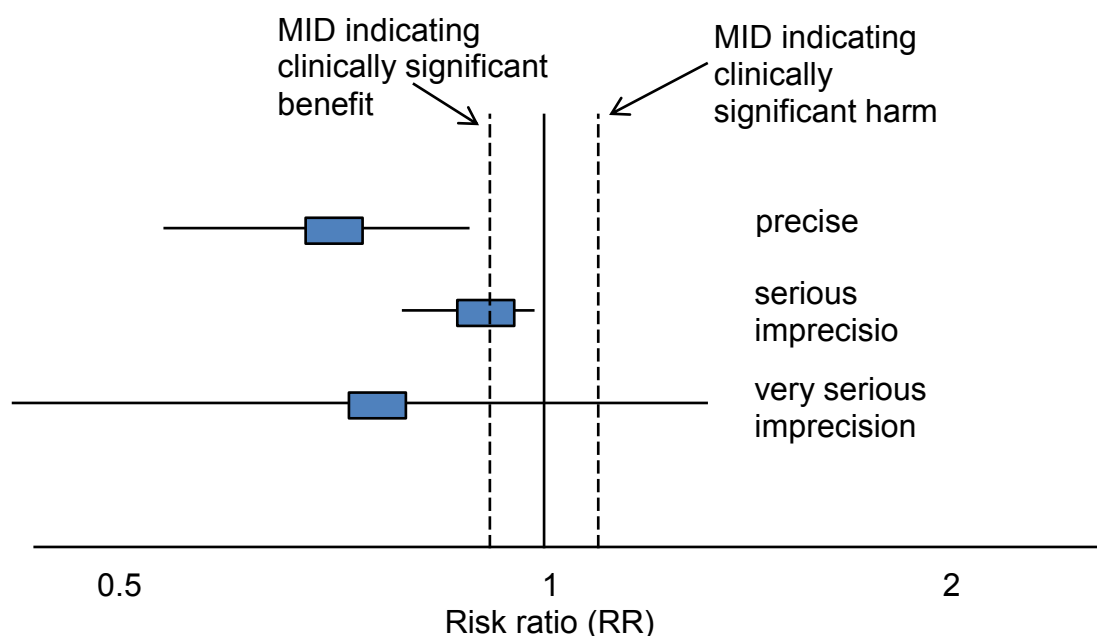
20 \* Note: this is not based on the literature and was a pragmatic decision for this guideline  
21 based on the SF-36 MIDs.

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

- 24 • For categorical outcomes the MIDs were taken to be RRs of 0.8 and 1.25. For ‘positive’  
25 outcomes such as ‘patient satisfaction’, the RR of 0.8 is taken as the line denoting the  
26 boundary between no clinically important effect and a clinically significant harm, whilst the  
27 RR of 1.25 is taken as the line denoting the boundary between no clinically important  
28 effect and a clinically significant benefit. For ‘negative’ outcomes such as ‘bleeding’, the  
29 opposite occurs, so the RR of 0.8 is taken as the line denoting the boundary between no  
30 clinically important effect and a clinically significant benefit, whilst the RR of 1.25 is taken  
31 as the line denoting the boundary between no clinically important effect and a clinically  
32 significant harm.

- 1 • For mortality any change was considered to be clinically important and the imprecision  
2 was assessed on the basis of the whether the confidence intervals crossed the line of no  
3 effect, that is whether the result was consistent with both benefit and harm.
- 4 • For continuous outcome variables the MID was taken as half the median baseline  
5 standard deviation of that variable, across all studies in the meta-analysis. Hence the MID  
6 denoting the minimum clinically significant benefit was positive for a 'positive' outcome (for  
7 example, a quality of life measure where a higher score denotes better health), and  
8 negative for a 'negative' outcome (for example, a visual analogue scale [VAS] pain score).  
9 Clinically significant harms were the converse of these. If baseline values were  
10 unavailable, then half the median comparator group standard deviation of that variable  
11 were taken as the MID.
- 12 • If standardised mean differences were been used, then the MID was set at the absolute  
13 value of +0.5. This follows because standardised mean differences are mean differences  
14 normalised to the pooled standard deviation of the 2 groups, and are thus effectively  
15 expressed in units of 'numbers of standard deviations'. The 0.5 MID value in this context  
16 therefore indicates half a standard deviation, the same definition of MID as used for non-  
17 standardised mean differences.
- 18 The default MID value was subject to amendment after discussion with the committee. If the  
19 committee decided that the MID level should be altered, after consideration of absolute as  
20 well as relative effects, this was allowed, provided that any such decision was not influenced  
21 by any bias towards making stronger or weaker recommendations for specific outcomes. The  
22 rationale for any such decision will be justified within the committee's discussion section of  
23 the relevant review.
- 24 For this guideline, no appropriate MIDs for continuous or dichotomous outcomes were found  
25 in the literature, and so the default method was adopted.

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



#### 2.3.4.1.51 Overall grading of the quality of clinical evidence

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

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

15 **Table 5: 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

### 2.3.5.1 Risk prediction tool studies

#### 2.3.5.1.2 Risk of bias and indirectness

3 Risk of bias and indirectness of evidence for prognostic risk tool data were evaluated by  
 4 study using the Prediction Study Risk of Bias Assessment tool (PROBAST) checklist which is  
 5 summarised in Table 6. PROBAST is still under development and the version used in this  
 6 guideline was acquired from the study author and adapted. One item concerning whether all  
 7 predictors were available at the time the risk tool would be used in practice was excluded  
 8 from the risk of bias assessment, and instead was incorporated into an assessment of  
 9 indirectness. Where the information required to complete PROBAST domains was not  
 10 reported in publications, this was taken into account for the risk of bias assessment. If the  
 11 majority of information was available but one domain had limited information there was no  
 12 obligate downgrade for risk of bias. If more than one domain had limited or no information to  
 13 inform its assessment, the risk of bias was given a 'serious' rating of -1. If very limited or no  
 14 information was provided for the majority of domains for the study, the risk of bias was given  
 15 a 'very serious' rating of -2. Ratings were assessed for the validation of risk tools; no ratings  
 16 were provided for the original development phase of the tools.

17 **Table 6: Summary of PROBAST**

Quality element	Description of cases where the quality measure would be downgraded
Participant selection	If case control rather than cohort, RCT or nested case-control, or if potential for selection bias
Predictors	If predictors were not defined or assessed in a similar way for all participants, if assessors were not blinded to outcome data
Outcome	If outcome was not defined or assessed in a similar way for all participants, if assessors were not blinded to predictor information, if predictors were included in the outcome definition
Sample size and participant flow	If there was a low event rate relative to the number of predictors, if there was an inappropriate time interval between predictor assessment and outcome, if risk of selection bias
Analysis	If analysis is not appropriate for the design, if relevant outcome measures were not reported
Applicability	If concerns that the study participants, predictors or outcome are dissimilar to those specified in the review protocol

#### 2.3.5.1.18 Inconsistency

19 Inconsistency refers to an unexplained heterogeneity of results for an outcome across  
 20 different studies. For discrimination, inconsistency was assessed by inspection of the AUC  
 21 value using the point estimates and 95% CIs of the individual studies.

22 For calibration scores inconsistency was assessed by examining whether the unpooled  
 23 results of individual studies had different findings, for example one showing poor and another  
 24 showing good calibration.

#### 2.3.5.1.25 Imprecision

26 For discrimination, imprecision was assessed according to the range of point estimates or, if  
 27 only one study contributed to the evidence, the 95% CI around the single study estimate of  
 28 the AUC. As a general rule (after discussion with the guideline committee) a variation of 0–  
 29 0.2 was considered precise, 0.2–0.4 serious imprecision, and >0.4 very serious imprecision.  
 30 Imprecision was assessed on the primary outcome measure for decision-making.

31 Imprecision could not be estimated for calibration scores.

### 2.3.5.1.31 Overall grading

2 Quality rating started at High for prospective and retrospective cohort studies, and each  
3 major limitation (risk of bias, indirectness, inconsistency and imprecision) was associated  
4 with a score of -1 or -2 and were summed to find the overall grading, with a minimum grade  
5 of Very Low, as explained for intervention reviews.

### 2.3.66 Assessing clinical importance

7 The committee assessed the evidence by outcome in order to determine if there was, or  
8 potentially was, a clinically important benefit, a clinically important harm or no clinically  
9 important difference between interventions. To facilitate this, binary outcomes were  
10 converted into absolute risk differences (ARDs) using GRADEpro<sup>1</sup> software: the median  
11 control group risk across studies was used to calculate the ARD and its 95% CI from the  
12 pooled risk ratio.

13 The assessment of clinical benefit, harm, or no benefit or harm was based on the point  
14 estimate of absolute effect for intervention studies, which was standardised across the  
15 reviews. The committee considered for most of the outcomes in the intervention reviews that  
16 if at least 100 more participants per 1000 (10%) achieved the outcome of interest in the  
17 intervention group compared to the comparison group for a positive outcome then this  
18 intervention was considered beneficial. The same point estimate but in the opposite direction  
19 applied for a negative outcome. For the critical outcome of mortality the committee  
20 considered any reduction represented a clinical benefit. For adverse events 50 events or  
21 more per 1000 (5%) represented clinical harm. For continuous outcomes if the mean  
22 difference was greater than the minimally important difference (MID) then this represented a  
23 clinical benefit or harm.

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

### 2.3.78 Clinical evidence statements

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

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

## 2.49 Identifying and analysing evidence of cost effectiveness

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

1 uncertainties must be offset by a compelling argument in favour of the recommendation. The  
2 cost impact or savings potential of a recommendation should not be the sole reason for the  
3 committee's decision.<sup>2</sup>

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

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

#### 2.4.1.18 Literature review

9 The health economists:

- 10 • Identified potentially relevant studies for each review question from the health economic  
11 search results by reviewing titles and abstracts. Full papers were then obtained.
- 12 • Reviewed full papers against prespecified inclusion and exclusion criteria to identify  
13 relevant studies (see below for details).
- 14 • Critically appraised relevant studies using economic evaluations checklists as specified in  
15 the NICE guidelines manual.<sup>2</sup>
- 16 • Extracted key information about the studies' methods and results into health economic  
17 evidence tables (which can be found in appendices to the relevant evidence reports).
- 18 • Generated summaries of the evidence in NICE health economic evidence profile tables  
19 (included in the relevant evidence report for each review question) – see below for details.

#### 2.4.1.20 Inclusion and exclusion criteria

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

25 Studies that only reported cost per hospital (not per patient), or only reported average cost  
26 effectiveness without disaggregated costs and effects were excluded. Literature reviews,  
27 abstracts, posters, letters, editorials, comment articles, unpublished studies and studies not  
28 in English were excluded. Studies published before 2002 and studies from non-OECD  
29 countries or the USA were also excluded, on the basis that the applicability of such studies to  
30 the present UK NHS context is likely to be too low for them to be helpful for decision-making.

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

36 For more details about the assessment of applicability and methodological quality see Table  
37 7 below and the economic evaluation checklist (appendix H of the NICE guidelines manual<sup>2</sup>)  
38 and the health economics review protocol, which can be found in each of the evidence  
39 reports.

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

#### 2.4.1.23 NICE health economic evidence profiles

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



1 report. The health economic evidence profile shows an assessment of applicability and  
2 methodological quality for each economic study, with footnotes indicating the reasons for the  
3 assessment. These assessments were made by the health economist using the economic  
4 evaluation checklist from the NICE guidelines manual.<sup>2</sup> It also shows the incremental costs,  
5 incremental effects (for example, quality-adjusted life years [QALYs]) and incremental cost-  
6 effectiveness ratio (ICER) for the base case analysis in the study, as well as information  
7 about the assessment of uncertainty in the analysis. See Table 7 for more details.

8 When a non-UK study was included in the profile, the results were converted into pounds  
9 sterling using the appropriate purchasing power parity.<sup>4</sup>

10 **Table 7: Content of NICE health economic evidence profile**

Item	Description
Study	Surname of first author, date of study publication and country perspective with a reference to full information on the study.
Applicability	An assessment of applicability of the study to this guideline, the current NHS situation and NICE decision-making: <sup>(a)</sup> <ul style="list-style-type: none"> <li>• 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.</li> <li>• Partially applicable – the study fails to meet 1 or more applicability criteria, and this could change the conclusions about cost effectiveness.</li> <li>• 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.</li> </ul>
Limitations	An assessment of methodological quality of the study: <sup>(a)</sup> <ul style="list-style-type: none"> <li>• 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.</li> <li>• Potentially serious limitations – the study fails to meet 1 or more quality criteria, and this could change the conclusions about cost effectiveness.</li> <li>• 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.</li> </ul>
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.

11 (a) *Applicability and limitations were assessed using the economic evaluation checklist in appendix H of the NICE*  
12 *guidelines manual*<sup>2</sup>

### 2.4.23 Undertaking new health economic analysis

14 No new health economic analyses were undertaken for this guideline update. The committee  
15 considered that sufficient existing evidence was identified for the high priority review question  
16 and new analysis was not required.

### 2.4.31 Cost-effectiveness criteria

2 NICE's report 'Social value judgements: principles for the development of NICE guidance'  
3 sets out the principles that committees should consider when judging whether an intervention  
4 offers good value for money.<sup>3</sup> In general, an intervention was considered to be cost effective  
5 (given that the estimate was considered plausible) if either of the following criteria applied:  
6 • the intervention dominated other relevant strategies (that is, it was both less costly in  
7 terms of resource use and more clinically effective compared with all the other relevant  
8 alternative strategies), or  
9 • the intervention cost less than £20,000 per QALY gained compared with the next best  
10 strategy.

11 If the committee recommended an intervention that was estimated to cost more than £20,000  
12 per QALY gained, or did not recommend one that was estimated to cost less than £20,000  
13 per QALY gained, the reasons for this decision are discussed explicitly in 'The committee's  
14 discussion of the evidence' section of the relevant evidence report, with reference to issues  
15 regarding the plausibility of the estimate or to the factors set out in 'Social value judgements:  
16 principles for the development of NICE guidance'.<sup>3</sup>

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

### 2.4.40 In the absence of health economic evidence

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

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

## 2.5.9 Developing recommendations

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

- 31 • Summaries of clinical and health economic evidence and quality (as presented in  
32 evidence reports [A–H]).
- 33 • Evidence tables of the clinical and health economic evidence reviewed from the literature.  
34 All evidence tables can be found in appendices to the relevant evidence reports.
- 35 • Forest plots and ordinal shift distribution graphs (in appendices to the relevant evidence  
36 reports).

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

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

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

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

- 23 • The actions health professionals need to take.
- 24 • The information readers need to know.
- 25 • The strength of the recommendation (for example the word 'offer' was used for strong  
26 recommendations and 'consider' for weaker recommendations).
- 27 • The involvement of patients (and their carers if needed) in decisions on treatment and  
28 care.
- 29 • Consistency with NICE's standard advice on recommendations about drugs, waiting times  
30 and ineffective interventions (see section 9.2 in the NICE guidelines manual<sup>2</sup>).

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

### **2.5.13 Research recommendations**

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

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

### **2.5.21 Validation process**

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

### **2.5.31 Updating the guideline**

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

### **2.5.45 Disclaimer**

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

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

### **2.5.53 Funding**

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

16

## 3<sub>1</sub> Additional information

### 2 Integration with previous evidence

3 This guideline includes a number of entirely new reviews and others that cover areas that  
4 were assessed in the previous guideline (CG68). When the latter is true, the studies that  
5 were included and excluded from the previous guideline were checked against the protocols  
6 for this update and new searches were run and sifted from the date of the previous reviews.

7

## 4<sub>1</sub> Acronyms and abbreviations

2

Acronym or abbreviation	Description
ABCD2	Age, blood pressure, clinical features, duration of TIA, and presence of diabetes
ABCD2-I	Age, blood pressure, clinical features, duration of TIA, and presence of diabetes plus imaging evidence of brain infarction
ABCD3	Age, blood pressure, clinical features, duration of TIA, presence of diabetes and dual TIA (the presence of $\geq 2$ TIA symptoms within 7 days).
ABCD3-I	Age, blood pressure, clinical features, duration of TIA, presence of diabetes and dual TIA (the presence of $\geq 2$ TIA symptoms within 7 days) plus the presence of abnormal findings on neuroimaging.
AQoL	Assessment of quality of life scale
ASPECTS	Alberta Stroke Program Early CT Score
CEA	Cost-effectiveness analysis
CI	Confidence interval
CTA	CT angiography
CUA	Cost-utility analysis
DSA	Digital subtraction angiography
DVT	Deep vein thrombosis
DW	Diffusion-weighted
DWI	Diffusion weighted imaging
ED	Emergency department
EVT	Endovascular therapy
GCS	Glasgow coma scale
GP	General practitioner
GRADE	Grading of Recommendations Assessment, Development and Evaluation
GTN	Glyceryl trinitrate
ICA	Internal carotid artery
ICER	Incremental cost-effectiveness ratio
IPD	Individual patient data
IQR	Interquartile range
MCA	Middle cerebral artery
MRA	Magnetic resonance angiography
mRS	Modified Rankin scale
NGC	National Guideline Centre
NICE	The National Institute for Health and Care Excellence
NIHSS	National Institutes of Health stroke scale
OECD	Organisation for economic co-operation and development
OR	Odds ratio
PE	Pulmonary embolism
PICO	Population, intervention, comparison and outcomes
PROBAST	Prediction study risk of bias assessment tool
PROBE	Prospective, randomized, open-label, controlled trial with blinded

<b>Acronym or abbreviation</b>	<b>Description</b>
	outcome evaluation
QALY	Quality-adjusted life years
RCT	Randomised controlled trial
RR	Relative risk
SD	Standard deviation
TFNE	Transient focal neurological episodes
TIA	Transient ischaemic attack
tPA	Tissue plasminogen activator

## 5<sub>1</sub> Glossary

2 The NICE Glossary can be found at [www.nice.org.uk/glossary](http://www.nice.org.uk/glossary).

### 5.1<sub>3</sub> Guideline-specific terms

Term	Definition
ABCD tools	<p>Prognostic score to identify people at high risk of stroke after a TIA. It is calculated based on:</p> <p>A – age (≥60 years, 1 point)</p> <p>B – blood pressure at presentation (≥140/90 mmHg, 1 point)</p> <p>C – clinical features (unilateral weakness, 2 points or speech disturbance without weakness, 1 point)</p> <p>D – duration of symptoms (≥60 minutes, 2 points or 10–59 minutes, 1 point).</p> <p>The calculation of ABCD2 also includes the presence of diabetes (1 point). Total scores range from 0 (low risk) to 7 (high risk).</p> <p>The calculation of ABCD2-I also includes acute diffusion-weighted imaging hyperintensity (3 points). Total scores range from 0 (low risk) to 10 (high risk).</p> <p>The calculation of ABCD3 also includes dual TIA (the presence of ≥2 TIA symptoms within 7 days, 2 points). Total scores range from 0 (low risk) to 9 (high risk).</p> <p>The calculation of ABCD3-I also includes imaging (acute diffusion-weighted imaging hyperintensity (2 points) and ipsilateral stenosis of the internal carotid artery by duplex ultrasound, or angiography (2 points)). Total scores range from 0 (low risk) to 13 (high risk).</p>
Alteplase	A drug used for thrombolysis
Alberta Stroke Program Early CT Score	<p>A 10-point quantitative topographic CT scan score used in patients with middle cerebral artery (MCA) stroke.</p> <p>Segmental assessment of the MCA vascular territory is made and 1 point is deducted from the initial score of 10 for every region involved:</p> <ul style="list-style-type: none"> <li>• caudate</li> <li>• putamen</li> <li>• internal capsule</li> <li>• insular cortex</li> <li>• M1: "anterior MCA cortex," corresponding to frontal operculum</li> <li>• M2: "MCA cortex lateral to insular ribbon" corresponding to anterior temporal lobe</li> <li>• M3: "posterior MCA cortex" corresponding to posterior temporal lobe</li> <li>• M4: "anterior MCA territory immediately superior to M1"</li> <li>• M5: "lateral MCA territory immediately superior to M2"</li> <li>• M6: "posterior MCA territory immediately superior to M3"</li> </ul>
Assessment of quality of life scale	Quality of life instrument designed to measure health-related quality of life and to be the descriptive system for a multi-attribute utility instrument. It measures 5 dimensions, each with 3 items: illness, independent living, social relationships, physical senses and psychological wellbeing.
CT angiography	Use of a CT scanner to produce detailed images of blood vessels and tissues.
Decompressive hemicraniectomy	A surgical procedure for the treatment of raised intracerebral pressure. A piece of the skull is removed to allow the swelling brain to expand.
Diffusion-weighted imaging	Specific MRI sequences and software used to generate images harnessing the diffusion of water molecules to generate contrast in MR



Term	Definition
	images.
Digital subtraction angiography	A fluoroscopy technique used in interventional radiology to clearly visualise blood vessels in a bony or dense soft tissue environment.
Endovascular therapy	Minimally invasive inter-arterial method that uses catheter-guided devices to assist restoration of blood flow in an occluded vessel through mechanically removing the clot from the site of occlusion.
Functional status	An individual's ability to perform normal daily activities required to meet basic needs, fulfil usual roles, and maintain health and well-being.
Glasgow coma scale	A scoring system used to describe the level of consciousness in a person following a traumatic brain injury. It measures eye opening, verbal response, and motor response. Scores range from 0 to 15, with low scores indicating a more severe deficit.
Infarct/infarction	An area of cell death due to the result of a deprived blood supply.
Intracerebral haemorrhage	A bleed in the brain as a result of a ruptured or bleeding blood vessel within the brain tissue or ventricles.
Intracerebral haemorrhage	A bleed in the brain as a result of a ruptured or bleeding blood vessel.
Ischaemia	A restriction of the blood supply that starves tissues of oxygen and nutrients.
Magnetic resonance angiography	Use of magnetic resonance imaging to evaluate blood vessels.
Mass lesion	A space-occupying growth.
mRS	<p>This is a functional outcome scale where a lower score indicates a better outcome.</p> <p>0: No symptoms at all</p> <p>1: No significant disability despite symptoms; able to carry out all usual duties and activities</p> <p>2: Slight disability; unable to carry out all previous activities, but able to look after own affairs without assistance</p> <p>3: Moderate disability; requiring some help, but able to walk without assistance</p> <p>4: Moderately severe disability; unable to walk without assistance and unable to attend to own bodily needs without assistance</p> <p>5: Severe disability; bedridden, incontinent and requiring constant nursing care and attention</p> <p>6: Dead</p>
National Institutes of Health stroke scale	<p>A tool to assess stroke severity consisting of 11 elements each scored between 0 and 2, 3 or 4. Higher scores indicate greater impairment and the scale range is 0-42. The items are:</p> <ul style="list-style-type: none"> <li>• Level of Consciousness</li> <li>• Horizontal Eye Movement</li> <li>• Visual field test</li> <li>• Facial Palsy</li> <li>• Motor Arm</li> <li>• Motor Leg</li> <li>• Limb Ataxia</li> <li>• Sensory</li> <li>• Language</li> <li>• Speech</li> <li>• Extinction and Inattention</li> </ul>

Term	Definition
Subarachnoid haemorrhage	A type of stroke caused by bleeding on the surface of the brain.
Stroke	The damaging or killing of brain cells starved of oxygen as a result of the blood supply to part of the brain being cut off. Types of stroke include ischaemic stroke caused by blood clots to the brain or haemorrhagic stroke caused by bleeding into/of the brain.
Stroke mimics	A term used to describe other clinical conditions which can mimic a stroke and confound diagnosis. Examples of these include brain tumours, epilepsy or subdural haematosis. Neurologic abnormalities similar to a stroke can also be the result of imbalances of glucose, sodium and calcium.
Thrombectomy	The interventional procedure of removing a blood clot (thrombus) from a blood vessel.
Thrombolysis	The use of drugs to break up a blood clot. Two examples of thrombolysis drugs are tPA and alteplase.
Tissue plasminogen activator	A drug used for thrombolysis.
Transient focal neurological episodes	Transient focal symptoms which often mimic transient ischaemic attack, but are more often related to bleeding (in particular superficial cortical siderosis or focal convexity sub-arachnoid haemorrhage) rather than ischaemia.
Transient ischaemic attack	A stroke which recovers within 24 hours of onset of symptoms.
Wake up stroke	People who go to sleep normal and awaken with stroke symptoms.

## 5.2.1 General terms

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

Term	Definition
	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	<p>A way to prevent researchers, doctors and patients in a clinical trial from knowing which study group each patient is in so they cannot influence the results. The best way to do this is by sorting patients into study groups randomly. The purpose of 'blinding' or 'masking' is to protect against bias.</p> <p>A single-blinded study is one in which patients do not know which study group they are in (for example whether they are taking the experimental drug or a placebo). A double-blinded study is one in which neither patients nor the researchers and doctors know which study group the patients are in. A triple blind study is one in which neither the patients, clinicians or the people carrying out the statistical analysis know which treatment patients received.</p>
Carer (caregiver)	Someone who looks after family, partners or friends in need of help because they are ill, frail or have a disability.
Case–control study	<p>A study to find out the cause(s) of a disease or condition. This is done by comparing a group of patients who have the disease or condition (cases) with a group of people who do not have it (controls) but who are otherwise as similar as possible (in characteristics thought to be unrelated to the causes of the disease or condition). This means the researcher can look for aspects of their lives that differ to see if they may cause the condition.</p> <p>For example, a group of people with lung cancer might be compared with a group of people the same age that do not have lung cancer. The researcher could compare how long both groups had been exposed to tobacco smoke. Such studies are retrospective because they look back in time from the outcome to the possible causes of a disease or condition.</p>
Case series	Report of a number of cases of a given disease, usually covering the course of the disease and the response to treatment. There is no comparison (control) group of patients.
Clinical efficacy	The extent to which an intervention is active when studied under controlled research conditions.
Clinical effectiveness	How well a specific test or treatment works when used in the 'real world' (for example, when used by a doctor with a patient at home), rather than in a carefully controlled clinical trial. Trials that assess clinical effectiveness are sometimes called management trials. Clinical effectiveness is not the same as efficacy.
Clinician	A healthcare professional who provides patient care. For example, a doctor, nurse or physiotherapist.
Cochrane Review	The Cochrane Library consists of a regularly updated collection of evidence-based medicine databases including the Cochrane Database of Systematic Reviews (reviews of randomised controlled trials prepared by the Cochrane Collaboration).
Cohort study	A study with 2 or more groups of people – cohorts – with similar characteristics. One group receives a treatment, is exposed to a risk

Term	Definition
	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.
Common odds ratio	Cumulative odds ratio as calculated by logistic regression (shift analysis). This provides a treatment effect in the form of a common estimate of the odds ratio for improvement over considered cut-points. This analysis relies on the assumption of an odds ratio behind any cut-off point.
Comorbidity	A disease or condition that someone has in addition to the health problem being studied or treated.
Comparability	Similarity of the groups in characteristics likely to affect the study results (such as health status or age).
Concordance	This is a recent term whose meaning has changed. It was initially applied to the consultation process in which doctor and patient agree therapeutic decisions that incorporate their respective views, but now includes patient support in medicine taking as well as prescribing communication. Concordance reflects social values but does not address medicine-taking and may not lead to improved adherence.
Confidence interval (CI)	<p>There is always some uncertainty in research. This is because a small group of patients is studied to predict the effects of a treatment on the wider population. The confidence interval is a way of expressing how certain we are about the findings from a study, using statistics. It gives a range of results that is likely to include the 'true' value for the population.</p> <p>The CI is usually stated as '95% CI', which means that the range of values has a 95 in a 100 chance of including the 'true' value. For example, a study may state that "based on our sample findings, we are 95% certain that the 'true' population blood pressure is not higher than 150 and not lower than 110". In such a case the 95% CI would be 110 to 150.</p> <p>A wide confidence interval indicates a lack of certainty about the true effect of the test or treatment – often because a small group of patients has been studied. A narrow confidence interval indicates a more precise estimate (for example, if a large number of patients have been studied).</p>
Confounding factor	<p>Something that influences a study and can result in misleading findings if it is not understood or appropriately dealt with.</p> <p>For example, a study of heart disease may look at a group of people that exercises regularly and a group that does not exercise. If the ages of the people in the 2 groups are different, then any difference in heart disease rates between the 2 groups could be because of age rather than exercise. Therefore age is a confounding factor.</p>
Consensus methods	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	<p>A group of people in a study who do not receive the treatment or test being studied. Instead, they may receive the standard treatment (sometimes called 'usual care') or a dummy treatment (placebo). The results for the control group are compared with those for a group receiving the treatment being tested. The aim is to check for any differences.</p> <p>Ideally, the people in the control group should be as similar as possible to those in the treatment group, to make it as easy as possible to detect any effects due to the treatment.</p>
Cost–benefit analysis	Cost–benefit analysis is one of the tools used to carry out an

Term	Definition
(CBA)	economic evaluation. The costs and benefits are measured using the same monetary units (for example, pounds sterling) to see whether the benefits exceed the costs.
Cost–consequences analysis (CCA)	Cost–consequences analysis is one of the tools used to carry out an economic evaluation. This compares the costs (such as treatment and hospital care) and the consequences (such as health outcomes) of a test or treatment with a suitable alternative. Unlike cost–benefit analysis or cost-effectiveness analysis, it does not attempt to summarise outcomes in a single measure (like the quality-adjusted life year) or in financial terms. Instead, outcomes are shown in their natural units (some of which may be monetary) and it is left to decision-makers to determine whether, overall, the treatment is worth carrying out.
Cost-effectiveness analysis (CEA)	Cost-effectiveness analysis is one of the tools used to carry out an economic evaluation. The benefits are expressed in non-monetary terms related to health, such as symptom-free days, heart attacks avoided, deaths avoided or life years gained (that is, the number of years by which life is extended as a result of the intervention).
Cost-effectiveness model	An explicit mathematical framework, which is used to represent clinical decision problems and incorporate evidence from a variety of sources in order to estimate the costs and health outcomes.
Cost–utility analysis (CUA)	Cost–utility analysis is one of the tools used to carry out an economic evaluation. The benefits are assessed in terms of both quality and duration of life, and expressed as quality-adjusted life years (QALYs). See also utility.
Credible interval (CrI)	The Bayesian equivalent of a confidence interval.
Decision analysis	An explicit quantitative approach to decision-making under uncertainty, based on evidence from research. This evidence is translated into probabilities, and then into diagrams or decision trees which direct the clinician through a succession of possible scenarios, actions and outcomes.
Deterministic analysis	In economic evaluation, this is an analysis that uses a point estimate for each input. In contrast, see Probabilistic analysis
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

Term	Definition
	estimate the benefits of a particular drug, programme or intervention.
Effect (as in effect measure, treatment effect, estimate of effect, effect size)	A measure that shows the magnitude of the outcome in one group compared with that in a control group. For example, if the absolute risk reduction is shown to be 5% and it is the outcome of interest, the effect size is 5%. The effect size is usually tested, using statistics, to find out how likely it is that the effect is a result of the treatment and has not just happened by chance (that is, to see if it is statistically significant).
Effectiveness	How beneficial a test or treatment is under usual or everyday conditions, compared with doing nothing or opting for another type of care.
Efficacy	How beneficial a test, treatment or public health intervention is under ideal conditions (for example, in a laboratory), compared with doing nothing or opting for another type of care.
Epidemiological study	The study of a disease within a population, defining its incidence and prevalence and examining the roles of external influences (for example, infection, diet) and interventions.
EQ-5D (EuroQoL 5 dimensions)	A standardised instrument used to measure health-related quality of life. It provides a single index value for health status.
Evidence	Information on which a decision or guidance is based. Evidence is obtained from a range of sources including randomised controlled trials, observational studies, expert opinion (of clinical professionals or patients).
Exclusion criteria (literature review)	Explicit standards used to decide which studies should be excluded from consideration as potential sources of evidence.
Exclusion criteria (clinical study)	Criteria that define who is not eligible to participate in a clinical study.
Extended dominance	If Option A is both more clinically effective than Option B and has a lower cost per unit of effect, when both are compared with a do-nothing alternative then Option A is said to have extended dominance over Option B. Option A is therefore cost effective and should be preferred, other things remaining equal.
Extrapolation	An assumption that the results of studies of a specific population will also hold true for another population with similar characteristics.
Follow-up	Observation over a period of time of an individual, group or initially defined population whose appropriate characteristics have been assessed in order to observe changes in health status or health-related variables.
Generalisability	The extent to which the results of a study hold true for groups that did not participate in the research. See also external validity.
Gold standard	A method, procedure or measurement that is widely accepted as being the best available to test for or treat a disease.
GRADE, GRADE profile	A system developed by the GRADE Working Group to address the shortcomings of present grading systems in healthcare. The GRADE system uses a common, sensible and transparent approach to grading the quality of evidence. The results of applying the GRADE system to clinical trial data are displayed in a table known as a GRADE profile.
Harms	Adverse effects of an intervention.
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	The term is used in meta-analyses and systematic reviews to

Term	Definition
or Lack of homogeneity	describe when the results of a test or treatment (or estimates of its effect) differ significantly in different studies. Such differences may occur as a result of differences in the populations studied, the outcome measures used or because of different definitions of the variables involved. It is the opposite of homogeneity.
Imprecision	Results are imprecise when studies include relatively few patients and few events and thus have wide confidence intervals around the estimate of effect.
Inclusion criteria (literature review)	Explicit criteria used to decide which studies should be considered as potential sources of evidence.
Incremental analysis	The analysis of additional costs and additional clinical outcomes with different interventions.
Incremental cost	The extra cost linked to using one test or treatment rather than another. Or the additional cost of doing a test or providing a treatment more frequently.
Incremental cost-effectiveness ratio (ICER)	The difference in the mean costs in the population of interest divided by the differences in the mean outcomes in the population of interest for one treatment compared with another.
Incremental net benefit (INB)	The value (usually in monetary terms) of an intervention net of its cost compared with a comparator intervention. The INB can be calculated for a given cost-effectiveness (willingness to pay) threshold. If the threshold is £20,000 per QALY gained then the INB is calculated as: $(£20,000 \times \text{QALYs gained}) - \text{Incremental cost}$ .
Indirectness	The available evidence is different to the review question being addressed, in terms of PICO (population, intervention, comparison and outcome).
Intention-to-treat analysis (ITT)	An assessment of the people taking part in a clinical trial, based on the group they were initially (and randomly) allocated to. This is regardless of whether or not they dropped out, fully complied with the treatment or switched to an alternative treatment. Intention-to-treat analyses are often used to assess clinical effectiveness because they mirror actual practice: that is, not everyone complies with treatment and the treatment people receive may be changed according to how they respond to it.
Intervention	In medical terms this could be a drug treatment, surgical procedure, diagnostic or psychological therapy. Examples of public health interventions could include action to help someone to be physically active or to eat a more healthy diet.
Intraoperative	The period of time during a surgical procedure.
Length of stay	The total number of days a participant stays in hospital.
Licence	See 'Product licence'.
Life years gained	Mean average years of life gained per person as a result of the intervention compared with an alternative intervention.
Long-term care	Residential care in a home that may include skilled nursing care and help with everyday activities. This includes nursing homes and residential homes.
Logistic regression or Logit model	In statistics, logistic regression is a type of analysis used for predicting the outcome of a binary dependent variable based on one or more predictor variables. It can be used to estimate the log of the odds (known as the 'logit').
Loss to follow-up	A patient, or the proportion of patients, actively participating in a clinical trial at the beginning, but whom the researchers were unable to trace or contact by the point of follow-up in the trial
Markov model	A method for estimating long-term costs and effects for recurrent or chronic conditions, based on health states and the probability of

Term	Definition
	transition between them within a given time period (cycle).
Meta-analysis	A method often used in systematic reviews. Results from several studies of the same test or treatment are combined to estimate the overall effect of the treatment.
Multivariate model	A statistical model for analysis of the relationship between 2 or more predictor (independent) variables and the outcome (dependent) variable.
Net monetary benefit (NMB)	The value in monetary terms of an intervention net of its cost. The NMB can be calculated for a given cost-effectiveness threshold. If the threshold is £20,000 per QALY gained then the NMB for an intervention is calculated as: (£20,000 × mean QALYs) – mean cost. The most preferable option (that is, the most clinically effective option to have an ICER below the threshold selected) will be the treatment with the highest NMB.
Non-randomised intervention study	A quantitative study investigating the effectiveness of an intervention that does not use randomisation to allocate patients (or units) to treatment groups. Non-randomised studies include observational studies, where allocation to groups occurs through usual treatment decisions or people's preferences. Non-randomised studies can also be experimental, where the investigator has some degree of control over the allocation of treatments.  Non-randomised intervention studies can use a number of different study designs, and include cohort studies, case-control studies, controlled before-and-after studies, interrupted-time-series studies and quasi-randomised controlled trials.
Number needed to treat (NNT)	The average number of patients who need to be treated to get a positive outcome. For example, if the NNT is 4, then 4 patients would have to be treated to ensure 1 of them gets better. The closer the NNT is to 1, the better the treatment.  For example, if you give a stroke prevention drug to 20 people before 1 stroke is prevented, the number needed to treat is 20. See also number needed to harm, absolute risk reduction.
Observational study	Individuals or groups are observed or certain factors are measured. No attempt is made to affect the outcome. For example, an observational study of a disease or treatment would allow 'nature' or usual medical care to take its course. Changes or differences in one characteristic (for example, whether or not people received a specific treatment or intervention) are studied without intervening.  There is a greater risk of selection bias than in experimental studies.
Odds ratio	Odds are a way to represent how likely it is that something will happen (the probability). An odds ratio compares the probability of something in one group with the probability of the same thing in another.  An odds ratio of 1 between 2 groups would show that the probability of the event (for example a person developing a disease, or a treatment working) is the same for both. An odds ratio greater than 1 means the event is more likely in the first group. An odds ratio less than 1 means that the event is less likely in the first group.  Sometimes probability can be compared across more than 2 groups – in this case, one of the groups is chosen as the 'reference category', and the odds ratio is calculated for each group compared with the reference category. For example, to compare the risk of dying from lung cancer for non-smokers, occasional smokers and regular smokers, non-smokers could be used as the reference category. Odds ratios would be worked out for occasional smokers compared with non-smokers and for regular smokers compared with non-smokers. See also confidence interval, risk ratio.



Term	Definition
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.
Ordinal shift analysis	Analysis that takes into account alterations in the distribution of patients over the entire range of possible outcomes on an ordinal outcome scale. This retains all information captured by an ordinal outcome scale and can improve study power.
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	<p>The p value is a statistical measure that indicates whether or not an effect is statistically significant.</p> <p>For example, if a study comparing 2 treatments found that one seems more effective than the other, the p value is the probability of obtaining these results by chance. By convention, if the p value is below 0.05 (that is, there is less than a 5% probability that the results occurred by chance) it is considered that there probably is a real difference between treatments. If the p value is 0.001 or less (less than a 1% probability that the results occurred by chance), the result is seen as highly significant.</p> <p>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.</p>
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.
Posterior distribution	In Bayesian statistics this is the probability distribution for a statistic based after combining established information or belief (the prior) with new evidence (the likelihood).
Postoperative	Pertaining to the period after patients leave the operating theatre, following surgery.
Power (statistical)	The ability to demonstrate an association when one exists. Power is related to sample size; the larger the sample size, the greater the power and the lower the risk that a possible association could be missed.
Preoperative	The period before surgery commences.
Pre-test probability	In diagnostic tests: The proportion of people with the target disorder in the population at risk at a specific time point or time interval. Prevalence may depend on how a disorder is diagnosed.
Prevalence	See Pre-test probability.
Prior distribution	In Bayesian statistics this is the probability distribution for a statistic based on previous evidence or belief.

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

Term	Definition
Resource implication	The likely impact in terms of finance, workforce or other NHS resources.
Retrospective study	A research study that focuses on the past and present. The study examines past exposure to suspected risk factors for the disease or condition. Unlike prospective studies, it does not cover events that occur after the study group is selected.
Review question	In guideline development, this term refers to the questions about treatment and care that are formulated to guide the development of evidence-based recommendations.
Risk ratio (RR)	<p>The ratio of the risk of disease or death among those exposed to certain conditions compared with the risk for those who are not exposed to the same conditions (for example, the risk of people who smoke getting lung cancer compared with the risk for people who do not smoke).</p> <p>If both groups face the same level of risk, the risk ratio is 1. If the first group had a risk ratio of 2, subjects in that group would be twice as likely to have the event happen. A risk ratio of less than 1 means the outcome is less likely in the first group. The risk ratio is sometimes referred to as relative risk.</p>
Secondary outcome	An outcome used to evaluate additional effects of the intervention deemed a priori as being less important than the primary outcomes.
Selection bias	<p>Selection bias occurs if:</p> <ul style="list-style-type: none"> <li>a) The characteristics of the people selected for a study differ from the wider population from which they have been drawn, or</li> <li>b) There are differences between groups of participants in a study in terms of how likely they are to get better.</li> </ul>
Sensitivity	<p>How well a test detects the thing it is testing for.</p> <p>If a diagnostic test for a disease has high sensitivity, it is likely to pick up all cases of the disease in people who have it (that is, give a 'true positive' result). But if a test is too sensitive it will sometimes also give a positive result in people who don't have the disease (that is, give a 'false positive').</p> <p>For example, if a test were developed to detect if a woman is 6 months pregnant, a very sensitive test would detect everyone who was 6 months pregnant, but would probably also include those who are 5 and 7 months pregnant.</p> <p>If the same test were more specific (sometimes referred to as having higher specificity), it would detect only those who are 6 months pregnant, and someone who was 5 months pregnant would get a negative result (a 'true negative'). But it would probably also miss some people who were 6 months pregnant (that is, give a 'false negative').</p> <p>Breast screening is a 'real-life' example. The number of women who are recalled for a second breast screening test is relatively high because the test is very sensitive. If it were made more specific, people who don't have the disease would be less likely to be called back for a second test but more women who have the disease would be missed.</p>
Sensitivity analysis	<p>A means of representing uncertainty in the results of economic evaluations. Uncertainty may arise from missing data, imprecise estimates or methodological controversy. Sensitivity analysis also allows for exploring the generalisability of results to other settings. The analysis is repeated using different assumptions to examine the effect on the results.</p> <p>One-way simple sensitivity analysis (univariate analysis): each parameter is varied individually in order to isolate the consequences</p>

Term	Definition
	<p>of each parameter on the results of the study.</p> <p>Multi-way simple sensitivity analysis (scenario analysis): 2 or more parameters are varied at the same time and the overall effect on the results is evaluated.</p> <p>Threshold sensitivity analysis: the critical value of parameters above or below which the conclusions of the study will change are identified.</p> <p>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).</p>
Significance (statistical)	<p>A result is deemed statistically significant if the probability of the result occurring by chance is less than 1 in 20 (<math>p &lt; 0.05</math>).</p>
Specificity	<p>The proportion of true negatives that are correctly identified as such. For example in diagnostic testing the specificity is the proportion of non-cases correctly diagnosed as non-cases.</p> <p>See related term 'Sensitivity'.</p> <p>In terms of literature searching a highly specific search is generally narrow and aimed at picking up the key papers in a field and avoiding a wide range of papers.</p>
Stakeholder	<p>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:</p> <ul style="list-style-type: none"> <li>• manufacturers of drugs or equipment</li> <li>• national patient and carer organisations</li> <li>• NHS organisations</li> <li>• organisations representing healthcare professionals.</li> </ul>
State transition model	<p>See Markov model</p>
Systematic review	<p>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.</p>
Time horizon	<p>The time span over which costs and health outcomes are considered in a decision analysis or economic evaluation.</p>
Transition probability	<p>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.</p>
Treatment allocation	<p>Assigning a participant to a particular arm of a trial.</p>
Univariate	<p>Analysis which separately explores each variable in a data set.</p>
Utility	<p>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).</p>

1  
2

## 1 **References**

2

3

4 1. GRADE Working Group. The Grading of Recommendations Assessment,  
5 Development and Evaluation (GRADE) Working Group website. 2011. Available from:  
6 <http://www.gradeworkinggroup.org/> Last accessed: 17/9/18.

7 2. National Institute for Health and Care Excellence. Developing NICE guidelines: the  
8 manual. London. National Institute for Health and Care Excellence, 2014. Available  
9 from:  
10 <http://www.nice.org.uk/article/PMG20/chapter/1%20Introduction%20and%20overview>

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

17 4. Organisation for Economic Co-operation and Development (OECD). Purchasing  
18 power parities (PPP). 2012. Available from: <http://www.oecd.org/sdd/prices-ppp/> Last  
19 accessed: 2/2/2018.

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

23