

**STUDIES TO INFORM THE METHODS FOR
COCHRANE SYSTEMATIC REVIEWS OF
DIAGNOSTIC ACCURACY IN STROKE
MEDICINE**

Miriam Brazzelli

Doctor of Philosophy by Research
University of Edinburgh
2010

To the memory of Gianluigi Brazzelli

Declaration

I hereby declare that:

- I entirely composed this thesis

- All the work contributing to this thesis was undertaken while I was in post at the Division of Clinical Neurosciences, University of Edinburgh, Western General Hospital

- The thesis contains examples of original research and has not been submitted in candidature for any other degree, diploma, or professional qualification

- My work for the preparation of this is outlined in the section entitled “My contribution to this thesis”. The role of my supervisors and co-authors is clarified as necessary in each chapter, and in the Acknowledgements.

Contents

ABSTRACT	7
1. INTRODUCTION	10
1.1 Evaluating diagnostic tests: methodological aspects	10
1.1.1 Electronic search strategies	11
1.1.2 Quality of reports of diagnostic accuracy studies	13
1.1.3 Statistical methods for synthesising results	14
1.1.4 Helping clinicians interpret summary findings	15
1.2 Stroke diagnosis	16
1.3 Imaging diagnosis in stroke medicine	17
1.3 The Cochrane Collaboration	19
1.4 Outline of the Thesis	20
2. THE QUALITY OF REPORTING OF DIAGNOSTIC ACCURACY STUDIES ON THE USE OF MAGNETIC RESONANCE IMAGING FOR THE EARLY DETECTION OF VASCULAR LESIONS IN STROKE PATIENTS	29
2.1 Background	29
2.2 Material and methods	31
2.2.1 Study selection	31
2.2.2 Data extraction	32
2.2.3 Statistical analysis	33
2.3 Results	33
2.4 Discussion	39
2.4.1 Implications for researchers	45
2.4.2 Implications for Cochrane diagnostic test accuracy (DTA) reviews and DTA methods groups	45
2.4.3 Implications for journal editors	46
3. IS THERE BIAS IN THE PROCESS OF PUBLICATION OF DIAGNOSTIC ACCURACY STUDIES IN STROKE SUBMITTED AS ABSTRACTS?	50
3.1 Introduction	50
3.2 Methods	52
3.3 Statistical analyses	54
3.4 Results	55
3.4.1 Abstracts published in full	57
3.4.2 Study characteristics and publication	59
3.5 Discussion	61
3.5.1 Implications for future research	66

4.	MAGNETIC RESONANCE IMAGING VERSUS COMPUTED TOMOGRAPHY FOR DETECTION OF ACUTE VASCULAR LESIONS IN PATIENTS PRESENTING WITH STROKE SYMPTOMS	70
4.1	Background	70
4.1.1	Target condition being diagnosed	70
4.1.2	Index test(s)	72
4.2	Objectives	74
4.2.1	Primary Objectives	74
4.2.2	Secondary Objectives	75
4.2.3	Investigation of sources of heterogeneity	75
4.3	Methods	76
4.3.1	Criteria for considering studies for this review	76
4.3.1.1	Types of studies	76
4.3.1.2	Participants	77
4.3.1.3	Index tests	78
4.3.1.4	Target conditions	78
4.3.1.5	Reference standard	78
4.3.2	Search methods for identification of studies	79
4.3.2.1	Electronic searches	79
4.3.2.2	Medline and Embase searches	80
4.3.2.3	Searching other resources	80
4.3.3	Data collection and analysis	82
4.3.3.1	Selection of studies	82
4.3.3.2	Data extraction and management	82
4.3.3.3	Assessment of methodological quality	83
4.3.3.4	Statistical analysis and data synthesis	84
4.4	Results	86
4.4.1	Results of the search	86
4.4.2	Methodological quality of included studies	91
4.4.2.1	Studies on ischaemic stroke	91
4.4.2.2	Studies on haemorrhagic stroke	94
4.4.3	Findings	97
4.4.3.1	Studies on ischaemic stroke	97
4.4.3.2	Studies on haemorrhagic stroke	97
4.4	Discussion	100
4.4.1	Findings on ischaemic stroke	101
4.4.2	Findings on haemorrhagic stroke	103
4.4.3	Summary of main results	104
4.4.4	Strengths and weaknesses of the review	105
4.4.5	Applicability of findings to clinical practice and policy	108
4.5	Authors' conclusions	109
4.5.1	Implications for practice	109
4.5.2	Implications for research	109
4.5.3	Implications for the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy	110
4.5.4	Summary of this review's findings	110

5.	TO ASSESS THE USEFULNESS OF TWO FORMS OF SUMMARIES OF A SYSTEMATIC REVIEW OF DIAGNOSTIC ACCURACY	121
5.1	Introduction	121
5.1.1	Objectives	125
5.2	Methods	125
5.2.1	Study Design	125
5.2.2	Participants	126
5.2.3	Development of Abstract and Summary of Findings table	126
5.2.4	Development of a questionnaire to assess clinicians' demographics, understanding of results, and preferences	128
5.2.5	Planned analyses	133
5.3	Results	134
5.3.1	Response rate	134
5.3.2	Demographics of participants	134
5.3.3	Summary documents: Abstract versus SoF table	135
5.3.4	Clinicians' preferences: Abstract versus SoF table	140
5.4	Discussion	142
5.4.1	Limitations of the study	146
5.4.2	Implications for practice and for the Cochrane Collaboration	148
6.	GENERAL DISCUSSION	153
6.1	Methodological relevance of this thesis	153
6.1.1	Evidence of reporting bias	154
6.1.2	Evidence of publication bias	156
6.1.3	Evidence of methodological biases in systematic reviews of test accuracy	158
6.1.4	Complex issues in systematic reviews of diagnostic test accuracy	160
6.1.5	The challenge of summarising findings of systematic reviews of test accuracy for clinicians	163
6.2	Implications of this thesis for research	164
6.3	Implications of this thesis for authors and consumers of diagnostic test accuracy reviews	165
6.4	Implications of this thesis for journal editors, publishers, professional organizations, peer reviewers	166
6.5	Implications of this thesis for the Cochrane Collaboration	167
6.6	Implications of this thesis for clinical practice	170
6.7	Future developments	170
	SUMMARY	174
	MY CONTRIBUTION TO THIS THESIS	180
	ACKNOWLEDGEMENTS	183
	APPENDICES	185

Abstract

Background

A variety of tests are used in clinical practice to help the diagnostic process and so improve patient care. Many aspects of stroke management depend on accurate and rapid diagnosis. Brain imaging, including CT or MRI, is necessary to identify the location and extent of the cerebral lesion, and to determine the pathological type of stroke and its likely cause. Current treatments - such as thrombolysis - for ischaemic stroke have increased the need for clear evidence on which imaging test is optimal for diagnosis in the acute phase of stroke.

Systematic reviews of diagnostic test accuracy may provide evidence on the best use of a diagnostic test in clinical practice and help clinicians to decide among alternative tests. The Cochrane Collaboration has recently included systematic reviews of diagnostic test accuracy within its remit. However, to prepare Cochrane systematic reviews of diagnostic test accuracy is challenging because the methods for such reviews are still in a state of flux.

Materials and methods

The research work undertaken for this thesis addresses four relevant methodological aspects of such reviews and, I hope, will contribute to informing the development of the methods for Cochrane systematic reviews of test accuracy:

- i) I assessed the quality of reporting of imaging studies in stroke medicine published between 1995 and 2008 with the current STAndards for the Reporting of Diagnostic accuracy studies (STARD) criteria;
- ii) I assessed the magnitude of publication bias in diagnostic accuracy studies in stroke medicine, by reviewing all diagnostic abstracts presented at two international stroke meetings between 1995 and 2004 and so evaluating the characteristics and findings of the identified abstracts;
- iii) I have evaluated the methods for preparing reviews of test accuracy by undertaking a pilot review according to the draft recommendations of the Cochrane Diagnostic Test Accuracy Working Group;
- iv) I conducted a survey to assess a) how well clinicians and health professionals interpret findings of Cochrane systematic reviews of diagnostic accuracy presented in summary documents; and b) what is the best format for summarising findings of Cochrane reviews of diagnostic accuracy.

Conclusions

In conclusion, methodological issues concerning the validity and reliability of findings of studies included in systematic reviews of diagnostic accuracy remain of fundamental importance. More empirical evidence is needed to address potential biases such as reporting bias and publication bias. To allow dissemination of diagnostic reviews findings in clinical practice better ways of communicating main characteristics and key results of systematic reviews of diagnostic accuracy should be considered.

In the current literature, the quality of reporting and methodological quality of imaging studies for the diagnosis of stroke is less than satisfactory and leaves room for improvement. This is worrying, especially if current health imaging policies are in fact based on poor quality evidence and hence scarce health resources may not be deployed as effectively as they could be.

1. Introduction

Good clinical care depends on good decision making by clinicians. Good decision making in health care should in turn be based on best evidence, which comes preferably from systematic reviews. Health care providers, policy makers, researchers and consumers need therefore somehow to keep up to date with the enormous volume of evidence that is rapidly accumulating in the medical literature.¹ Systematic reviews provide a concise, up to date and efficient synthesis of all the available relevant evidence.² Since the results of systematic reviews are used to inform and guide decision making on health care, it is crucial that their methodology is precise, reliable, reproducible and makes every effort to avoid bias. Although the methods for preparing, assessing, and interpreting systematic reviews of interventions are well established; the methods for conducting systematic reviews of diagnostic accuracy studies are not yet fully developed.

1.1 Evaluating diagnostic tests: methodological aspects

In the medical field, diagnostic tests are used to reduce uncertainty about the presence or absence of target clinical conditions. Over the past few decades many new diagnostic technologies have been developed; in imaging methods, biochemistry-based test procedures, and genetics to name a few. Systematic reviews of diagnostic accuracy are therefore potentially useful in informing decisions about the use of particular diagnostic tests in clinical practice. ‘Diagnostic accuracy’ indicates the amount of agreement between the results of the index test and those of a reference standard, which is the best available method to establish the presence or

absence of the condition of interest or target condition. The optimal design for studies of diagnostic test accuracy is a cross-sectional study in which the participants' test results are cross-classified against the results of the reference standard in order to distinguish individuals with the target condition from those without the target condition. A main objective of systematic reviews of diagnostic test accuracy is to compare the relative performance of two or more tests. Diagnostic tests can be compared in the same studies in the same patient populations (direct comparison) or in separate studies in different patient populations (indirect comparison). Systematic reviews of test accuracy that compare tests directly provide useful information to assist clinical decisions on whether one test is superior to its alternative tests or on whether a new test may replace an existing one. In Chapter 4 I report a pilot systematic review conducted according to the recommendations of the Cochrane Diagnostic Test Accuracy Working Group.

Some of the key methodological challenges of conducting systematic reviews of diagnostic test accuracy are considered below.

1.1.1 Electronic search strategies

Although the clinical epidemiological methods for the evaluation of a diagnostic test are well established, the methods to synthesise the evidence from a series of test accuracy studies in a systematic review are not so well developed as those for systematic reviews of randomised clinical trials of interventions.³ Systematic reviews

of diagnostic test accuracy pose a number of methodological challenges. For example, searching the medical literature to identify studies of diagnostic accuracy is more challenging than for randomised controlled studies.⁴ Diagnostic accuracy studies are not uniformly and unequivocally indexed in major electronic databases such as MEDLINE and EMBASE. The MeSh heading ‘sensitivity and specificity’ was introduced in MEDLINE in 1991 to replace the previous term ‘sensitivity, specificity (epidemiology)’ that was introduced for the first time only in 1987. However, the term ‘sensitivity and specificity’ is often applied inconsistently and it may fail to identify all relevant diagnostic accuracy reports or, on the contrary, wrongly retrieve reports which are not diagnostic. The use of methodological search filters to restrict the search to reports of test accuracy are usually not recommended because they can miss potential relevant studies and may affect the sensitivity of the search.⁵

In systematic reviews of randomised clinical trials the assessment of whether studies with positive results are more likely to be published than studies with negative results (publication bias) and the potential impact of this type of bias on the overall findings of the review have been extensively studied.⁶⁻¹⁰ In contrast to the substantial evidence available for randomised clinical trials, at present there is little evidence on publication bias in diagnostic research. The association between publication and positive findings has been demonstrated for randomised clinical trials by i) direct evidence which involved the follow up of cohort of studies registered by ethics committees; and ii) indirect evidence which included the observation of a high proportion of studies with positive findings in the published literature as well as the

association between sample size and effect size. The magnitude, determinants, and impact of publication bias for diagnostic accuracy studies have yet to be assessed. Location of a reliable datasource for diagnostic test accuracy studies it is, however, problematic as often diagnostic studies rely on data collected primarily as part of clinical care and do not require formal registration. Irwig and colleagues claimed that publication bias could be indeed more of a problem for studies of diagnostic test accuracy than for randomised clinical trials because of the absence of a clear record of attempted clinical evaluations.¹¹ Moreover, the definition of what constitutes a 'positive result' is more complex in studies of diagnostic test accuracy. Deeks and colleagues maintained, in fact, that the determinants of publication bias are likely to be different for diagnostic test accuracy studies as their statistical analyses involve the estimation of measures of accuracy (e.g. sensitivity and specificity) instead of the computation of a p-value following the formulation of a null hypothesis.¹² Furthermore, the methods we have from randomised controlled trials for evaluating publication bias do not seem to work well with test accuracy studies.¹² In Chapter 3 I evaluated the scale of publication bias in studies of diagnostic test accuracy by assessing the full publication status of research initially presented as abstracts in scientific meetings.

1.1.2 Quality of reports of diagnostic accuracy studies

Studies of diagnostic accuracy may appear of poor methodological quality because they are either poorly reported or poorly conducted.¹³ Therefore, the assessment of the quality of reporting of primary studies together with the assessment of their methodological quality are crucial steps for identifying potential biases which might

affect the validity and generalisability of findings and conclusions of diagnostic accuracy reviews.

1.1.3 Statistical methods for synthesising results

Systematic reviews of diagnostic test accuracy differ from those of randomised controlled trials also regarding their summary statistics and the methods for pooling results across studies. There are many different ways for summarising test accuracy results in a systematic review. To synthesis diagnostic test results across studies is particularly challenging due to the fact that there is no a single measure to define 'accuracy'. In studies of diagnostic test accuracy two or more statistics are used to express the performance of a test: sensitivity and specificity, positive and negative predictive values, likelihood ratios of test results, or Receiver Operating Characteristics curve (ROC). The recommended method for synthesising data in the context of meta-analyses of diagnostic accuracy studies is the summary ROC curve (SROC) which, in contrast for example with the simple pooling of sensitivities and specificities, has the advantage to take into account the underlying relationship between these measures of accuracy. The most advanced methods for fitting summary ROC curves are the hierarchical summary ROC model and the bivariate random-effects model.^{14,15} These models allow for both within and between study variability, and also for a statistical correlation between pairs of sensitivity and specificity estimates across studies. The parameters estimate from either model can be used to fit a SROC curve and produce a summary operating point. However, both models are very complex, more accurate when the number of studies is large, and not

at all easy to use. They require statistical skills to be fitted and interpreted and sophisticated softwares, which allow fitting of mixed models. To synthesise and interpret results of studies of test accuracy is difficult and there is a clear need for a consensus on which method is the most appropriate for conducting meta-analyses.

In this thesis I did not focus on the statistical aspects of the current recommended methods for synthesising diagnostic data. Nevertheless, I had to consider these methods in preparing the Cochrane systematic review presented in Chapter 4.

1.1.4 Helping clinicians interpret summary findings

The way in which the test accuracy results are synthesised in systematic reviews has important implications for clinical practice. There is evidence for example that clinicians may confuse measures of diagnostic accuracy.¹⁶ The main objective of systematic reviews of diagnostic accuracy is to help clinicians and health professionals to make informed decisions about the use of medical tests. The main findings of a systematic review are presented in a ‘summary’ that is simple, concise, and easy to interpret at the beginning of the review. Summaries aim at summarising the key features, results, and conclusions of systematic reviews and can be read as stand-alone documents.¹⁷ They can have a verbal structure (abstract) or a tabular format (summary of findings tables). The Cochrane Collaboration recommends now including ‘summary of findings’ tables in the structure of its reviews to provide information on the quality of evidence and main studies results.¹⁶ However, there is hardly any evidence in the literature on how clinicians and health professionals interpret the findings of systematic reviews of diagnostic accuracy summarised in

abstracts or summary of findings tables, and on their preferences as to the contents and format of these summary documents.

1.2 Stroke diagnosis

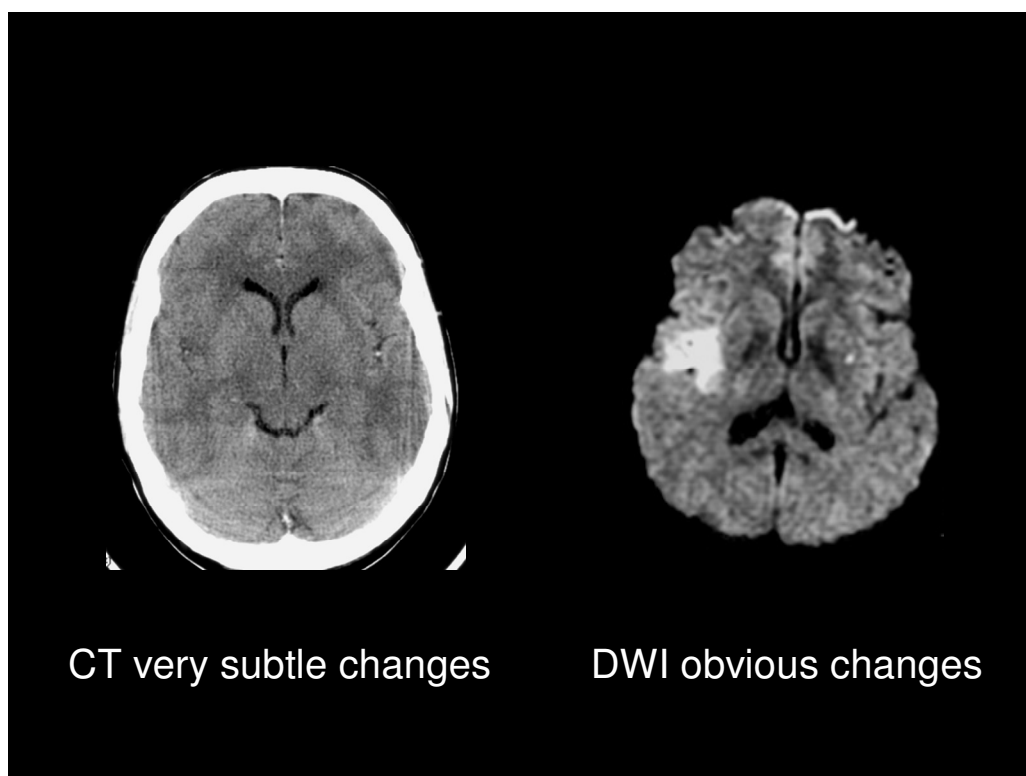
Much of stroke management is dependent on accurate diagnosis. Many areas of stroke management (acute treatment, secondary prevention, and rehabilitation) require diagnostic methods that can be proven to be reliable, repeatable and valid; this applies equally to laboratory tests as to bedside clinical procedures. In particular, new treatments to be given in the first few hours after stroke onset depend clearly on accurate diagnosis to identify the nature and location of the cerebral vascular lesion and ensure appropriate interventions. A very wide range of diagnostic technologies is now employed in stroke medicine including CT and MRI scanning. Some of these diagnostic technologies are not only expensive, but their availability varies across hospitals and countries.¹⁸ Developing diagnostic strategies that are not only cost-effective but also appropriately and equitably available to the whole population (including remote and rural areas) is still a major challenge for any health care system. Healthcare professionals and policy makers have to face regularly decisions concerning the best choice of diagnostic tests and interpretation of their findings. Inappropriate tests or wrong interpretations of their findings may lead to possible dangerous delays in making a correct diagnosis and therefore starting proper patient management or, on the hand, may lead to false diagnoses and unnecessary treatments.

1.3 Imaging diagnosis in stroke medicine

Recent technological developments in imaging have greatly contributed to the diagnosis of stroke by allowing visualization and characterisation of the brain vascular lesions, which may explain the patients' stroke symptoms. Non-contrast or plain computed tomography (CT) is the most commonly used imaging method to assess patients with signs and symptoms of acute stroke. CT produces axial images of a number of sections through the whole brain. Signs of abnormality are shown on CT scans as abnormal tissue density and hyperattenuated artery.^{19,20} CT, when performed early after stroke onset, can accurately detect intracerebral haemorrhage. Although the capability of CT to detect ischaemic changes early after acute stroke has improved since the technology was first introduced in clinical practice, overall in the acute phase CT shows the appropriate ischaemic lesion in only half of the patients with an acute ischaemic stroke.^{21,22} Larger, more extensive ischaemic lesions are more likely to show than small ischaemic lesions.^{22,23} Therefore, an early 'negative' CT for ischaemic signs or lesions does not rule out a stroke diagnosis in patients presenting with stroke neurological symptoms. CT is considered the first-line investigation for stroke patients, as it is widely available in many hospitals, in many countries and is simple and quick to perform, even in uncooperative patients.^{24,25} Magnetic resonance imaging (MRI) is a non-invasive method to image the brain which employs radiofrequency radiation in the presence of magnetic fields.²⁶ MRI is considered more sensitive than CT for detecting early ischaemic stroke (Figure 1.1).²⁷ In particular, MR diffusion weighted imaging (DWI) has recently gained a major role in the evaluation of stroke patients due to the fact that it can show even very small ischaemic lesions within a few minutes of stroke onset.²⁸⁻³⁰

DWI measures the translational movement of water molecules and is usually performed by means of single shot echoplanar imaging (EPI). DWI is commonly interpreted in conjunction with an apparent diffusion coefficient (ADC) map, which is derived from DWI data and is displayed as a grey scale image.²⁴ Abnormal diffusion is shown as a signal of hyperintensity and a reduced ADC.³¹ It is important, for correct interpretation of DWI images to view the DWI sequences together with the corresponding ADC map. This enables the radiologist to distinguish between acute, subacute and chronic ischaemic lesions and high DWI signal that is not associated with restricted diffusion (this can occur if the T2 weighted images show a particularly intense signal - T2 'shine through'). In clinical practice DWI can also be used to detect the area of diffusion-perfusion mismatch representing the ischaemic penumbra (i.e. ischaemic tissue that is not destined become necrotic and hence has the potential to be saved by reperfusion therapy).³² Overall, MRI is considered useful to evaluate the extent, the anatomical distribution and age of ischaemic lesions.²⁴ However, MRI may fail to identify hyperacute intracerebral haemorrhage³³ and the claim that MRI should be the preferred diagnostic imaging technique for the early diagnosis of patients with suspected acute stroke³⁴ remains to be proven. Furthermore, large numbers of acute ischaemic patients either do not tolerate MRI or have a medical contraindication to exposure to high magnetic field (e.g. cardiac pacemaker).³⁵

Figure 1.1. CT taken within a few hours of onset of stroke symptoms shows very subtle ischaemic changes. MR DWI performed shortly after CT shows clear high signal abnormalities.



1.4 The Cochrane Collaboration

At present, the key international organisation which produces and promotes systematic reviews to help people making well-informed decisions about health care is the Cochrane Collaboration (www.cochrane.org). The Cochrane Collaboration recognised the methodological challenges posed by studies of diagnostic test accuracy and started to develop a programme of work to include systematic reviews of diagnostic accuracy within its remit in 2001, during the Cochrane Colloquium in Lyon. Soon after, a Cochrane Diagnostic Test Accuracy Working Group was formed by the Steering Group of the Cochrane Collaboration. The main remit of this Working Group was to develop the methodology and implement the publication of

systematic reviews of diagnostic test accuracy within the Collaboration. In October 2007, during the Cochrane Colloquium in Sao Paulo, the implementation of systematic reviews of diagnostic test accuracy was officially launched (<http://srdta.cochrane.org>). In particular, the work of the Screening and Diagnostic Methods Group and of the Diagnostic Test Accuracy Working Group has provided inspiration and impetus for the completion of this thesis.

1.5 Outline of the Thesis

This thesis addresses specific methodological areas related to systematic reviews of diagnostic accuracy in stroke medicine: the quality of reporting of studies of diagnostic accuracy; the magnitude of publication bias in studies of diagnostic accuracy; methods for preparing reviews of test accuracy; and, interpretation of summary findings of Cochrane reviews of diagnostic test accuracy.

Chapter 2 consists of a study that assesses the accuracy and completeness of reporting of diagnostic accuracy studies in stroke medicine using the current STAndards for Reporting of Diagnostic Accuracy (STARD) criteria. The chapter focuses, specifically, on the use of magnetic resonance imaging for the early diagnosis of ischaemic stroke. **Chapter 3** consists of a study that assesses the scale of publication bias in studies of diagnostic test accuracy in stroke medicine. **Chapter 4** reports a systematic review of diagnostic test accuracy that evaluates two imaging methods for the early diagnosis of stroke. This systematic review served as a pilot review to test and evaluate draft methods proposed by the Cochrane Diagnostic Test Accuracy Working Group. **Chapter 5** consists of a survey which aims to evaluate how clinicians and health professionals interpret findings of systematic reviews of

diagnostic test accuracy presented in two summary documents. **Chapter 6** summarises the main findings of this thesis, together with their implications for clinical practice and future research.

References

1. Egger M, Davey Smith G, O'Rourke K. Rationale, potentials, and promise of systematic reviews. In Egger M, Davey Smith G, Altman DG (Editors): Systematic reviews in health care. Meta-analysis in context. BMJ Publishing Group 2001
2. Mulrow CD. Systematic reviews: rationale for systematic reviews. *BMJ* 1994; 309: 597-9
3. Deeks JJ. Systematic reviews of evaluations of diagnostic and screening tests. In: Egger M, Davey Smith G, Altman DG (editors), Systematic Reviews in Health Care. Meta-analysis in context. London (UK): BMJ Books 2001
4. Haynes RB, Wilczynski N, McKibbon KA, Walker CJ, Sinclair JC. Developing optimal search strategies for detecting clinically sound studies in MEDLINE. *J Am Med Inform Assoc* 1994; 1: 447-58
5. Leeflang MMG, Deeks JJ, Gatsonis C, Bossuyt PMM, on behalf of the Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. *Annals of Internal Medicine* 2008; 149: 889-97
6. Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990; 263: 1385-9

7. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991; 337: 867-72
8. Dickersin K, Min YI. NIH clinical trials and publication bias. *Online Journal of Clinical Trials* 1993; 50: 4967
9. Dickersin K, Olson CM, Rennie D, Cook D, Flanagan A, Zhu Q, Reiling J, Pace B. Association between time interval to publication and statistical significance. *JAMA* 2002; 287: 2829-31
10. Liebeskind DS, Kidwell CS, Sayre JW, Saver JL. Evidence of publication bias in reporting acute stroke clinical trials. *Neurology* 2006; 67: 973-9
11. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *Journal of Clinical Epidemiology* 1995; 48: 119-30
12. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of Clinical Epidemiology* 2005; 58: 882-93

13. Deville' WL, Buntinx F. Guidelines for conducting systematic reviews of studies evaluating the accuracy of diagnostic tests. In Knottnerus JA (editor), The Evidence Base of Clinical Diagnosis. London (UK): BMJ Books 2002
14. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluation. *Statistics in Medicine* 2001; 20: 2865-84
15. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* 2005; 58: 982-90
16. Steurer J, Fischer JE, Bachmann LM, Koller M, ter Riet G. Communicating accuracy of tests to general practitioners: a controlled study. *BMJ* 2002; 324: 824-6
17. Schunemann HJ, Oxman AD, Higgins JPT, Vist GE, Glasziou P, Guyatt GH on behalf of the Cochrane Applicability and Recommendations Methods Group and the Cochrane Statistical Methods Group. Presenting results and 'Summary of Findings' tables. In: Higgins JPT and Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester (UK): John Wiley & Sons, 2008

18. Kane I, Whiteley WN, Sandercock PAG, Wardlaw JM. Availability of CT and MR for assessing patients with acute stroke. *Cerebrovascular Diseases* 2008; 25: 375-7

19. Armstrong P, Wastie ML. *Diagnostic Imaging*. Oxford: Blackwell Scientific Publications, 1987

20. Wardlaw JM, Mielke O. Early signs of brain infarction at CT: observer reliability and outcome after thrombolytic treatment - systematic review. *Radiology* 2005; 235: 444-453; doi:10.1148/radiol.2352040262

21. Warlow C, van Gijn J, Dennis M, Wardlaw J, Bamford J, Hankey G, Sandercock P, Rinkel G, Langhorne P, Sudlow C, Rothwell P. What pathological type of stroke is it, cerebral ischemia or hemorrhage? In: *Stroke, A practical guide to management*. 2nd Edition. Oxford: Blackwell Publishing, 2008

22. Wardlaw JM, West TM, Sandercock PAG, Lewis SC, Mielke O. Visible infarction on computed tomography is an independent predictor of poor functional outcome after stroke, and not of haemorrhagic transformation. *Journal of Neurology, Neurosurgery & Psychiatry* 2003; 74: 452-458
doi:10.1136/jnnp.74.4.452

23. Lindgren A, Norrving B, Rudling O, Johansson BB. Comparison of clinical and neuroradiological findings in first-ever stroke: a population-based study. *Stroke*, 1994; 25: 1371-7
24. Rich PM, Jäger HR. Cerebrovascular disease and nontraumatic intracranial haemorrhage. In: Adam A and Dixon AK Editors. *Diagnostic Radiology: A textbook of Medical Imaging*. Fifth Edition, Vol. 2. Elsevier Limited, 2008
25. Moustafa RR, Baron JC. Clinical review: Imaging in ischaemic stroke – implications for acute management. *Critical Care* 2007; 11: 227
(doi:10.1186/cc5973)
26. Wilkinson ID, Paley MNJ. Magnetic resonance imaging: basic principles. In: Adam A and Dixon AK Editors. *Diagnostic Radiology: A textbook of Medical Imaging*. Fifth Edition, Vol. 2. Elsevier Limited, 2008
27. Keir SL, Wardlaw JM. Systematic review of diffusion and perfusion imaging in acute ischemic stroke. *Stroke*, 2000; 31: 2723 - 2731.

28. Hjort N, Christensen S, Solling C, Ashkanian M, Wu O, Rohl L, Gydensted C, Andersen G, Ostergaard L. Ischemic injury detected by diffusion imaging 11 minutes after stroke. *Annals of Neurology* 2005; 58: 462-65
29. Gonzalez RG, Schaefer PW, Buonanno FS, Schwamm LH, Budzik RF, Rordorf G, Wang B, Sorensen AG, Koroshetz WJ. Diffusion-weighted MR imaging: diagnostic accuracy in patients imaged within 6 hours of stroke symptom onset. *Radiology*, 1999; 210: 155-62
30. Muir KW, Buchan A, von Kummer R, Rother J, Baron JC. Imaging of acute stroke. *Lancet Neurology*, 2006; 5: 755-68
31. Baird AE, Benfield A, Schlaug G, Siewert B, Lovblad KO, Edelman RR, Warach S. Enlargement of human cerebral ischemic lesion volumes measured by diffusion-weighted magnetic resonance imaging. *Annals of Neurology* 1997; 41: 581-9
32. Kesavadas C, Fiorelli M, Gupta AK, Pantano P, Bozzao L, Kapilamoorthy TR. Diffusion weighted magnetic resonance imaging in acute ischemic stroke. *Indian J Radiol Imaging* 2003; 13: 433-40

33. Wardlaw JM, Farrall AJ. Diagnosis of stroke on neuroimaging. “Scan all immediately” strategy improves outcome and reduce costs. *BMJ*, 2004; 328: 655-6
34. Chalela JA, Kidwell CS, Nentwich LN, Luby M, Butman JA, Demchuk AM, et al. Magnetic resonance imaging and computed tomography in emergency assessment of patients with suspected acute stroke: a prospective comparison. *Lancet* 2007; 369: 293-8
35. Hand PJ, Wardlaw JM, Rowat AM, Haisma JA, Lindley RI, Dennis MS. Magnetic resonance brain imaging in patients with acute stroke: feasibility and patient related difficulties. *Journal of Neurology, Neurosurgery & Psychiatry* 2005; 76: 1525-1527 doi:10.1136/jnnp.2005.062539

2. The quality of reporting of diagnostic accuracy studies on the use of magnetic resonance imaging for the early detection of vascular lesions in stroke patients

2.1 Background

Effective and efficient management of patients with acute stroke relies heavily upon a rapid and reliable diagnostic process. Current imaging technology (e.g. MRI and CT) provides essential diagnostic information which contributes to optimal therapeutic choices. Diagnostic accuracy studies on imaging are used to assess the accuracy of a test in diagnosing the presence of a vascular brain lesion and distinguishing between ischaemic and haemorrhagic lesions. There is evidence in the literature that studies of diagnostic test accuracy may fail to fulfil essential methodological standards.^{1,2} Methodologically biased studies may affect estimates of diagnostic accuracy and provide misleading results. Lijmer and colleagues in 1999³ and Rutjes and colleagues in 2006⁴ demonstrated that differences in study design and patient selection are associated with variations in the estimates of accuracy. Case-control design, retrospective data collection, and non-consecutive inclusion of patients were amongst the characteristics associated with higher estimates of diagnostic accuracy. Methodological characteristics of diagnostic accuracy studies may, however, be difficult to ascertain if the quality of reporting is poor. For studies of diagnostic test accuracy, complete and accurate reporting is an essential prerequisite on which to judge the potential occurrence of methodological flaws. Following the successful CONSORT initiative for enhancing the quality of reports of

randomised controlled trials,⁵ the STARD (STAndards for the Reporting of Diagnostic accuracy studies) group in 2003 published a statement to improve the accuracy and completeness of reporting of studies of diagnostic accuracy.^{6,7} The STARD statement consists of a checklist of 25 items and recommends the use of a flow diagram which provides the exact number of participants at each stage of the study. The STARD checklist has been published and promoted by several medical journals. Research studies have also been conducted to assess the quality of reporting of articles of diagnostic test accuracy published before 2003.⁸⁻¹⁰ Overall the quality of reporting varied greatly across STARD items. Two published reviews (one assessing all diagnostic articles published in journals with a high impact factor and one assessing diagnostic articles published in ophthalmic journals) found that only about 40% of the identified diagnostic articles reported on more than 50% of the STARD items.^{8,9} Smidt and colleague¹¹ have also assessed whether the quality of reporting of diagnostic accuracy studies has improved since publication of the STARD statement. They looked at the quality of reporting of all diagnostic accuracy studies published in 12 medical journals in 2000 (pre STARD) and 2004 (post-STARD) and concluded that the quality of reporting of diagnostic accuracy studies has only marginally improved over time.¹¹

The purpose of this study was to evaluate the extent to which studies on diffusion-weighted magnetic resonance imaging for the early diagnosis of stroke published between 1999 and 2008 comply with the STARD criteria and to explore whether the introduction of the STARD statement has contributed to a better quality of reporting.

2.2 Materials and Methods

2.2.1 Study Selection

I searched the MEDLINE and EMBASE electronic databases to identify reports published between 1999 and 2008 (five years before STARD publication and five years after STARD publication). Reports were included if 1) they focused on magnetic resonance imaging including diffusion-weighted sequences for the early detection of ischaemic vascular brain lesions, 2) they investigated patients suspected of stroke, 3) they were primary studies of diagnostic test accuracy. The search strategies were developed and tested in close collaboration with the Cochrane Stroke Group Trial Search Coordinator, Brenda Thomas, who is an internationally recognised expert in the field and has been working for the Cochrane Collaboration since 1996. The searches were limited to reports focused on human research, published in English or Italian. In particular, the language restrictions were applied a) because of the limited resources available (i.e. no funding for professional translations of abstracts or full papers; the Cochrane Stroke Group could not provide access to volunteer translations) and b) to follow the methods used in previous assessments of compliance with the STARD criteria, where selection of papers was restricted to the English language.^{8,9,11,12}

Details of the MEDLINE and EMBASE searches are reported in Appendix 1.

All citations of all potential eligible reports were screened by a single reviewer (myself). Only full-text reports were deemed suitable for inclusion.

2.2.2 Data extraction

The STARD checklist was used to assess the quality of reporting of all relevant reports of diagnostic accuracy. A single reviewer (myself) established whether each item of the STARD checklist was adequately described in the text of the identified reports. Only the quality of reporting was assessed and not the risk of potential methodological biases. I was not blind to the source and details of publications. Studies of diagnostic accuracy were defined as studies in which the results of one (or more tests) were compared with those of a suitable reference standard in the same patient population. Either clinical assessments accompanied by imaging follow-up, or autopsy findings were considered suitable reference standards.

Following Wilczynski,¹² for the purpose of this study item 13 (“Describe methods for calculating test reproducibility, if done”), item 23 (“Report estimates of variability of diagnostic accuracy between subgroups of participants, readers or centres, if done”), and item 24 (“Report estimates of test reproducibility, if done”) were removed from the STARD checklist as they all contain a hypothetical clause (“if done”). It was considered impracticable to establish for the identified diagnostic reports whether the lack of reporting was because the items were assessed but not reported or because they were simply not assessed. Thus, only 22 of the original 25 STARD items were considered for this assessment. Equal weights were applied to each item in the list with the exception of items 8, 9, 10, and 11. These four items concern the index test as well as the reference standard. Thus, each of them was split into two statements, one for the index test and one for the reference standard. Each statement was evaluated separately and was assigned a maximum score of 0.5 point (see Table 2.2).

Reports published between 1999 and 2003 (pre STARD checklist) were evaluated separately from reports published between 2004 and 2008 (post STARD checklist).

2.2.3 Statistical Analysis

For each individual item of the STARD checklist the total number of reports that adequately described the elements needed to satisfy that item was calculated for the years 1999-2003 and for the years 2004-2008. For each item, the percentage difference between pre- and post-STARD studies was calculated. For each report a total STARD score was calculated by summing up the number of all reported items (0-22 score range). Higher score indicated better quality of reporting. The overall mean scores and standard deviations of reports published 1999-2003 and reports published 2004-2008 were calculated. The difference in the overall mean score between pre-STARD reports and post-STARD reports was calculated by means of a two-tailed t-test for independent samples. Statistical significance was set at $p = 0.05$. The analyses were carried out using MINITAB (MinitabR 15.1.20.0.).

2.3 Results

The search strategies identified 2408 citations. After screening the titles and abstracts 59 reports were considered potentially relevant and retrieved in full. Thirty-four reports were subsequently excluded because they did not meet the inclusion criteria. I assessed a total of 25 diagnostic reports published in 14 different journals (Table 2.1). Eighteen reports were published between 1999 and 2003 and seven reports between 2004 and 2008. The mean 2-year impact factor of journals that published

reports between 1999 and 2003 was similar to that of journals that published reports between 2004 and 2008 (2.969 versus 3.246). Twenty-four of the assessed reports were cohort studies and one was a case-control study (published in 2001). The quality of reporting of the individual STARD items according to years of publication is shown in Table 2.2.

There was a wide variation in the quality of reporting of individual STARD items (0-100%). At an initial descriptive assessment, 10 of the 26 STARD items evaluated (items 8, 9, 10 and 11 counted twice) were more frequently reported in diagnostic reports published between 1999 and 2003, whilst 14 were more frequently reported in reports published between 2004 and 2008. Item 25 (discussion on applicability of results) and item 8 (technical specification of the index test – i.e. magnetic resonance sequences) were reported in all studies pre and post STARD.

Table 2.1. Number of diagnostic reports according to Journal type and years of publication

Journals	1999-2003	2004-2008
Acta Radiologica	2	-
American Journal of Roentgenology	1	-
American Journal of Neuroradiology	2	2
Archives of Neurology	1	-
European Neurology	-	1
Journal of Clinical Imaging	-	1
Journal of the Formosan Medical Association	1	-
Journal of Neurology Neurosurgery & Psychiatry	-	1
La Radiologia Medica	1	-
Lancet	-	1
Neurology	2	-
Neuroradiology	3	1
Radiology	2	-
Stroke	3	-

Note: mean 2-year impact factor for journals that published 1999-2003 reports = 2.96;9 mean 2-year impact factor for journals that published 2004-2008 reports = 3.246

Table 2.2. Reporting of individual STARD items in diagnostic accuracy reports published between 1999-2003 and between 2004-2008

STARD items Total = 22	Studies for 1999-2003 Total = 18	Studies for 2004 -2008 Total = 7	Difference between pre- and post- STARD studies
Title abstract and keywords			
1 Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	13 (72%)	5 (71%)	- 1%
2 State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups	16 (89%)	6 (86%)	- 3%
Methods			
3 Describe the study population: The inclusion and exclusion criteria, setting and locations where the data were collected.	9 (50%)	4 (57%)	+ 7%
4 Describe participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	15 (83%)	6 (86%)	+ 3%
5 Describe participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in items 3 and 4? If not, specify how participants were further selected.	5 (28%)	6 (86%)	+ 58%
6 Describe data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	14 (78%)	5 (71%)	- 7%
7 Describe the reference standard and its rationale	15 (83%)	7 (100%)	+ 17
8 Describe technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	18 (100%)	7 (100%)	0
a) for index test	17 (94%)	7 (100%)	+ 6%
b) for reference standard			
9 Describe definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard.	14 (78%)	6 (71%)	- 7%
a) for index test	13 (72%)	6 (71%)	- 1%
b) for reference standard			
10 Describe the number, training and expertise of the persons executing and reading the index tests and the reference standard.	16 (88%)	7 (100%)	+ 12%
a) for index test	14 (78%)	5 (71%)	- 7%
b) for reference standard			

Table 2.2. Reporting of individual STARD items in diagnostic accuracy reports published between 1999-2003 and between 2004-2008 - *continued*

STARD items Total 22	Studies for 1999-2003 Total = 18	Studies for 2004 -2008 Total = 7	Difference between pre- and post- STARD studies
11 Describe whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.	15 (83%)	7 (100%)	+ 17%
a) for index test	11 (61%)	5 (71%)	+ 10%
b) for reference standard			
12 Describe methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	3 (17%)	3 (43%)	+ 26%
Results			
14 Report when study was done, including beginning and ending dates of recruitment.	11 (61%)	7 (100%)	+ 39%
15 Report clinical and demographic characteristics of the study population (e.g. age, sex, spectrum of presenting symptoms, comorbidity, current treatments, recruitment centers).	8 (44%)	3 (43%)	- 1%
16 Report the number of participants satisfying the criteria for inclusion that did or did not undergo the index tests and/or the reference standard; describe why participants failed to receive either test (a flow diagram is strongly recommended).	11 (61%)	3 (43%)	- 18%
17 Report time interval from the index tests to the reference standard, and any treatment administered between.	6 (33%)	3 (43%)	+ 10%
18 Report distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	11 (61%)	2 (28%)	- 33%
19 Report a cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	3 (17%)	2 (28%)	+ 11%
20 Report any adverse events from performing the index tests or the reference standard.	0	1 (14%)	+ 14%
21 Report estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	9 (50%)	5 (71%)	+ 21%
22 Report how indeterminate results, missing responses and outliers of the index tests were handled.	4 (22%)	0	- 22%
25 Discuss the clinical applicability of the study findings.	18 (100%)	7 (100%)	0

About 70% of reports in 1999-2003 and 2004-2008 were identified as “diagnostic” mainly due to information contained in their abstracts. Only 33% of reports in 1999-2003 and 14% in 2004-2008 included diagnostic terms (e.g. ‘diagnosis’, ‘sensitivity’, ‘diagnostic accuracy’) in their titles. Items related to the description of results were poorly reported (< 30%) in both diagnostic reports pre-STARD and diagnostic reports post-STARD. In particular, reports lacked several important items: a cross tabulation of results of the index test by the results of the reference standard; a description of how indeterminate results were handled; a description of diagnoses in patients without the target condition; and a description of any adverse events related to performing the index test or the reference standard (see Table 2.2). About 45% of the diagnostic reports published pre- and post-STARD lacked a detailed description of the study population in which the imaging tests were administered (item 4) and about 55% failed to describe adequately the clinical and demographic characteristics of the patient population (item 15). The participant sampling was more frequently reported in studies post-STARD (86%) compared to studies pre-STARD (28%). None of the assessed diagnostic reports, however, included a flow diagram of study participants (as recommended in item 16 of the list). The methods for calculating or comparing measures of diagnostic accuracy were reported in only 17% of pre-STARD studies and 43% of post-STARD studies and estimates of diagnostic accuracy with 95% confidence intervals were reported in 50% of pre-STARD studies and 71% of the post-STARD studies. The time interval from imaging tests and reference standard was reported in 33% of pre-STARD studies and 43% of post-STARD studies. The maximum number of reported items in a single diagnostic report was similar for pre- and post-STARD reports (16 vs 17).

Table 2.3 shows that the mean total number of reported STARD items for diagnostic reports published between 1999 and 2003 was 12.83 (SD 2.54) with a range of 8-16 and for reports published between 2004 and 2008 was 14.43 (SD 2.35) with a range of 9.5-17. The mean difference was not statistically significant ($p = 0.165$).

Table 2.3. STARD mean score according to years of publication

	Mean Score/22 (SD)	Score Range	Significance Level
1999-2003	12.83 (2.54)	8-16	$p = 0.165$
2004-2008	14.43 (2.35)	9.5-17	

SD: standard deviation

2.4 Discussion

To my knowledge this is the first study that has appraised the quality of reporting of diagnostic accuracy studies of diffusion-weighted resonance imaging for the diagnosis of stroke. The topic is clinically relevant, and is the subject of ongoing research to determine the cost-effectiveness of this imaging test, which is expensive and has only limited availability in many centres in the UK.¹³ It is therefore of considerable interest to assess the quality of the evidence upon which current NHS policy is based. This study indicates that the quality of reporting in diagnostic accuracy studies of imaging in this aspect of stroke medicine is modest and that there is a wide variation in the quality of reporting of individual STARD items. The mean number of reported STARD items in diagnostic reports published between 2004 and 2008 (14.43) was not significantly different to that of diagnostic reports published

between 1999 and 2003 (12.83). These findings are in line with the results on quality of reporting from current diagnostic research and other clinical specialities.^{9,11,12,14} Smidt and colleagues found that, in journals with a high impact factor (≥ 4), the quality of reporting of diagnostic accuracy improved only slightly in 2004, after publication of the STARD checklist.¹¹ More recently Wilczynski assessed diagnostic studies published in 12 journals in 2001, 2002, 2004, and 2005 and found that after publication of the STARD checklist the quality of reporting remained similar to the pre-STARD standards.¹² Similarly, Coppus and colleagues evaluated the extent to which test accuracy studies published in two major reproductive medicine journals in 1999 and 2004 adhered to the STARD criteria. They did not find a significant improvement in the number of reported items for the reports published after introduction of the STARD statement.¹⁴

The current study assessed reports on magnetic resonance imaging for the diagnosis of stroke published five years before STARD publication (1999-2003) and five years after STARD publication (2004-2008) but failed to detect a statistically significant overall improvement in the quality of reporting. The lack of improvement observed in this study may be explained by the very small sample of diagnostic reports identified in the literature. In particular, only seven reports were identified for the period 2004-2008, after the STARD publication, compared with 18 reports for the period 1999-2003, pre-STARD. The descriptive assessment of individual STARD items showed that for 10 items the quality of reporting appeared slightly worse after publication of the STARD checklist (see Table 2.2). This may indicate that the design principles of a high quality diagnostic accuracy report in stroke and imaging

journals are not yet uniformly established or that the editorial control on the quality of diagnostic test accuracy reports is weak. Two of the 14 journals considered in this study had reports published pre- and post-STARD publication but the number of reported items was the same in pre- and post-STARD reports (13.5 versus 13.5). More interesting, only two of the six journals where the seven post-STARD reports were published include the STARD checklist in their instructions for authors. Not even the two reports published in the journals which require application of the STARD criteria actually mentioned the STARD checklist in their methods section. Although this assessment is based on small numbers of reports, it is worth noticing that the range of the scores of the two studies published in the journals adopting the STARD checklist (range 15-17, average: 16) overlapped that of the five reports published in the journals which did not adopt STARD (range 9.5-15.5, average: 13.6). The lack of improvement in the quality of reporting and the fact that the STARD checklist is not mentioned in the methods section of diagnostic reports published in journals which adopt the list may suggest that the editors and peer reviewers have not so far been successful in implementing the STARD guidelines for diagnostic research.

The quality of reporting in studies of diagnostic accuracy is crucial for evaluating the validity and generalizability of the results and for revealing potential methodological flaws in the study design and conduct. It is surprising that about 50% of diagnostic reports failed to provide a detailed description of the study inclusion and exclusion criteria and of the clinical and demographic characteristics of the participants. For imaging studies in stroke medicine this may imply potential spectrum bias and may

seriously hamper the applicability of findings in clinical practice. Moreover, none of the assessed reports included a flow diagram to illustrate the design of the study even though this is strongly recommended by the STARD group.^{6,8} Overall results were scantily reported in both reports published pre-STARD and reports published post-STARD. Even though estimates of diagnostic accuracy and confidence intervals were better reported after STARD publication (71% versus 50% in pre-STARD reports) I believe that this is far from satisfactory. It is indeed very worrying that of the seven post-STARD reports: two failed to report estimates of accuracy and measures of statistical uncertainty (i.e. confidence intervals); four did not describe the methods used for calculating measures of diagnostic accuracy; and, five did not provide a cross-tabulation of test results.

About 70% of reports in 1999-2003 and in 2004-2008 were identified as studies of diagnostic accuracy because the diagnostic terms (i.e. sensitivity, diagnosis, accuracy) appeared in the abstract. Diagnostic terms were only rarely used in the titles of reports. The STARD group recommends the use of the term 'diagnostic accuracy' as publication type in the title and abstract of reports that compare the results of one or more test versus a reference standard.⁶ They also recommend the term 'post-test probability' as a MeSH term in addition to 'sensitivity and specificity'.⁸ The MeSH heading 'sensitivity and specificity' was revised in MEDLINE in 1991 to facilitate retrieval of diagnostic studies. However, the use of this heading is still far from optimal as they may fail to allow the retrieval of studies of diagnostic accuracy or wrongly retrieve reports that are not diagnostic.^{15,16}

The current study has some limitations. The identification of studies of diagnostic accuracy in the literature is difficult.¹⁷ Even though I searched MEDLINE and EMBASE with comprehensive search strategies I cannot exclude the possibility I have missed potentially relevant studies. I chose to restrict the scope of this study to the topic of the Cochrane review for two reasons. Firstly, the diagnostic literature in stroke is very substantial (hence a very broad review might have not been feasible), and secondly, this search would place the methodological quality of the studies of my Cochrane review in an appropriate context. As it turned out (slightly to my surprise, given the wide use of the technology in stroke care) the searches identified only a small number of diagnostic reports on magnetic resonance imaging for acute stroke patients. Therefore, the findings of this study are greatly limited by the small sample of suitable studies available. However, it is worth noting that the findings in this field are comparable to those in other fields of medicine. Table 2.4 summarises the outcome of published reviews assessing the quality of reporting of diagnostic accuracy studies before and after the STARD publication. These studies used a methodology similar to my own. From the table it is clear that, although their sample sizes and periods of assessment varied, they show overlapping results, indicating that the quality of reporting is only slightly but not significantly improved after publication of the STARD statement, and that the magnitude of this improvement is comparable between studies.

Because of resource constraints, only one reviewer (myself) was available to assess the identified diagnostic reports using the STARD checklist and therefore evaluation of the reproducibility of the list proved impossible. Substantial disagreement has

been, however, documented for some of the STARD individual items.¹⁸ Some STARD items are open to interpretation and cover multiple aspects of the same domain. There is therefore the risk to process the same item in slightly different way across studies. This risk is likely to be higher if the assessment is carried out by a single reviewer.

Table 2.4. Results of research studies assessing the quality or reporting pre- and post-STARD publication

Study	STARD items assessed	Year of publication - pre-STARD (no. of reports)	Year of publication – post-STARD (no. of reports)	% of reported items in pre-STARD and post-STARD reports	Difference
Smidt 2006¹¹	25	2000 (124)	2004 (141)	47% vs 54%	+ 7%
Coppus 2006¹⁴	25	1999 (24)	2004 (27)	48% vs 50%	+ 2%
Wilczynski 2008¹²	13	2001-2002 (120)	2004-2005 (120)	61% vs 65%	+ 4%
This study 2010	22	1999-2003 (18)	2004-2008 (7)	56% vs 63%	+ 7%

In conclusion, in stroke medicine, diagnostic accuracy reports on magnetic resonance imaging for detection of stroke lesions were not uniformly reported and there is still ample room for improvement. The methodology of diagnostic research has greatly improved over the past seven years, also thanks to the efforts and enthusiasm of the Screening and Diagnostic Methods Group within the Cochrane Collaboration. If the scientific community believes that STARD is an achievement and a way forward to improve the quality of reporting of diagnostic studies, we all should make a clearer

effort to implement its guidelines, from authors, to reviewers, editors, professional bodies.

The STARD checklist provides a useful guide for designing, writing and reviewing diagnostic reports. Editors and peer reviewers should enforce the use of the STARD checklist in all diagnostic manuscripts they consider for publication. In particular they should encourage studies' authors to use the term 'diagnostic accuracy' in the title of their reports more consistently, include a flow diagram to provide the exact number of participants at each stage of the study, and provide a clear description of findings including a cross-tabulation of tests results.

2.4.1 Implications for researchers

- There is a need for authors of individual studies to adhere better to STARD criteria when reporting their methods and results.
- Research funding and ethical approval agencies should consider making adherence to STARD guidelines a mandatory requirement for funding or ethical approval for new studies of diagnostic test accuracy.

2.4.2 Implications for Cochrane diagnostic test accuracy (DTA) reviews and DTA methods groups

- The Cochrane Diagnostic Handbook should provide guidance for review authors on a) how to assess the quality of reporting and b) how to incorporate the quality

of reporting in the exploration of sources of heterogeneity between studies in systematic reviews of diagnostic test accuracy.

2.4.3 Implications for journal editors

- Journal editors should adopt and implement the STARD criteria in their instructions for authors.
- Journal editors should not accept manuscripts reporting studies of diagnostic accuracy which do not comply with current STARD criteria.

References

1. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *Jama* 1995; 274: 645-51.
2. Harper R, Reeves B. Compliance with methodological standards when evaluating ophthalmic diagnostic tests. *Invest Ophthalmol Vis Sci* 1999; 40: 1650-7.
3. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JHP, Bossuyt PMM. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; 282: 1061-6.
4. Rutjes AWS, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PMM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006; 174(4). DOI:10.1503/cmaj.050090.
5. Moher D, Jones A, Lepage L, for the CONSORT group. Use of the CONSORT statement and quality of reports of randomised trials: a comparative before- and after-evaluation. *JAMA* 2001; 285:1992-9
6. Bossuyt PMM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Moher D, Rennie D, de Vet HCW, Lijmer JG. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem* 2003a; 49: 7-18
7. Bossuyt PMM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HCW for the STARD Group. Towards complete and accurate reporting of studies of diagnostic accuracy; the STARD initiative. *Radiology* 2003b; 226: 24-8

8. Smidt N, Rutjes AWS, van der Windt DAWM, Ostelo RWJG, Reitsma JB, Bossuyt PM, Bouter LM, de Vet HCW. Quality of reporting of diagnostic accuracy studies. *Radiology* 2005; 235: 347-53
9. Siddiqui MAR, Azuara-Blanco A, Burr J. The quality of reporting of diagnostic accuracy studies published in ophthalmic journals. *Br J Ophthalmol* 2005; 89: 261-5
10. Stengel D, Bauwens K, Rademacher G, Mutze S, Ekkernkamp A. Association between compliance with methodological standards of diagnostic research and reported test accuracy: meta-analysis of focused assessment of US for Trauma. *Radiology* 2005; 236:102-11
11. Smidt N, Rutjes AWS, van der Windt DAWM, Ostelo RWJG, Bossuyt PM, Reitsma JB, Bouter LM, de Vet HCW. The quality of diagnostic accuracy studies since the STARD statement; Has it improved? *Neurology* 2006; 67: 792-7.
12. Wilczynski NL. Quality of reporting of diagnostic accuracy studies: no change since STARD statement publication – before and after study. *Radiology* 2008; 248: 817-23
- 13 Kane I, Whiteley WN, Sandercock PAG, Wardlaw JM. Availability of CT and MR for assessing patients with acute stroke. *Cerebrovascular Diseases* 2008; 25: 375-7
- 14 Coppus SFPJ, van der Veen F, Bossuyt PMM, Mol BWJ. Quality of reporting of test accuracy studies in reproductive medicine: impact of the Standards for Reporting of Diagnostic Accuracy (STARD) initiative. *Fertility and Sterility* 2006; 86(5): 1321-9

- 15 Deville' WL, Bezemer PD, Bouter LM. Publications on diagnostic test evaluation in family medicine journals: an optimal search strategy. *Journal of Clinical Epidemiology* 2000; 53: 65-9
- 16 Leeflang MMG, Deeks JJ, Gatsonis C, Bossuyt PMM, on behalf of the Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. *Annals of Internal Medicine* 2008; 149: 889-97
- 17 Deeks JJ, Bossuyt PMM, Gatsonis C (editors). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Oxford: The Cochrane Collaboration [Forthcoming].
- 18 Smidt N, Rutjes AWS, Van der Windt DAWM, et al. Reproducibility of the STARD checklist: an instrument to assess the quality of reporting of diagnostic accuracy studies. *BMC Medical Research Methodology* 2006; 6:12

3. Is there bias in the process of publication of diagnostic accuracy studies in stroke submitted as abstracts?

3.1 Introduction

Publication bias is defined as the ‘tendency to publish research results based on the strength and direction of a study’s findings’.¹ It has been demonstrated that publication of both observational and experimental studies is influenced by the characteristics of the study, and research findings are less likely to be published if they are shown to be negative rather than positive or if they are based on small patient populations.²⁻⁵ The term ‘publication bias’ is also used in the literature to refer to ‘other biases related to the time, type and language of publication, and multiple publications’.⁶ Publication and other related biases may lead to an overestimation of the magnitude of treatment effects and consequently may affect decisions about patient management. They also represent a serious threat to the reliability of systematic reviews which focus primarily on evidence from the published literature. Whilst there is a substantial literature on publication bias in systematic reviews of randomised controlled trials, there is little empirical evidence on the frequency and determinants of publication bias in systematic reviews of studies of diagnostic test accuracy.⁷ Determinants of publication bias in studies of diagnostic test accuracy are likely to differ from those in clinical trials.⁸ Since many aspects of stroke management (e.g. acute treatment, secondary prevention, investigation of complications) depend on the results of diagnostic tests, we

considered it important to determine whether or not there was evidence of publication bias.

Publication bias in randomised controlled clinical trials has been evaluated by tracing cohorts of trials identified from ethics committees, and investigating determinants of publication within the cohort.¹⁻⁴ Analogous cohorts of studies of diagnostic test accuracy are more difficult to be identified as formal registration of diagnostic research and consequently ethical approval has not uniformly been required. I have followed an alternative approach of a cohort of studies at a point further down the research process, where they are presented as conference abstracts, and investigating determinants of future full publication. Although this approach does not capture the full magnitude of publication bias, I hypothesize that the determinants of full publication in this stage would be similar.

I therefore sought (i) to assess publication bias by determining what proportion of studies of diagnostic accuracy presented as abstracts at international stroke meetings were subsequently published in full in peer-reviewed journals and (ii) to assess which factors were predictive of time to publication.

I focused on abstracts presented at international stroke meetings as 'stroke' is the underlying disease theme of this methodological thesis. I did not consider abstracts presented at international *imaging* conferences because a) they usually cover a variety of different clinical conditions, and individual studies might contain a mixture of stroke and non-stroke cases, which would have introduced unwanted

heterogeneity to the data set; b) hand-searching of proceedings from imaging conferences would have enormously broadened the scope of this study and hence required for greater resources than were available.

3.2 Methods

I reviewed all proceedings of the International Stroke Conference and the European Stroke Conference between 1995 and 2004. All abstracts submitted to both stroke conferences were peer reviewed blind to authorship prior to acceptance. Acceptance rates varied from year to year, with about 60% of all submitted proceedings accepted for presentation in most recent years.⁹ These proceedings were published as abstracts in special issues of *Stroke and Cardiovascular Diseases*. Abstracts were selected for inclusion only if they reported findings of studies of diagnostic test accuracy. Studies of diagnostic test accuracy were defined as studies assessing the diagnostic performance of a single test or two or more tests against a reference standard. I extracted information on study characteristics from selected abstracts. For each relevant abstract the following information was recorded: sensitivity and specificity values, clinical utility of observed results, retrospective versus prospective method of patient recruitment; number of authors, region of origin of the corresponding author, number of clinical centres, sample size, type of test performed, blinding of test results, if there was assessment of inter-observer agreement, and congress presentation (abstracts presented at the International Stroke Conference versus abstracts presented at the European Stroke Conference). Details on whether or not ethical approval was granted were not recorded.

According to the conclusions reported by the original authors and, if given, the estimates of sensitivity and specificity, I categorised the clinical utility of test results as 'accurate', 'possibly useful', or 'non-informative'. Test results were classified as 'accurate' if the diagnostic accuracy of the test(s) under investigation was high enough to recommend its use in clinical practice. Test results were classified as 'possibly useful' if the test(s) under investigation proved to have a good sensitivity but not necessarily a good specificity (and vice versa) or if the test(s) proved to perform well only in certain circumstances (e.g. in severe patients; or patients with a particular type of stroke or cerebrovascular disease). Test results were classified as 'non-informative' if the accuracy of the test under investigation was not good enough to recommend its use in clinical practice or - for comparative studies - if the accuracy of the experimental test was equivalent to or not better than that of an existing alternative test. Furthermore, whenever it was possible, Youden's Index, a measure of test performance (sensitivity + specificity - 1), was calculated. According to the affiliations reported in the abstracts, I grouped abstracts' authors in three categories: authors from English speaking countries (i.e. Australia, Canada, New Zealand, UK, and USA), authors from European countries other than UK, and authors from Asia and South America.

I searched MEDLINE and EMBASE electronic databases up to November 2006 to identify all study reports published in full in peer-reviewed journals. Searches were performed using initially the lead author's surname and initials. If no subsequent publication was identified, additional searches were performed using the surname and initials of the other authors or appropriate keywords from the title of the abstract.

The date and journal of publication were noted. For abstracts for which a full-text publication could not be located with confidence in MEDLINE or EMBASE, a questionnaire was sent or e-mailed to the first author. If the post or e-mail address of the first author was not available the questionnaire was sent to another author of the abstract. The questionnaire asked whether the study had been published or whether a manuscript was submitted, rejected, planned to be submitted, or no longer planned to be completed. If it had been published, authors were asked to provide a full citation of the publication. Studies were considered unpublished when no articles could be located in the international published literature and either the contacted authors confirmed no publication status or no questionnaire was returned.

3.3 Statistical analyses

A Kaplan-Meier survival analysis was conducted to examine the relationship between abstract presentation and time to publication. Potential factors predictive of time of publication were examined - one at a time - using univariate Cox regression analyses and results were expressed as hazard ratios. Time to publication was defined as the time from the date the study abstract was published in Stroke or Cerebrovascular Diseases to the date of the first peer-reviewed full-publication identified in the literature. Studies whose findings were published before abstract presentation were excluded from the survival analysis. Unpublished studies were censored at the time the last literature search was performed (i.e. November 2006). The study factors included in the model were the following: type of study design, number of authors, multi-centre status, sample size, blinding of tests results, inter-observer agreement, type of diagnostic test, clinical utility of results, Youden's

Index, and region of origin of the corresponding author. All *p* values were two tailed with a significance level of 0.05. Missing data were excluded from the analysis. Statistical analyses were performed using SAS 9.1 for Windows.

3.4 Results

One hundred and sixty abstracts met the specified inclusion criteria. Seventy-six percent (121) of all abstracts did not report on blinding and 65% (104) did not mention study design. Approximately half of all abstracts did not provide estimates of sensitivity and specificity. Eighty-eight percent (141) reported 'positive' diagnostic test results whilst only 6% (9) reported 'non-informative' test results.

I was able to locate a full-text publication for 117 abstracts in MEDLINE or EMBASE. Study authors completed 15 questionnaires (35%) out of the 43 sent for unpublished studies. Authors' responses to the questionnaire indicate that four abstracts were subsequently published in full. Thus a total of 121/160 (76%) diagnostic stroke abstracts were published as full-text reports.

Of the remaining 11 completed questionnaires, six authors stated that an original manuscript was initially rejected and a new manuscript was in preparation, four authors had not yet prepared or had no intention to submit a full-text manuscript, and one author stated that several manuscripts were published but was unable to provide precise information on the full-text publication related to the abstract presentation. Twenty-eight authors of the 43 abstracts not published in full did not return the questionnaire.

Table 3.1. Study characteristics reported in diagnostic abstracts on stroke

Study characteristics	Total no. of abstracts (%)	No. published in full (%)
<i>Study Design</i>		
Prospective	42 (26)	28 (23)
Retrospective	14 (9)	10 (8)
Not given	104 (65)	83 (69)
<i>No. of authors (median 6)</i>		
Above median	82 (51)	65 (54)
Below median	78 (49)	56 (46)
<i>No. of centres</i>		
Multi-centre	13 (8)	10 (8)
Single centre	147 (92)	111 (92)
<i>Sample size (median 43)</i>		
Above median	81 (51)	58 (48)
Below median	78 (49)	62 (51)
Not given	1	1
<i>Blinding of test result</i>		
Blinded	39 (24)	34 (28)
No blinding reported	121 (76)	87 (72)
<i>Inter-observer agreement</i>		
Assessed	17 (11)	15 (12)
Not stated if assessed	143 (89)	106 (88)
<i>Type of test</i>		
Imaging tests	140 (87)	107 (88)
Clinical assessment	20 (12)	14 (12)
<i>Clinical utility of test results*</i>		
Accurate	141 (88)	107 (88)
Possibly useful	10 (6)	7 (6)
Non-informative	9 (6)	6 (6)
<i>Sensitivity (median 0.91)</i>		
Above median	49 (31)	38 (32)
Below median	46 (31)	34 (28)
Not given	65 (41)	49 (40)
<i>Specificity (median 0.91)</i>		
Above median	42 (26)	30 (25)
Below median	37 (23)	27 (22)
Not given	81 (51)	64 (53)
<i>Region of origin of corresponding author</i>		
Australia, Canada, NZ, UK, USA	71 (44)	54 (45)
Europe (except UK)	79 (49)	62 (51)
Asia and South America	10 (6)	5 (4)
<i>Congress presentation</i>		
International Stroke Conference (USA)	79 (49)	63 (52)
European Stroke Conference	81 (51)	58 (48)

* See 'Methods' section for definitions

3.4.1 Abstracts published in full

Amongst the 160 identified diagnostic abstracts, 121 (76%) were subsequently published in full. The characteristics of the abstracts published in full are shown in Table 3.1. Fifty two percent of the abstracts subsequently published in full were originally presented at the International Stroke Conference whilst 48% were presented at the European Stroke Conference. In 55% of the abstracts subsequently published in full, the corresponding author was affiliated with a non-English speaking European institution whilst 45% of published articles generated from an English speaking institution (i.e. UK, United States, Canada, New Zealand, or Australia). Within the non-English speaking European countries, the highest number of abstracts subsequently published in full came from Germany (42 out of 62), whilst amongst the English-speaking countries, the USA had more published articles (n= 33) than UK (n= 11), Canada (n= 6), Australia (n= 3) or New Zealand (n= 1). Abstracts were published as papers in 35 different refereed journals (see Table 3.2). The top three journals where abstracts were published were: Stroke (n = 39), Cerebrovascular Diseases (n = 11), and Journal of Neuroimaging (n = 11).

Amongst the abstracts on diagnostic tests published in full, 13% (16/121) were published before abstract presentation, 62% (75/121) were published between 0 and 24 months after presentation, 8% (10/121) between 24 and 36 months, and 17% (20/121) after this time. Median time to publication, excluding abstracts published before meetings presentation, was 16 months and the mean time to publication was 20 months. The cumulative publication rate is illustrated by the Kaplan-Meier survival curve in Figure 3.1.

Table 3.2. Journals in which full-text reports were published

Journal	Full-text papers	Journal	Full-text papers
<i>Stroke</i>	39	<i>Australian Prescriber</i>	1
<i>Cerebrovascular Diseases</i>	11	<i>Brain</i>	1
<i>Journal of Neuroimaging</i>	11	<i>Canadian Journal of Neurological Sciences</i>	1
<i>Journal of Neurology, Neurosurgery, and Psychiatry</i>	8	<i>European Journal of Neurology</i>	1
<i>American Journal of Neuroradiology</i>	8	<i>European Radiology</i>	1
<i>Journal of Neuroradiology</i>	3	<i>Journal of Cerebral Blood Flow & Metabolism</i>	1
<i>Neurology</i>	3	<i>Journal of Computed Assisted Tomography</i>	1
<i>Archives of Neurology</i>	3	<i>Journal of Magnetic Resonance Imaging</i>	1
<i>Radiology</i>	3	<i>Journal of Neurological Sciences</i>	1
<i>Journal of Neurosurgery</i>	3	<i>Journal of Stroke & Cerebrovascular Diseases</i>	1
<i>Acta Neurologica Scandinavica</i>	2	<i>Journal of Ultrasound in Medicine</i>	1
<i>European Neurology</i>	2	<i>JAMA</i>	1
<i>Radiologe</i>	2	<i>Klinische Neuroradiologie</i>	1
<i>Ultrasound in Medicine and Biology</i>	2	<i>Neurosurgery</i>	1
<i>Acta Clinica Croatica</i>	1	<i>Physiotherapy Research International</i>	1
<i>Acta Radiologica</i>	1	<i>Prehospital Emergency Care</i>	1
<i>Age and Ageing</i>	1	<i>Rinsho Shinkeigoku</i>	1
<i>American Journal of Cardiology</i>	1	Published papers	121

3.4.2 Study characteristics and publication

The results of the univariate Cox regression analyses for the association between study characteristics and publication status are shown in Table 3.3. The hazard ratio for publication for studies in which inter-observer agreement was assessed compared to studies in which no inter-observer agreement was reported was 2.10 (95% CI 1.21 to 3.64) ($P = 0.02$). No other potential predictor was statistically significantly associated with full-text publication. In particular, the clinical utility of results, and Youden's Index had no statistically significant influence on publication. There was a non-significant trend for smaller studies to be more likely to be published than larger ones ($p = 0.09$).

Figure 3.1. Analysis of time to full publication of 121 diagnostic studies published as abstracts. Kaplan-Meier survival analysis. The line does not cross the vertical axis at 0%, as 16 papers were published in full prior to publication of the abstract.

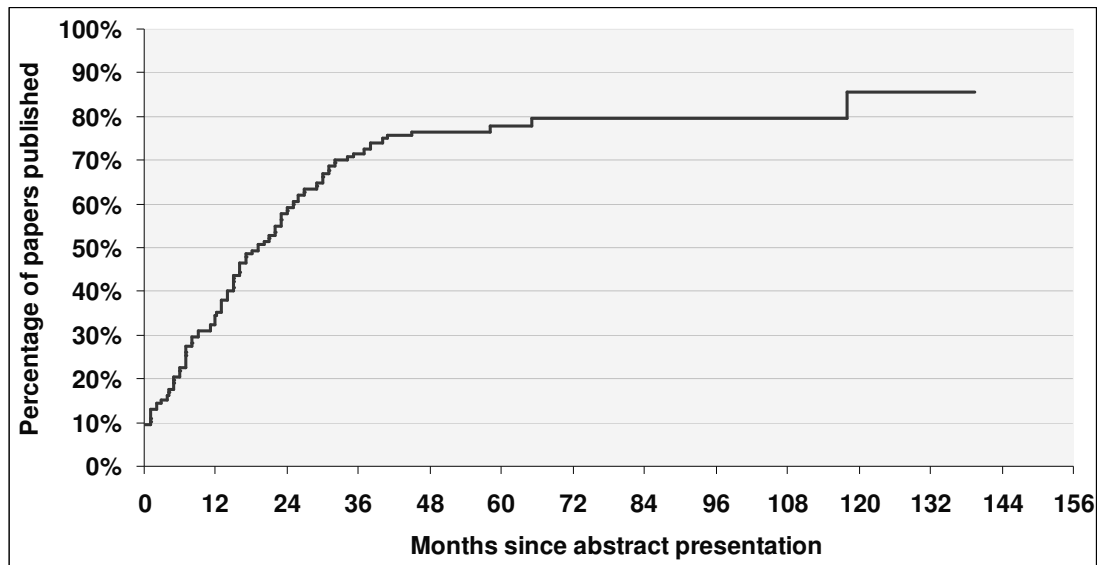


Table 3.3. Predictive factors of full-text publication using univariate Cox regression analyses

Variables	p-value (change in Log likelihood)	Hazard ratio (95% CI)
<i>Study design</i>	0.19	
Not stated		1.52 (0.95 to 2.43)
Retrospective		1.18 (0.53 to 2.64)
Prospective		comparator
<i>No. of authors</i>	0.48	1.03 (0.95 to 1.13)
<i>No. of centres</i>	0.84	
Multi-centre		1.07 (0.56 to 2.06)
Single centre		comparator
<i>Sample size (per 100 patients)</i>	0.09	0.90 (0.77 to 1.05)
<i>Blinding of test results</i>	0.15	
Blinded		1.38 (0.90 to 2.10)
No blinding reported		comparator
<i>Inter-observer agreement</i>	0.02	
Assessed		2.10 (1.21 to 3.64)
Not stated if assessed		comparator
<i>Type of test</i>	0.32	
Imaging tests		1.34 (0.74 to 2.40)
Clinical assessment		comparator
<i>Clinical utility of test results</i>	0.69	
Possibly useful		1.08 (0.50 to 2.33)
Non-informative		0.70 (0.28 to 1.72)
Accurate		comparator
<i>Youden's Index</i>	0.91	
Youden's Index score		0.71 (0.15 to 3.33)
Youden's Index missing		1.03 (0.69 to 1.53)
<i>Region of origin of corresponding author</i>	0.21	
Europe (except UK)		1.14 (0.76 to 1.68)
Asia and South America		0.50 (0.18 to 1.40)
Australia, Canada, NZ, UK, USA		comparator
<i>Congress presentation</i>	0.13	
International Stroke Conference (USA)		1.35 (0.92 to 1.99)
European Stroke Conference		comparator

* See 'Methods' section for definitions

3.5 Discussion

In systematic reviews it is crucial to identify all relevant published studies in the literature and minimize possible publication and related biases. There is significant empirical evidence of publication biases for randomised clinical trials. However, publication and related biases for studies of diagnostic accuracy have not been investigated to any great extent, including in the field of stroke, where diagnostic testing plays a major role in clinical practice. This study examined the frequency and determinants of publication of studies of diagnostic accuracy based on a large sample of abstracts presented at two major scientific stroke meetings. Abstracts varied in terms of study characteristics and methodological quality. In particular, 76% of all abstracts did not report on blinding, 65% on type of study design, and about half on measures of diagnostic accuracy.

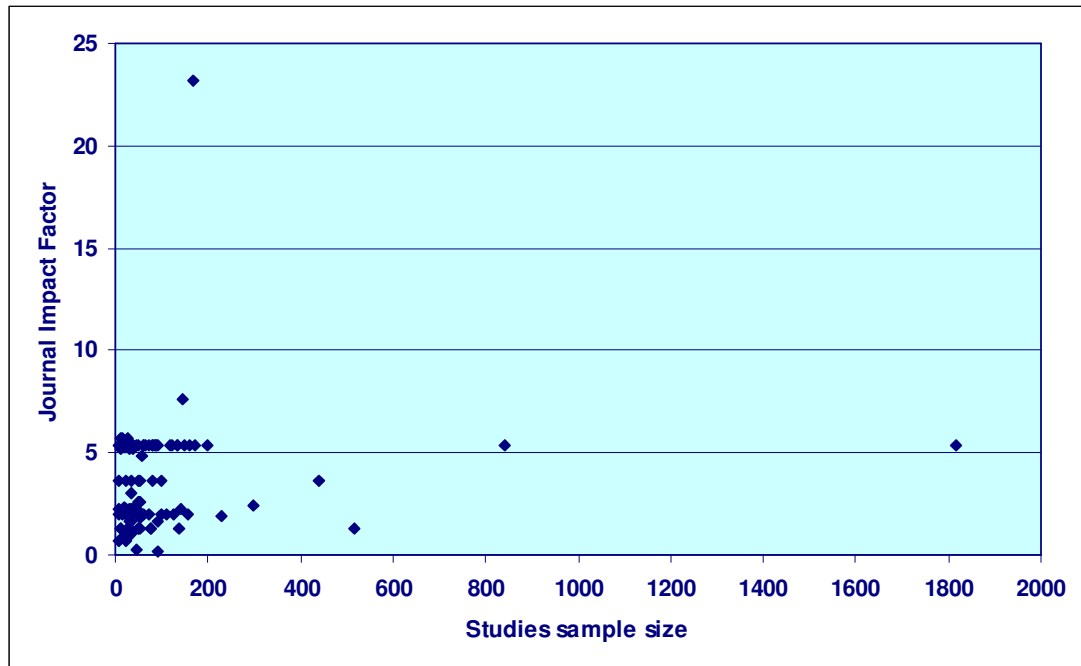
The findings of this study indicate that approximately 60% of diagnostic abstracts presented at international stroke meetings are published in refereed journal within two years from presentation and 70% were fully published within three years. The mean time of approximately 1.5 years to publication reflects most likely the peer review process that scientific data must undergo in order to be published. The publication rate of abstracts in our study is consistent with that reported by a recent study which assessed the frequency of non-publication of abstracts presented at the 2000 Annual Stroke Conference of the American Stroke Association. The study's authors included a broad class of stroke studies and did not focus specifically on diagnostic studies of test accuracy. They found, however, that overall 62.3% of all 2000 abstracts resulted in full-text publications, with a median time to publication of 15 months. No positive-outcome bias was detected across stroke studies.¹⁰

In this study the most important factor to be associated to full publication was the reporting of inter-observer agreement between test readers. This is likely to be explained by the overall higher methodological quality of investigations reporting information on inter-observer agreement. Interestingly, I found no evidence that full publication of diagnostic abstracts was influenced by the clinical utility of the findings. In particular, higher values of sensitivity and specificity results were not significantly associated with full publication. However, it is worth noticing that the median value of both sensitivity and specificity for all published studies was 0.91. Whilst this result could simply mean that the diagnostic tests applied in stroke medicine really are highly sensitive and specific, other explanations are possible. Firstly, authors might only submit a diagnostic abstract for presentation if the study showed high ‘accuracy’ or ‘positive results’. Indeed the majority of abstracts (88%) had ‘positive’ diagnostic test results and only 6% had ‘non-informative’ diagnostic test results. I suspect, but of course cannot confirm, that authors of ‘non-informative’ studies of diagnostic accuracy would be less likely to submit their work as an abstract than authors of ‘non-informative’ randomised controlled trials. Alternatively, this may reflect a bias in the selection process of abstracts for presentation to international stroke meetings.

Youden’s Index did not seem to predict publication in our study. However, in the analyses I had aggregate data across all types of test. This is not ideal, since Youden’s Index is more reliable if stratified by type of test.

I was surprised to note the tendency for smaller studies to be published in full. This is a marked contrast with the direction of publication bias in reports of clinical trials, where smaller studies are less likely to be published in full. However, as the difference did not reach statistical significance, this finding may purely be due to chance. On the other hand, it has been suggested that, in diagnostic research, smaller studies may be coupled with better methodology.⁸ Jon Deeks maintains for example, that “large retrospective studies in which investigators obtain test results from clinical databases may be more biased than smaller prospective studies in which clinicians carefully recruit patients presenting with a specific clinical problem”. In diagnostic research large studies are usually retrospective as large prospective studies (for example on new imaging tests) are expensive, difficult to perform, and currently not seen as priority. It is also likely that participants in larger diagnostic studies, rather than in smaller studies, are verified by different reference standards (i.e. verification bias) due to inadequate resources or practical problems.⁸ To examine further whether smaller studies might be of higher quality, I used the impact factor of the journal in which the full report was published as a surrogate marker of study quality. I have not found any clear relation between sample size of published studies and journal impact factor (Figure 3.2). Further evidence from future research is needed to clarify whether smaller studies of diagnostic accuracy are more likely to be published in full. In particular, it would be worth investigating the presence of publication bias in small and large studies after adjusting for the quality of individual studies. I did not find any significant association between either the region of origin of the corresponding author or the numbers of authors and full publication status on the univariate regression analyses.

Figure 3.2. Scatter plot of sample size versus journal impact factor for the 121 abstracts subsequently published in full.



Possible limitations of this study include the low response rate in the authors' survey (32%). There are several reasons for failing to publish a full-text manuscript following a conference presentation. Undoubtedly the reviewing process for a full publication is much more challenging than vetting for a conference presentation. Moreover, it should be noted that training programmes underwrite the costs of attending a conference only if one submits an abstract;¹¹ indeed in several institutions this is true independently of the status of the author. Hence, authors may put some effort in submitting their abstract but not pursue it afterwards. In fact, 4 out of the 11 of authors who replied to our questionnaire admitted that they had no intention of preparing a full-text publication. Alternatively, the 28 authors who did not reply to our questionnaire could have been the same authors who saw their papers rejected due to a publication bias. However, most (26/28) of these authors reported 'accurate' diagnostic test results in their abstract. I was also not able to assess the extent of a

possible bias in the selection of abstracts for presentation. The presence of a positive-outcome bias has been previously documented in both the submission and the selection processes of studies presented at scientific meetings.¹² Future creation of a prospective register for diagnostic studies (as are being created for randomised clinical trials) would allow evaluation of publication bias at all stages, but may fail to include diagnostic studies which are undertaken using retrospective designs.

To assess the methodological quality of abstracts was challenging due to the limited information given and the absence of a standard reporting structure. Essential information was missing in many abstracts, possibly due to the space restrictions. In particular, several abstracts failed to include methodological factors that become crucial in the quality assessment of diagnostic studies as defined by the STARD criteria such as inclusion of consecutive patients, a clear description of both the experimental tests and the reference standard, together with information on type of study design, patient sample, blinding of tests readers, and measures of test performance.¹³ Therefore, in assessing the quality of abstracts I may have run the risk of judging the quality of what was *reported* rather than the quality of what was actually *done*. There is no evidence suggesting that high quality work was undermined by poor abstracts. Indeed, there may be an association between poor quality work and overestimation of test accuracy, which could lead to an association between poor reporting and overestimates of test performance.

In conclusion, these results provide some reassurance that for abstracts submitted to major stroke conferences, there is no evidence of substantial bias in the publication

once an abstract has been accepted for publication. It seems likely that if publication bias is present, it is more likely to occur at the level of abstract submission or selection.

3.5.1 Implications for future research

- Further research is needed to provide additional empirical evidence on the size and direction of publication bias in diagnostic research. In particular, further studies are needed to confirm or refute the findings of this study in other clinical areas. Further studies should also investigate whether publication bias is more likely to occur at the level of abstract submission or selection.
- Further empirical evidence is needed to inform the development of search strategies to identify studies to be included in systematic reviews of diagnostic test accuracy.
- Further research is needed to develop graphical and statistical methods for detecting publication bias in systematic reviews of diagnostic test accuracy.
- Further research should also develop methods for reducing publication bias such as the registration of all diagnostic studies in publicly accessible databases, classified according on whether they gather data prospectively or retrospectively.

This study has been published in the Journal of Clinical Epidemiology:
Brazzelli M, Lewis SC, Deeks JJ, Sandercock PAG. No evidence of bias in the process of publication of diagnostic accuracy studies in stroke submitted as abstracts. *Journal of Clinical Epidemiology* 2009; 62:425-30

References

1. Dickersin K, Olson CM, Rennie D, Cook D, Flanagan A, Zhu Q, Reiling J, Pace B. Association between time interval to publication and statistical significance. *JAMA* 2002; 287: 2829-31
2. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991; 337: 867-72
3. Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990; 263: 1385-9
4. Dickersin K, Min YI. NIH clinical trials and publication bias. *Online Journal of Clinical Trials* 1993; 50: 4967
5. Liebeskind DS, Kidwell CS, Sayre JW, Saver JL. Evidence of publication bias in reporting acute stroke clinical trials. *Neurology* 2006; 67: 973-9
6. Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technology Assessment* 2000; 4 (10)
7. Scher RW, Dickersin K, Langenberg P. Full publication of results initially presented in abstracts: a meta-analysis. *JAMA* 1994; 272: 158-62

8. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of Clinical Epidemiology* 2005; 58: 882-93
9. Mohr JP. Report of the 15th European Stroke Conference, Brussels, May 16-19, 2006. *Cerebrovascular Diseases* 2006; 22: 459-60
10. Sanossian N, Ohanian AG, Saver JL, Kim LI, Ovbiagele B. Frequency and determinants of nonpublication of research in the stroke literature. *Stroke* 2006; 37: 2588-92
11. Marx WF, Cloft HJ, Do HM, Kallmes DF. The fate of neuroradiologic abstracts presented at national meetings in 1993: rate of subsequent publication in peer-reviewed, indexed journals. *American Journal of Neuroradiology* 1999; 20:1173-7
12. Callaham ML, Wears RL, Weber EJ, Barton C, Young G. Positive-outcome bias and other limitations in the outcome of research abstracts submitted to a scientific meeting. *JAMA* 1988; 280: 254-7
13. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HCW. STAndards for Reporting of Diagnostic accuracy (STARD) steering group. Towards complete and

accurate reporting of studies of diagnostic accuracy: the STARD initiative.

BMJ 2003; 326: 41-4

4. Magnetic resonance imaging versus computed tomography for detection of acute vascular lesions in patients presenting with stroke symptoms

4.1 Background

4.1.1 Target condition being diagnosed

Stroke is the third leading cause of death in Western societies and the leading cause of long-lasting severe disability.¹ The overall incidence of stroke is about 2.4 per 1000 population, with modest geographical variations.²

The two main categories of stroke are ischaemic and haemorrhagic. The latter occurs when a blood vessel in the brain ruptures causing bleeding either within the brain (intracerebral haemorrhage) or between the brain and the thin membrane that surrounds the brain (subarachnoid haemorrhage). Haemorrhagic strokes account for about 20% of all strokes. An ischaemic stroke occurs when an artery in the brain becomes blocked and blood flow suddenly decreases or stops causing a brain infarction. It is the most common form of stroke and accounts for about 80% of all strokes.

According to the location of the vascular event, ischaemic strokes may be classified as: (1) lacunar syndromes (LACS); (2) posterior circulation syndromes (POCS); (3)

total anterior circulation syndromes (TACS); and (4) partial anterior circulation syndromes (PACS).³

A transient ischaemic attack (TIA) starts suddenly, like a stroke, but symptoms last only for a short period of time (usually minutes or hours) and then resolve without leaving any noticeable sign or deficit. Symptoms of a TIA disappear completely within 24 hours from onset but they are associated with a high risk of a subsequent stroke, especially within the first few weeks (8% to 11.5% within the first seven days).⁴

Stroke is usually diagnosed by a combination of clinical examination and imaging procedures, though it is accepted that some patients with a clinically definite stroke may have normal brain imaging appearances.

An accurate and timely diagnosis is crucial in acute stroke both for decision-making and for establishing appropriate patient management. The continuing advances in neuroimaging techniques and the advent of thrombolytic therapy and other emergency neurointerventional procedures for ischaemic stroke (for example mechanical clot retrieval and intra-arterial thrombolysis), whose efficacy is greatest when given within a few hours of stroke onset,^{5,6} have increased the need for a rapid and reliable diagnosis. The main objectives of neuroimaging in acute stroke are to distinguish between stroke and non-stroke lesions (for example brain tumour, abscess), to distinguish haemorrhagic from ischaemic stroke, and to identify the

anatomic localisation of the vascular lesion. In particular, positive signs of acute cerebral ischaemic lesions on imaging may contribute to diagnosis within the first few hours. For example, if the clinical symptoms and radiological localisation of stroke match, or if the scan gives a clue to aetiology (such as the hyperdense artery sign suggesting large artery occlusion), this may give clinicians greater confidence to administer thrombolysis. However, it is not known whether the appearance of the acute ischaemic lesion influences stroke treatment, management, or outcome.

4.1.2 Index test(s)

Non-contrast computed tomography (CT) is a cost-effective and widely used neuroimaging method for the initial evaluation of patients presenting with stroke symptoms.⁷ In particular in the acute phase of stroke, when patients are scanned within a few hours of symptom onset, CT is quick to perform, easy to tolerate, and is known to be very reliable for the detection of intracerebral haemorrhage. Early detection of haemorrhage is essential since the presence of blood in the brain or subarachnoid space is the main contraindication both for the administration of antiplatelet, anticoagulants, or thrombolytic therapy.⁸ In contrast, the initial signs of cerebral infarction on CT can be subtle and consequently difficult to detect. It is now widely recognised that patients with an acute ischaemic stroke (especially lacunar stroke and stroke in the brainstem) can have normal CT appearances.

Magnetic resonance imaging (MRI) with diffusion-weighted sequences (DWI) has been increasingly used in the assessment of patients with stroke and TIA because of

its sensitivity in detecting the early changes associated with ischaemia,⁹ especially in patients with mild events or with small infarcts (for example lacunar or brainstem infarcts). On the other hand, the detection of acute cerebral haemorrhage on MRI is not as straightforward as on CT. There have been suggestions that MR gradient-echo sequences (GRE) can be as sensitive as CT for excluding intracerebral haemorrhage before the administration of thrombolysis.¹⁰⁻¹³ However, the utility of routine MRI to replace CT as the primary imaging for patients with suspected acute stroke has yet to be fully demonstrated. Moreover, in many countries, MRI is not available for stroke patients in many hospitals, it is contraindicated in patients with pacemakers and metal implants, and severely ill or confused patients may not tolerate MR scanning.

We therefore conducted a systematic review of the literature to estimate the accuracy of DWI compared with CT for the diagnosis of acute ischaemic stroke and to assess the accuracy of MRI (all feasible sequences, for example diffusion-weighted, gradient-echo sequences) for the early detection of haemorrhagic stroke.

We restricted the scope of this review to assess the diagnostic accuracy of both CT and MRI in patients suspected of acute stroke. We did not address the issue of the use of imaging methods (such as perfusion-diffusion mismatch) to identify acute stroke patients who might benefit from thrombolytic therapy outside the conventional therapeutic time window. Nor did we address the consequences of identifying an acute ischaemic lesion on imaging in terms of management or outcome.

4.2 Objectives

4.2.1 Primary objectives

- To compare MR diffusion-weighted images (DWI) and CT scans with respect to the accuracy of the localisation of acute ischaemic lesions. In particular, the review aimed to assess whether DWI could be considered superior to CT for the detection of acute ischaemic lesions within 12 hours (replacement of CT with DWI) or as an additional investigation for patients with negative or inconclusive CT scans.
- To assess the accuracy of MRI for the detection of acute haemorrhagic lesions within 12 hours.

The 12 hours time window was selected because this is the time when antithrombotic or thrombolytic therapy is most likely to be beneficial for patients with ischaemic stroke (and hence the reliable exclusion of haemorrhage is a clinically urgent priority). The choice of this time window is, however, necessarily arbitrary. At the time this review was planned, intravenous thrombolytic therapy was approved for use within 3 hours, but studies seeking to extend this time window to 9 hours (and possibly further) were already underway. Other acute stroke studies were recruiting to later time windows - 24 or 48 hours (e.g. the ENOS trial <http://www.enos.ac.uk/>). We opted for a time window of 12 hours because sufficient published acute stroke studies had an inclusion window of 12 hours. An unduly restrictive time window would have further limited the number of studies eligible for inclusion. Although thrombolytic therapy is one of the most important therapies, others such as intra-

arterial thrombolysis, novel thrombolytic agents, hypothermia, and neuroprotective drugs may have a therapeutic window of 12 hours or longer and therefore the choice of a wider time window is likely to have a greater clinical relevance. This choice has been further supported by recent published evidence-based guidelines which have adopted the same 12 hour cut off for DWI studies.¹⁴ Besides, in clinical practice clinicians need to diagnose people even if they are not going to treat them, so including studies across a range of times is more relevant to determining diagnostic accuracy.

4.2.2 Secondary objectives

As imaging results may vary depending upon the technical characteristics of the imaging test and the time when the imaging test is performed, we planned to analyse diagnostic data according to:

- the time of imaging from onset of symptoms (e.g. patients scanned within three, six, and 12 hours of symptoms onset);
- choice of imaging test for detection of haemorrhagic stroke (e.g. diffusion-weighted sequences, gradient-echo sequences).

4.2.3 Investigation of sources of heterogeneity

We were not able to investigate methodological and clinical sources of heterogeneity due to the relatively limited number of studies included in this review.

4.3 Methods

4.3.1 Criteria for considering studies for this review

4.3.1.1 *Types of studies*

We aimed to include studies published in any language. However, we did not include non-English articles for which a full-text translation or evaluation could not be obtained. For the detection of ischaemic lesions, studies were eligible if:

- i) both DWI and CT were evaluated in the same patient population (direct comparison) against an acceptable reference standard (as defined later), or if patients were randomised within a study to DWI or CT;
- ii) clinical and imaging assessments were performed within 12 hours of onset of symptoms;
- iii) the absolute numbers of observations of true positives, false positives, false negatives, and true negatives were available or derivable from the data reported in the primary studies.

For the detection of haemorrhagic lesions, studies were eligible if:

- i) MRI sequences were evaluated against a clinical diagnosis of stroke supported by CT findings (reference standard) in cross-sectional studies;
- ii) clinical and imaging assessments were performed within 12 hours of symptoms onset;

iii) the absolute numbers of observations of true positives, false positives, false negatives, and true negatives were available or derivable from the data reported in the primary studies.

We included both prospective and retrospective studies.

We excluded studies that focused on patients presenting exclusively with a clinical syndrome suggesting either subarachnoid haemorrhage or isolated intraventricular haemorrhage since they are very distinctive clinical syndromes not directly relevant to patients presenting with the focal neurological deficits of acute stroke.

We also excluded studies that: addressed specific anatomical, metabolic, microvascular, or volumetric aspects of stroke; focused on specific technical aspects of CT and MRI; analysed perfusion versus diffusion imaging differences in patients with acute cerebral ischaemia.

Where investigators published several reports based on data from a single study population, we selected the updated or most complete report.

4.3.1.2 Participants

Adult patients with clinical symptoms suggestive of acute stroke, including patients in whom the subsequent diagnosis proved to be TIA.

4.3.1.3 Index tests

- DWI and CT performed within 12 hours of onset of symptoms for the detection of ischaemic brain lesions (CT is regarded here as the alternative test for detection of ischaemic lesions).
- MRI (all suitable sequences) performed within 12 hours of onset of symptoms for the detection of haemorrhagic brain lesions.

4.3.1.4 Target conditions

- Acute ischaemic stroke
- Acute haemorrhagic stroke

4.3.1.5 Reference standard

A single 'gold standard' for the diagnosis of stroke does not exist. In clinical practice however, expert assessment based on the combination of clinical features, imaging appearances, laboratory tests, and clinical follow up does provide the most comprehensive diagnosis. A diagnosis of stroke based only on the clinical and imaging data available to the clinician within the first few hours is unlikely to be sufficiently accurate.

For the diagnosis of acute ischaemic stroke we considered an acceptable reference standard to be: a combination of clinical and imaging information supported by clinical or imaging follow up (CT or MRI) or autopsy. Any elaboration of this definition was however deemed suitable for inclusion (e.g. studies that relied exclusively on a clinical diagnosis or exclusively on a CT or MRI follow up).

For the diagnosis of acute haemorrhagic stroke we considered a valid reference standard to be: a clinical diagnosis supported by CT or autopsy.

Note: in some studies, patients whose symptoms lasted less than 24 hours, but who had evidence of an ischaemic lesion on imaging, were counted as having had strokes and hence analysed as true positive cases. In other studies, however, patients with symptom duration less than 24 hours and an ischaemic lesion on imaging were analysed as being false positive cases.

4.3.2 Search methods for identification of studies

4.3.2.1 Electronic searches

We identified eligible studies by searching the following electronic databases:

- Cochrane Stroke Group Trials Register (last searched by the Trials Register Administrator in March 2009);
- Cochrane Central Register of Controlled Trials (CENTRAL) (*The Cochrane Library* Issue 1, 2009);
- MEDLINE - Ovid (from January 1995 to March 2009);
- EMBASE - Ovid (from January 1995 to March 2009);
- MEDION (last searched in March 2009 using the 'Systematic Reviews and Diagnostic Studies' search filter, the ICPC code = 'Neurological' and the signssymp = 'Medical Imaging').

4.3.2.2 MEDLINE and EMBASE searches

We searched indexed records which appeared in MEDLINE (January 1995 to March 2009). The choice of this time period was justified by the introduction in the early 1990s of MR diffusion-weighted and gradient-echo sequences into clinical practice. The MEDLINE search strategy included both subject headings (MeSH terms) and text words for the target condition (stroke) and the imaging techniques under investigation (MRI and CT). We also included a methodological filter for studies of diagnostic accuracy. Our methodological filter was based on the diagnostic component of the search strategy developed and validated by Astin and colleagues to identify diagnostic accuracy studies on imaging.¹⁵ We did not apply any language restrictions. We adapted the MEDLINE search to search EMBASE. In particular, we 'translated' the MEDLINE MeSH terms into the corresponding terms available in the Emtree vocabulary. Full details of both the MEDLINE and EMBASE search strategies together with a brief summary of the MEDLINE search strategy are presented in Appendix 2. We imported all citations identified by the MEDLINE and EMBASE search strategies into the Reference Manager bibliographic database.¹⁶

4.3.2.3 Searching other resources

We handsearched all proceedings of the International Stroke Conference and the European Stroke Conference (1995 to 2004). These proceedings were published as abstracts in special issues of two peer-reviewed journals: *Stroke* and *Cerebrovascular Diseases*. We also searched the following websites using terms for the target condition (stroke, acute stroke) and for the two imaging techniques under investigation (magnetic resonance imaging, computer tomography):

- American College of Cardiology - ACC (www.acc.org) (last searched March 2009);
- American College of Radiology - ACR (www.acr.uk) (last searched March 2009);
- American Heart Association - AHA (www.americanheart.org) (last searched March 2009);
- American Stroke Association - ASA (www.strokeassociation.org) (last searched March 2009);
- Department of Radiology & Diagnostic Imaging - University of Alberta, Canada (www.radiology.med.ualberta.ca) (last searched March 2009);
- National Institute of Neurological Diseases and Stroke - NINDS (www.ninds.nih.gov) (last searched March 2009);
- National Stroke Association - NSA (www.stroke.org) (last searched March 2009);
- Royal College of Physicians (www.rcplondon.ac.uk) (last searched March 2009);
- Royal College of Radiologists - RCR (www.rcr.ac.uk) (last searched March 2009);
- Scottish Intercollegiate Guidelines Network - SIGN (www.sign.ac.uk) (last searched March 2009).

We perused the reference lists of all relevant articles to identify further published studies for possible inclusion in the review. We also contacted experts in the field to enquire about ongoing or completed but yet not published diagnostic studies.

4.3.3 Data collection and analysis

4.3.3.1 Selection of studies

One author (MB) initially screened the titles and abstracts of the search results and retrieved all potentially relevant reports in full. Three review authors (MB, MGC, ER) independently reviewed all relevant reports according to the pre-defined inclusion criteria. We resolved any disagreements by consensus or arbitration. The same three authors extracted data from the selected reports.

4.3.3.2 Data extraction and management

We designed a data abstraction form specifically to collect details from selected studies (see Appendix 3). We recorded the following information for each individual study (without concealing the study authorship or other publication details): journal name, year of publication, study design and method of recruitment (systematic review, randomised controlled trial, cross-sectional survey; prospective study, retrospective study), setting, number and characteristics of participants (age, sex, ethnicity, previous history of stroke, concomitant diseases), classification of stroke, definition of abnormal CT and MR images, time of imaging, the reference standard by which the final diagnosis was established, time interval from index test and comparator tests, time interval from index test(s) and established diagnosis of stroke, technical characteristics of MRI and CT, information related to the clinicians who read and interpreted imaging results (background speciality, level of expertise) and to the clinicians who established a clinical diagnosis of stroke. We resolved any disagreements by consensus or arbitration.

4.3.3.3 Assessment of methodological quality

Four authors (MB, MGC, ER, NA) independently assessed the methodological quality of each included study using the QUality Assessment of Diagnostic Accuracy Studies (QUADAS) tool developed by the NHS Centre for Reviews and Dissemination at the University of York, UK.¹⁷ The QUADAS tool is structured in a series of questions which should be answered 'yes', 'no', or 'unclear', and aims to evaluate the presence of spectrum bias, bias associated with the choice of reference standard, disease progression bias, verification bias, review bias, clinical review bias, incorporation bias, and bias associated with study withdrawals and indeterminate results. In particular, we considered a representative spectrum of patients to be: female and male patients of all ages presenting with mild, moderate, or severe stroke symptoms; with or without previous history of stroke; scanned within a few hours of onset of symptoms. We defined an appropriate reference standard likely to correctly classify the target condition as: an expert clinical assessment coupled with clinical and imaging follow up. We also considered up to seven days an acceptable time period between MRI and CT for the detection of haemorrhagic stroke. For ischaemic stroke we accepted any time period reported by the studies' investigators between CT and MRI, and the follow up reference test. We also decided to add to the recommended QUADAS questions the following items pertinent to the purpose of this systematic review: expertise of the person interpreting the imaging results; whether the scans were read blind to clinical information; and whether the sequence of imaging tests was determined at random (the modified QUADAS checklist is reported in Appendix 4). As regards the expertise of the person interpreting imaging results, we distinguished radiologists and neuroradiologists, who by definition are

experts in reading imaging test results, from neurologists, geriatricians, and general medicine doctors. We resolved any disagreements by consensus or arbitration. For each individual study we tabulated the agreed results of the quality assessment.

4.3.3.4 Statistical analysis and data synthesis

We extracted or derived indices of diagnostic performance from data presented in each primary study for each imaging test. We constructed 2 X 2 contingency tables of true positive cases, false positive cases, false negative cases, and true negative cases. We considered patients with ischaemic stroke as false positives or true negatives when analysing the performance for detecting haemorrhagic stroke, and we counted patients with haemorrhagic stroke as false positives or true negatives when analysing the performance for detecting ischaemic stroke. We calculated sensitivity and specificity with 95% confidence intervals (CI) for each imaging test in each study. We tabulated results for studies on ischaemic stroke separately from those for studies on haemorrhagic stroke.

We drew forest plots to show the variation of sensitivity and specificity estimates together with their 95% CI. For studies on ischaemic stroke where DWI was compared with CT versus a reference standard of clinical diagnosis and imaging follow up we also plotted the imaging test results on a receiver operating characteristic (ROC) plot of true positive rate (sensitivity) against false positive rate (1 - specificity).

We explored the heterogeneity of the sensitivity and specificity estimates amongst studies on ischaemic stroke by examining both the forest plots and the ROC plot. As almost all estimates of specificity were at 'ceiling level' (specificity of 1) there was no evidence of heterogeneity and it was not possible to use statistical methods that rely on estimating correlations between sensitivity and specificity to enable estimation of a summary ROC curve.^{18,19} Rather, we separately pooled estimates of sensitivities and specificities across the studies. For CT sensitivity and DWI sensitivity and specificity we undertook meta-analyses using maximum likelihood estimation of a random-effects model to pool logit transformed proportions and allow for within-study binomial variation. We computed confidence intervals using MCMC sampling. We used a fixed-effect analysis to estimate the pooled specificity of CT as a specificity of 1 was observed in every study (score method was used to compute confidence intervals). We used the DiagMeta package within the R software (The R Foundation for Statistical Computing Version 2.7.1) to carry out the analyses. We were not able to perform a formal statistical comparison between tests due to the zero cell issues and small sample sizes. An informal comparison between tests was made by meta-analysing each test separately and examining the results. As all studies in the analysis evaluated both tests in all patients this comparison should not be biased by differences between the studies.

We did not include study-level covariates in the analyses to assess factors that might have contributed to heterogeneity (such as time of imaging) as in small meta-analyses this is likely to produce unreliable estimates.

We did not calculate overall estimates for studies on haemorrhagic stroke as we only identified two small studies of such different methodological quality that formal meta-analysis was not appropriate.

4.4 Results

4.4.1 Results of the search

The MEDLINE and EMBASE searches identified 9961 citations. Of these, we considered 112 relevant to the purpose of our review and we retrieved the full-text articles (Figure 4.1). We subsequently excluded 103 articles (see the Characteristics of excluded studies table in Appendix 3). The most common reason for exclusion was that the study was either not a primary diagnostic study of test accuracy or it did not involve appropriate test comparisons. Eight studies, published in nine reports,²⁰⁻²⁸ with a total of 306 participants fulfilled our inclusion criteria. Six studies focused on the comparison between DWI and CT for the detection of ischaemic lesions,²⁰⁻²⁵ one study estimated the accuracy of MRI for detection of haemorrhagic lesions,²⁶ and one study assessed the use of MRI compared with CT for detection of both ischaemic and haemorrhagic lesions.²⁷ Thus, seven studies contributed to the assessment of acute ischaemic stroke and two studies contributed to the assessment of haemorrhagic stroke. The details of all included studies are reported in the Table 4.1.

Figure 4.1. Flow of studies through the selection process

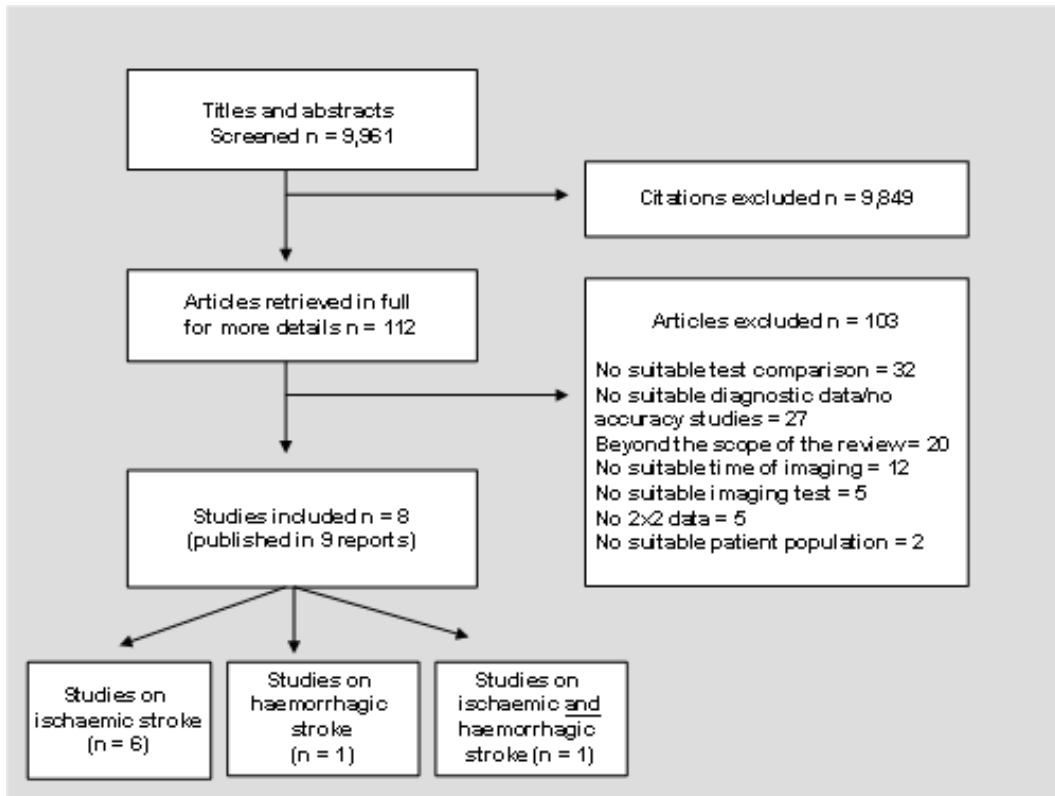


Table 4.1. Characteristics of included studies

Study ID	Clinical features and setting	Participants	Study Design	Target condition and reference standard	Index and comparator tests	Follow-up	Comments
Barber 1999	Patients with suspected acute ischaemic stroke in the middle cerebral artery territory who were studied with both DWI and CT within 6 hours of symptom onset	17 patients (53% men) presenting with stroke symptoms Mean age: 68.5 years Mean Canadian Neurological Scale score: 5.8 (range 1.5 to 11)	Prospective	Ischaemic stroke Clinical diagnosis and imaging follow up	DWI versus CT	T2-weighted imaging performed at 90 days	None of the patients were treated with thrombolysis One patient was unable to tolerate MRI and was not included
Bozzao 1999	Patients with suspected acute ischaemic stroke who underwent imaging within 12 hours of symptom onset	15 stroke patients (40% men) Mean age 67.6 years (range: 54 to 81 years)	Prospective	Acute ischaemic stroke Clinical diagnosis and imaging follow up	DWI versus CT	CT at 8 days	Haemorrhage excluded Severity of stroke not reported Stroke vascular territory not specified One patient could not undergo MRI because of agitation
Chalela 2007	Patients with suspected acute stroke who underwent both DWI and CT within 3 hours of symptom onset Patients selection was not restricted to MCA strokes	90 patients presenting with stroke symptoms Median age 76 years (range 21 to 100 years) Median score at NIHSS = 3 (range 0 to 37)	Prospective	Acute stroke Final diagnosis based on all available evidence including acute and follow-up imaging	DWI versus CT for detection of ischaemic stroke MRI sequences for detection of haemorrhagic stroke	Imaging	None of the patients were treated with thrombolysis TIAs with imaging evidence of infarction were counted as true positive cases The distribution of patients was skewed towards mild cases Patients who could not tolerate MRI or with uninterpretable imaging results were excluded

Table 4.1. Characteristics of included studies – continued

Study ID	Clinical features and setting	Participants	Study Design	Target condition and reference standard	Index and comparator tests	Follow-up	Comments
Gonzalez 1999	<p>Patients with suspected ischaemic stroke and with a negative or inconclusive CT scan and for whom MRI was deemed essential for establishing proper management</p> <p>Imaging was performed within 6 hours of symptom onset</p> <p>Most of the patients had a stroke in the MCA territory</p>	<p>22 patients (55% men) with acute stroke</p> <p>Mean age: 66.2 years</p>	<p>Retrospective</p> <p>Original scans were re-examined de novo by study investigators</p>	<p>Acute ischaemic stroke</p> <p>Clinical and imaging follow up</p>	<p>DWI versus CT</p>	<p>Clinical assessment and imaging</p>	<p>Haemorrhage excluded</p> <p>Severity of stroke not reported</p> <p>Three patients were excluded because they did not undergo CT</p>
Oppenheim 2005	<p>Patient details were retrospectively extracted from the acute stroke database of 2 university hospitals which used MRI as the first imaging modality for patients reaching hospital within 6 hours of symptoms onset</p> <p>Only patients with a stroke severity of ≥ 3 points on the NIHSS were deemed suitable for inclusion</p>	<p>86 patients (64%) with and without haemorrhagic stroke</p> <p>Mean age: 68.8 years</p>	<p>Retrospective study</p> <p>Original scans were re-examined de novo by study investigators</p>	<p>Acute haemorrhagic stroke</p> <p>Clinical and imaging follow up</p>	<p>DWI and GRE MR sequences</p>	<p>Clinical assessment and imaging</p>	<p>Only a minority of patients had a CT scan and the diagnosis of acute intracerebral haemorrhage was based on multisequence MRI (incorporation bias)</p>

Table 4.1. Characteristics of included studies – continued

Study ID	Clinical features and setting	Participants	Study Design	Target condition and reference standard	Index and comparator tests	Follow-up	Comments
Saur 2003	Patients with acute ischaemic stroke in the middle cerebral artery territory for whom DWI and CT were performed within 6 hours of stroke onset and with a time interval of less than 45 minutes	46 stroke patients (67% men) Mean age: 62.8 years (range: 35 to 89 years) Mean NIHSS score: 13.3 (range: 3 to 23)	Retrospective Original scans were re-examined de novo by study investigator	Ischaemic stroke (middle cerebral artery territory) Clinical diagnosis and imaging follow up	DWI versus CT	Clinical assessment and imaging	Haemorrhage excluded
Sorensen 1996	Patients with suspected stroke for whom imaging was performed within 12 hours of symptoms onset Patients with intracerebral haemorrhage were excluded	11 patients (73% men) with acute ischaemic stroke Mean age: 64 years (range 47 to 91 years)	Prospective	Ischaemic stroke Clinical diagnosis and imaging follow up	DWI versus CT	MR imaging	Haemorrhage excluded Severity of stroke not reported Stroke vascular territory not specified
Urbach 2000	Patients with acute ischaemic stroke in the middle cerebral artery territory for whom DWI and CT were performed within 6 hours of stroke onset	30 patients (60% men) with acute ischaemic stroke Mean age: 52 years (range 18 to 76 years)	Retrospective	Acute ischaemic stroke Clinical diagnosis and imaging follow up	DWI versus CT	Clinical assessment and imaging	Severity of stroke not reported

CT: computed tomography; DWI: diffusion-weighted magnetic resonance imaging; GRE: gradient-echo; MCA: middle cerebral artery; MR or MRI: magnetic resonance imaging; NIHSS: National Institute of Health Stroke Scale; TIA: transient ischaemic attack.

4.4.2 Methodological quality of included studies

4.4.2.1 Studies on ischaemic stroke

Seven studies compared CT with DWI in the same patients for the detection of acute cerebral ischaemia. The total number of assessed patients was 226. Sample size ranged from 11 to 90 patients (mean 32 patients). The reported mean age was 65.1 years (range 21 to 100 years). The proportion of men ranged from 40% to 73%, with no information on gender distribution in one study. Only three studies clearly reported stroke severity and only three reported the number of patients who were excluded because they could not tolerate MRI. CT and DWI were performed within three hours of symptoms onset in one study, within six hours in four studies, and within 12 hours in the two remaining studies. In all but one study CT was performed before DWI. In five studies the average delay between CT and DWI was 55.3 minutes (SD 24.4 minutes). One study reported that the median interval between the two imaging techniques was 34 minutes, and the remaining study did not provide this information but stated that the interval between DWI and stroke onset was 4.2 hours. In four studies selection of patients was restricted to middle cerebral artery stroke; in two studies the stroke vascular territory was not given even though it is likely that they predominantly enrolled patients with anterior circulation stroke; and in one study patients were not selected according to the type of stroke.

The quality of the seven included studies varied (Figure 4.2 and Figure 4.3). Four studies collected patients' data prospectively (132 patients in total) and three studies retrospectively (94 patients in total). In all three retrospective studies the original MR

and CT images of acute stroke, obtained from the patients' hospital records, were reviewed de novo by the study investigators. In these three studies, even though brain images were reviewed de novo, there was still a risk of bias due to the retrospective selection of patients' records. Four studies clearly described their inclusion criteria but only one study appeared to include a representative spectrum of stroke patients (a consecutive series of patients referred to hospital because of a clinical suspicion of stroke and irrespective of gender, age, previous medical history, co-morbidity, symptom severity, or final diagnosis) (90 patients).²⁷ However, the extremely mild strokes and the absence of any stroke mimics in this study indicated that some clinical exclusion criteria must have been applied after hospital admission and before study inclusion and scanning. Furthermore, in this study patients presenting with TIA but in whom DWI showed a new ischaemic lesion had their diagnosis changed to stroke (incorporation bias). In all the included studies the reference standard for diagnosis of stroke was a clinical diagnosis supported by imaging follow up. The reference standard was independent of the index text in six studies (the acute images were not used in the final diagnosis). Readers of DWI and CT acute images were reported to be blind to patients' clinical details and final diagnosis in only three studies. Information on blinding of the reference standard results was not clearly reported in five studies and in two studies interpretation of the follow up images was not blind to the findings of the acute images. Information on the expertise of clinicians reading imaging results was available in all but one study. None of the studies use formal randomisation methods to determine the sequence of the imaging tests.

Figure 4.2. Methodological quality of the seven included studies on ischaemic stroke

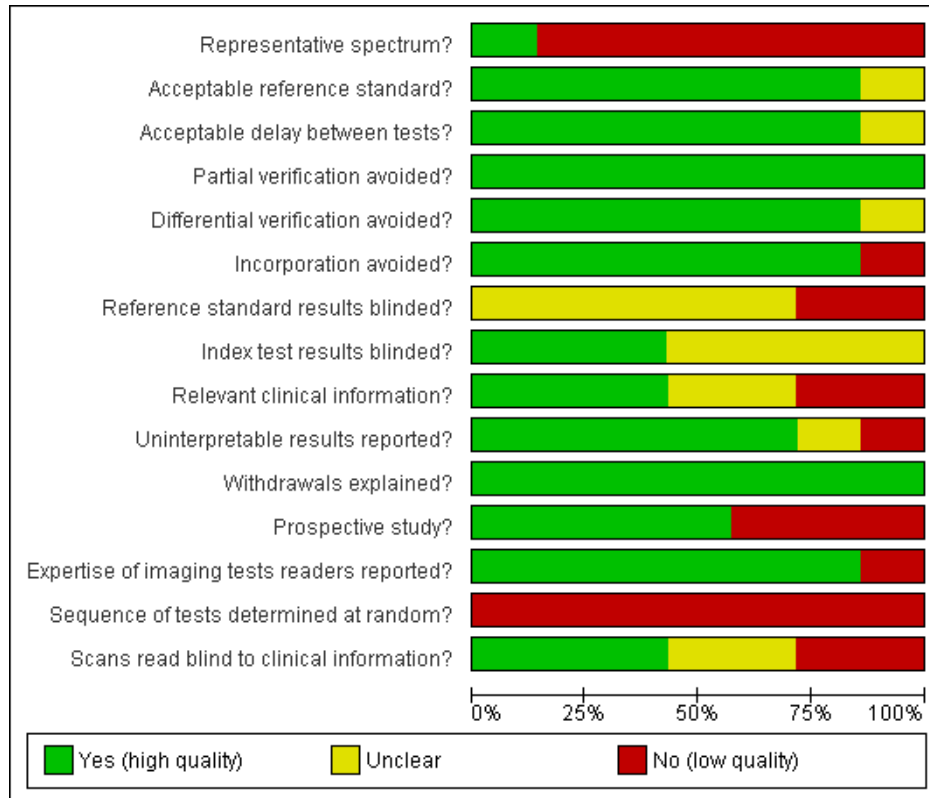


Figure 4.3. Methodological quality summary: review authors' judgment on each individual QUADAS item for the seven included studies on ischaemic stroke

	Representative spectrum?	Acceptable reference standard?	Acceptable delay between tests?	Partial verification avoided?	Differential verification avoided?	Incorporation avoided?	Reference standard results blinded?	Index test results blinded?	Relevant clinical information?	Uninterpretable results reported?	Withdrawals explained?	Prospective study?	Expertise of imaging tests readers reported?	Sequence of tests determined at random?	Scans read blind to clinical information?
Barber 1999	+	+	+	+	+	+	?	+	+	+	+	+	+	+	+
Bozzao 1999	+	?	+	+	+	+	?	?	?	?	+	+	+	+	?
Chalela 2007	+	+	?	+	?	+	+	+	+	+	+	+	+	+	+
Gonzalez 1999	+	+	+	+	+	+	?	?	+	+	+	+	+	+	+
Saur 2003	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Sorensen 1996	+	+	+	+	+	+	?	?	?	+	+	+	+	+	?
Urbach 2000	+	+	+	+	+	+	?	?	+	+	+	+	+	+	+

4.4.2.2 Studies on haemorrhagic stroke

The characteristics of the two studies assessing the use of acute MRI for detection of haemorrhagic lesions are summarised in Table 4.2.

The prospective study by Chalela and colleagues²⁷ compared non-contrast CT with MRI (diffusion-weighted and susceptibility-weighted images) in 450 patients referred for emergency assessment of suspected stroke, 90 of whom were scanned within three hours from the onset of symptoms. The patients' median severity score, assessed by the National Institute of Health Stroke Scale (NIHSS), was 3 (range 0 to 37) indicating the presence of predominantly mild stroke deficits. The proportion of patients with primary cerebral haemorrhage was 13% (12/90). Overall the proportion of patients who could not tolerate MRI, amongst a predominantly mild stroke population, was 11% (49/450).

The retrospective study by Oppenheim and colleagues²⁶ used data extracted from the acute databases of two university hospitals to evaluate the accuracy of five MR sequences (T₁, GRE, FLAIR, T₂-EPI, and DWI) to identify within 86 stroke patients those with (43 patients) and without (43 patients) intracerebral haemorrhage. Patients were included if they presented with a stroke severity score of ≥ 3 points on the NIHSS and if they underwent imaging within six hours of stroke onset. The patients' final diagnosis incorporated all clinical, pathological, and imaging investigations. However, as only a small number of patients

Table 4.2. Characteristics and diagnostic results of the two included studies on haemorrhagic stroke

Study	Participants (% men)	Participants assessed	Age (range)	Stroke severity	Time of imaging	MRI results (95% CI)
<u>Chalela 2007</u> *	450 (unknown)	90	Median 76 y (21 to 100 y)	Median score at NIHSS = 3 (range 0 to 37)	Within 3 hours of stroke onset	GRE and DWI sensitivity 0.83 (0.52 to 0.98) GRE and DWI specificity 1.00 (0.95 to 1.00)
<u>Oppenheim 2005</u> **	86 (64)	82	Mean 68.8 y	Mean score at NIHSS = 11.25	Within 6 hours of stroke onset (mean time 2.6 hours)	DWI sensitivity 1.00 (0.91 to 1.00) DWI specificity 1.00 (0.91 to 1.00) GRE sensitivity 1.00 (0.91 to 1.00) GRE specificity 0.98 (0.87 to 1.00)

*Note: *: prospective **: retrospective; 95% CI: 95% confidence intervals; DWI: diffusion-weighted imaging; GRE: gradient-echo; MRI: magnetic resonance imaging; NIHSS: National Institute of Health Stroke Scale*

underwent CT, the reference standard was highly inconsistent and the final diagnosis of intracerebral haemorrhage was primarily based on multisequence MRI (incorporation bias). Not all patients completed all five MR sequences and we assessed results from 82 patients who completed both the gradient-echo and diffusion-weighted sequences. Figure 4.4 summarises the results of the quality assessment of the two studies on haemorrhagic stroke. Only five of the 15 QUADAS items were met by both studies. Information on the spectrum of patients was not clearly reported in the Oppenheim study.²⁶ In both studies the reference standard was a final clinical diagnosis of haemorrhagic stroke supported by all available imaging investigations including the acute images. However, in both studies it was unclear whether all patients were verified by the same reference standard and there was clear evidence of incorporation bias as MRI findings contributed to the final diagnosis. In both studies the follow up images were not read blind to the findings of acute images. The reading order of MRI examinations was determined at random only in the Oppenheim study.²⁶

Figure 4.4. Methodological quality summary: review authors' judgement on each individual QUADAS item for the two included studies on haemorrhagic stroke

	Representative spectrum?	Acceptable reference standard?	Acceptable delay between tests?	Partial verification avoided?	Differential verification avoided?	Incorporation avoided?	Reference standard results blinded?	Index test results blinded?	Relevant clinical information?	Uninterpretable results reported?	Withdrawals explained?	Prospective study?	Expertise of imaging tests readers reported?	Sequence of tests determined at random?	Scans read blind to clinical information?
Chalela 2007	+	+	?	+	?	-	-	+	+	-	+	+	+	-	+
Oppenheim 2005	?	?	?	+	?	-	-	+	-	-	+	-	+	+	+

4.4.3 Findings

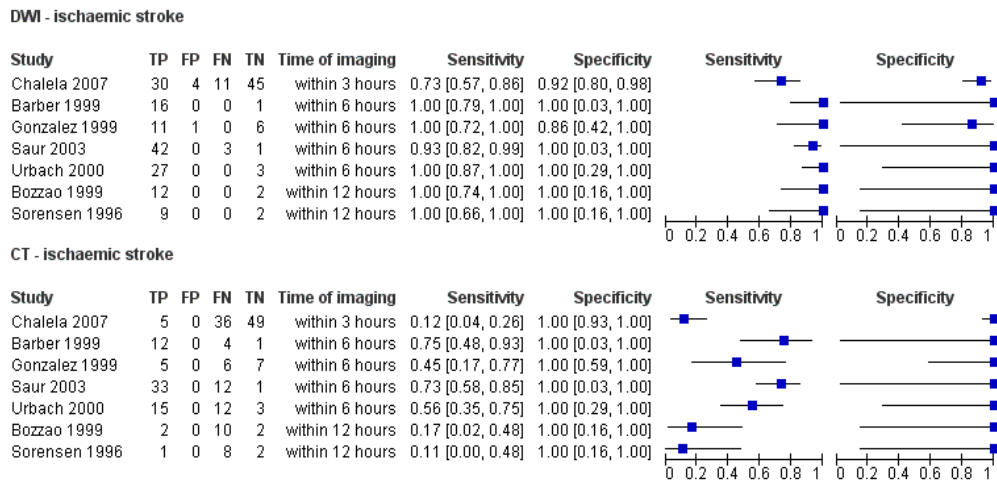
4.4.3.1 Studies on ischaemic stroke

Figure 4.5 shows the forest plots of the sensitivity and specificity estimates for DWI and CT for the seven studies that assessed patients with ischaemic stroke. Sensitivity estimates for DWI ranged from 0.73 to 1.00 (median 1.00) and the sensitivity estimates for CT ranged from 0.11 to 0.75 (median 0.45). Specificity estimates for DWI ranged from 0.86 to 1.00 (median 1.00) whilst specificity estimates for CT were all at 'ceiling level' (1.00 specificity). The pairs of observed values of sensitivity and specificity for DWI and CT are presented in a ROC space in Figure 6. The pooled estimates for DWI sensitivity and specificity were 0.99 (95% CI 0.23 to 1.00) and 0.92 (95% CI 0.83 to 0.97) respectively, whilst the pooled estimates for CT sensitivity and specificity were 0.39 (95% CI 0.16 to 0.69) and 1.00 (95% CI 0.94 to 1.00) respectively.

4.4.3.2 Studies on haemorrhagic stroke

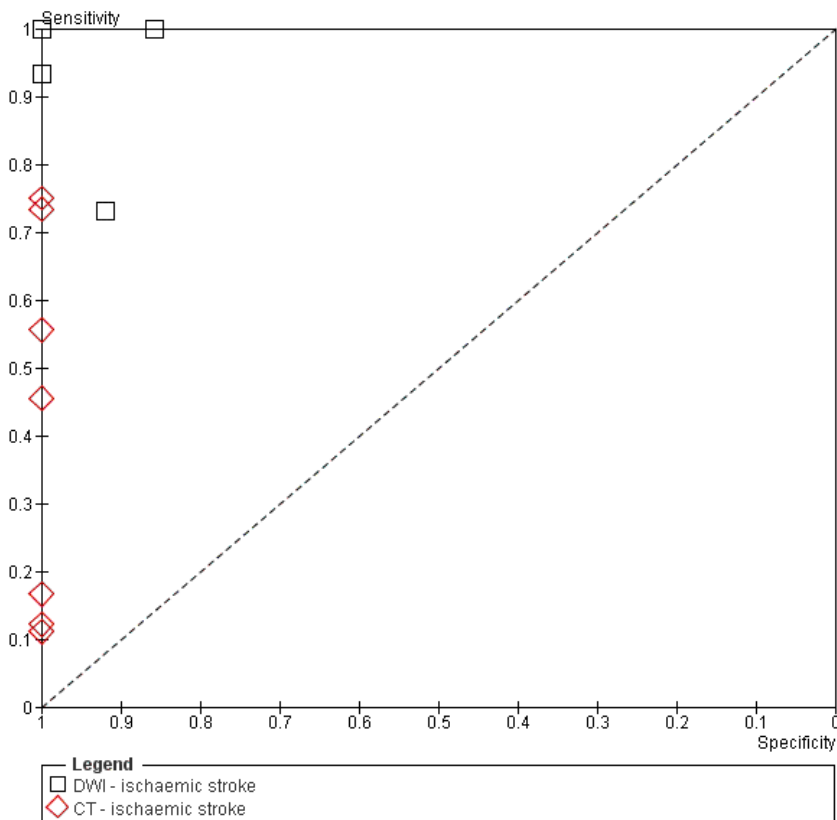
As data for the assessment of haemorrhagic stroke were derived from only two studies of low methodological quality with clear evidence of incorporation bias, we did not perform a meta-analysis of measures of test accuracy. The findings of the two studies suggested that MRI sequences may distinguish between patients with and without acute intracerebral haemorrhage with reasonably high sensitivity and specificity (see Figure 4.7).

Figure 4.5. Forest plots of sensitivity and specificity estimates for DWI and CT studies on ischaemic stroke. The squares represent each individual study; the black horizontal lines represent the 95% CIs.



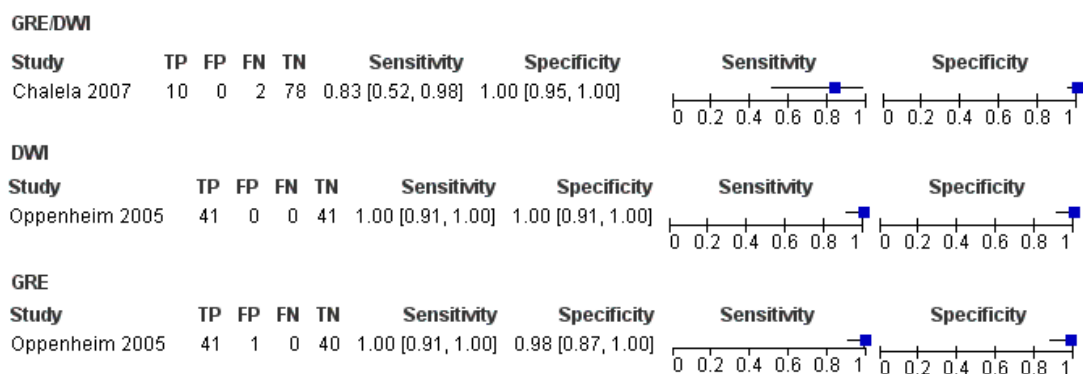
Note: TP = true positives; FP = false positives; FN = false negatives; TN = true negatives. The 'Sensitivity' and 'Specificity' columns list the numerical values of sensitivity and specificity estimates with 95% CIs for each study

Figure 4.6. ROC plot of sensitivity versus specificity for the seven studies that compared DWI with CT for the early detection of ischaemic stroke. The dotted line indicates the curve of an uninformative test.



The Chalela study²⁷ showed a sensitivity of 0.83 (95% CI 0.52 to 0.98) and a specificity of 1.00 (95% CI 0.95 to 1.00) for gradient-echo and diffusion-weighted MRI in patients assessed within three hours of stroke, versus CT and clinical assessment. The sensitivity estimate was, however, based on only 12 patients (13% of the 90 patients investigated) who were found to have acute cerebral haemorrhage. Similarly, the Oppenheim study²⁶ reported 1.00 (95% CI 0.91 to 1.00) for both the sensitivity and specificity of diffusion-weighted MRI and 1.00 (95% CI 0.91 to 1.00) sensitivity and 0.98 (95% CI 0.87 to 1.00) specificity for gradient-echo MRI performed within six hours of symptoms onset, but with no CT comparator in most of the patients. However, the high proportion (50%) of patients with haemorrhage in this study as compared to the proportion of intracerebral haemorrhage observed in the typical clinical population (10% to 15%) indicated the presence of spectrum bias (highly selected patient sample) that was likely to have increased sensitivity and influenced specificity.

Figure 4.7. Forest plots of sensitivity and specificity estimates for DWI and GRE sequences for the two studies on haemorrhagic stroke. The squares represent each individual study; the black horizontal lines represent the 95% CIs



Note: TP = true positives; FP = false positives; FN = false negatives; TN = true negatives. The 'Sensitivity' and 'Specificity' columns list the numerical values of sensitivity and specificity estimates with 95% CIs for each study

4.4 Discussion

The emergency management of patients with acute stroke relies heavily on accurate and rapid diagnosis. Early identification of patients with stroke and the distinction between ischaemic and haemorrhagic stroke are crucial for therapeutic decision making and, in particular, for selecting patients for thrombolytic treatment. CT and MRI are both used in clinical practice to identify patients with acute stroke who might benefit from thrombolytic therapy (which is absolutely contraindicated in patients with stroke due to intracerebral haemorrhage). We conducted a systematic review to compare the accuracy of these two imaging methods for detection of acute ischaemic and haemorrhagic stroke. Six studies for the detection of ischaemic stroke, one study for the detection of haemorrhagic stroke, and one study which assessed both haemorrhagic and ischaemic stroke fulfilled our inclusion criteria. For the assessment of ischaemic stroke we included only comparative studies that evaluated both imaging techniques in the same patients, as they provide the best evidence on which to judge the relative performance of CT and MRI for the detection of ischaemic stroke lesions. In this limited cohort of studies MRI had higher sensitivity than CT but similar specificity (see Figure 4.5). The two studies that contributed to the assessment of haemorrhagic stroke provide similar sensitivity and specificity estimates for MRI (see Table 4.1 and Figure 4.7). We could not assess whether CT and MRI were equally good at identifying stroke mimics as the majority of studies excluded non-stroke patients.

4.4.1 Findings on ischaemic stroke

Our findings are in line with the well-established claim in the literature that, in patients subsequently confirmed to have acute stroke, diffusion-weighted MR sequences are more sensitive for detecting acute ischaemia than plain CT; especially in the first hours after symptoms onset and chiefly in patients with mild stroke.²⁸⁻³² There are, however, important considerations to be made. The characteristics of the patient population varied between studies and all but one study included a very narrow spectrum of stroke patients. The study with a broader spectrum of patients²⁷ still only included mostly mild strokes (median score on NIHSS = 3 – a scale with a maximum score of 44) and, therefore, was not representative of the typical population being assessed for thrombolysis. The evaluation of mild cases may increase the sensitivity of DWI, which is known to be particularly useful in detecting small ischaemic lesions.³³ The exclusion of more severe cases may further disadvantage CT; more severely affected patients are more likely to have a CT-visible lesion, which may help to explain the large difference between CT and DWI sensitivity estimates in this study.²⁷ The severity of stroke was rarely reported in the remaining included studies. It is also known that many patients with severe stroke do not tolerate MRI.³⁴ Information related to the patients who were excluded because they either could not tolerate MRI or had contraindications was provided in only three studies. Of these, the largest study reported that about 11% of the patients initially screened for inclusion were subsequently excluded due to MRI contraindications.²⁷ Moreover, the majority of included studies enrolled only patients with typical anterior circulation stroke. Negative DWI findings have been reported to occur more often in posterior circulation stroke during the first 24 hours.^{35,36} Thus in

an unselected population of stroke patients, the inclusion of posterior circulation strokes could well reduce the apparently greater diagnostic accuracy of DWI compared with CT. In all but one study, CT was performed about an hour before DWI. Thus since ischaemic lesions become more visible with the passage of time, the lesions on CT could have been less conspicuous and more difficult to detect than they would have been at a later stage (a more rigorous approach would have been to determine the sequence of tests by random allocation). Variable timing of CT and DWI could also have contributed to the variability in the observed CT sensitivity estimates. Studies generally had very small sample sizes, which may have jeopardised blinding and had an effect on the estimates of accuracy, especially for sensitivity. Furthermore, CT and DWI were evaluated in a highly selected group of patients in all but one study. Most of the patients had a final diagnosis of ischaemic stroke or TIA. TIA cases were usually counted as 'stroke negative' cases, except in one study²⁷ where TIA cases with evidence of ischaemic lesions on DWI were reclassified as 'strokes' (true positive cases). This 'reclassification' might have added a negative effect on CT and did clearly switch the reference standard to an MRI diagnosis. The incorporation of DWI findings in the reference standard (incorporation bias) was likely to have inflated the observed DWI estimates of sensitivity. Moreover, although it has been demonstrated that DWI may show an apparent acute cerebral infarction in approximately half of patients with TIA³⁷ the clinical significance of DWI-positive TIAs remains uncertain. Since DWI-positive lesions can resolve, the subsequent scans may reveal no evidence of infarction. Specificity estimates were very high in all studies. Indeed, in most of the studies stroke mimics (for example cerebral neoplasms, systemic infections) or patients with

other cerebrovascular lesions were not included in the spectrum of patients assessed. This renders the sample poorly representative of the acute patients typically seen in clinical practice, where 15% to 30% of patients with an initial clinical diagnosis of stroke are ultimately found to have stroke-mimic pathologies.^{8,38-40} In turn, this makes it difficult to be certain that these estimates of accuracy apply in routine clinical practice to a wider spectrum of patients and provides no information on the accuracy of CT and MRI in detecting mimics.

4.4.2 Findings on haemorrhagic stroke

CT is the imaging modality most commonly used to distinguish the acute presentation of intracerebral haemorrhage from ischaemic stroke in the evaluation of potential candidates for thrombolytic therapy. More recently, it has been suggested that MRI, including diffusion-weighted and gradient-echo sequences, could detect haemorrhage in the first hours after stroke.^{10,13,41-43} However, methodologically rigorous data on haemorrhagic stroke are scanty and even the two studies that met our pre-defined inclusion criteria^{26,27} suffered from major methodological biases and limitations. Hence, there is insufficient evidence on which to draw any sound conclusions on the accuracy of MRI for detection of haemorrhagic stroke in routine practice. In both included studies the reference standard was a hospital discharge diagnosis which incorporated all available clinical and imaging data (including acute imaging data) but without CT in many cases, thereby leading to a comparison of MRI with itself. The presence of this incorporation bias may have overestimated the reported MRI diagnostic accuracy. Furthermore, in both studies patients with non-

stroke lesions were not assessed and it would have been useful to know how well MRI (compared with CT) could distinguish haemorrhagic lesions from non-stroke lesions (for example neoplasms). Thus, while the ability of CT to distinguish acute haemorrhagic lesions from non-stroke lesions is well established the accuracy of MRI assessment of suspected acute stroke is still somewhat unclear.

4.4.3 Summary of main results

In conclusion, we identified only a limited number of studies that directly compared MRI versus CT for the early detection of stroke lesions. The overall methodological quality of these studies was poor. Our results suggest that diffusion-weighted MRI is probably more sensitive than CT, but not more specific, for the early detection of ischaemic stroke in highly selected patient populations. Our data do not allow any comments to be made on the merits of MRI for the detection of haemorrhagic stroke. Moreover, estimates of diagnostic accuracy of CT and MRI were obtained from well-defined groups of patients with a final diagnosis of stroke so may be of limited clinical utility as they may not be applicable to the broad range of patients with suspected acute stroke usually seen in routine clinical practice. Neither practicality nor cost-effectiveness was effectively taken into consideration in the included studies. Additional well-designed studies are needed to estimate more reliably whether MRI can be used as the primary imaging modality for patients presenting with suspected acute stroke.

4.4.4 Strengths and weaknesses of the review

For the detection of ischaemic stroke, we focused exclusively on comparative studies that evaluated both CT and MRI versus a reference standard of clinical diagnosis and imaging follow up in the same patients, which is known to provide the best evidence about the diagnostic accuracy of two different methods. We searched major electronic databases to identify all relevant studies. Three review authors with different expertise (a methodologist and two neurologists) independently selected studies and extracted data. Four review authors (a methodologist, two neurologists, and a radiologist) independently assessed the quality of the included studies.

Our review has some limitations. Overall, our findings are limited by the relatively small number of comparative studies available in the literature; incomplete reporting of studies' characteristics and results; limited methodological quality; and relatively small sample sizes. Diagnostic imaging studies seem to be particularly prone to these problems.^{44,45} Shortcomings in study design may affect the estimates of diagnostic accuracy resulting in an overestimation, particularly in studies including non-representative samples of patients and invalid reference standards.^{44,45} Future studies should include an appropriate spectrum of patients; a consistent reference standard independent of the imaging modalities under investigation, to reduce incorporation bias; and blind interpretation of tests results. They should also comply with the Standards for Reporting of Diagnostic Accuracy (STARD) recommendations for improving the quality of reporting of diagnostic studies.⁴⁶

With regard to our literature searches to identify relevant studies to include in our review, a couple of points are worth raising. We included a methodological search filter in our MEDLINE and EMBASE searches to identify studies of diagnostic accuracy. The use of a search filter, even though it may have reduced the overall sensitivity of the MEDLINE search, was justified by the fact that a literature search combining MeSH terms and text words for the target condition with those for the diagnostic tests under evaluation (as for the current recommendation of the Screening and Diagnostic Tests Methods Group) would have retrieved an unmanageable number of hits; CT is an imaging test used very frequently in clinical practice and therefore referred to in many research papers. We did not search additional electronic databases, such as BIOSIS, LILACS, or Science Citation Index, firstly because the number and relevance of indexed journals in these databases are limited compared to those indexed in MEDLINE and EMBASE and secondly we are confident we have enhanced the sensitivity of our literature searches by searching 'specialised' databases (for example MEDION for systematic reviews of diagnostic accuracy) and professional bodies' websites (for example National Stroke Association, American Stroke Association, Royal College of Radiology), handsearching all conference proceedings of two major international stroke conferences for a 10-year period, and contacting experts in the field. In this way, even though relying on limited resources we have maximised sensitivity and specificity for identifying comparative studies on the use of CT and MRI for detection of acute stroke lesions.

We could not use the currently recommended summary ROC curve methodology to compare the performance of CT and DWI for the early diagnosis of ischaemic stroke

as our data were insufficient to fit this complex statistical model.^{18,19,47} Similarly, due to the limited number of identified studies, we were not able to perform sensitivity analyses to assess which methodological aspects may have contributed to clinical heterogeneity (for example time of imaging, characteristics of patient population) or heterogeneity related to study design (for example prospective versus retrospective studies, presence of incorporation bias).

We were unable to address practicality and applicability issues as only three studies mentioned the number of patients who were excluded because they could not undergo MRI. However, practical difficulties in performing MRI, instead of CT, have been documented in about 20% of stroke patients and many patients with severe stroke do not tolerate MRI.^{34,48,49} Similarly we did not assess the cost-effectiveness of MRI compared with CT as a first-line test for the early detection of stroke and we did not consider the relative impact on clinical outcomes of a policy of routine MRI versus routine CT. However, in deciding whether MRI could substitute for CT as the primary method for early imaging of patients with suspected ischaemic or haemorrhagic stroke it is important to consider the relative diagnostic accuracy of each imaging test together with practicality and cost-effectiveness issues. In many countries CT is known to be rapid, easy to tolerate, and more readily available in most emergency settings. On the other hand MRI is not immediately available in many hospitals and is more expensive, contraindicated for patients with pacemakers and metal implants, and can be unpleasant or difficult to tolerate especially for patients with more severe strokes. A recent survey conducted in the UK showed that even though 78% of all acute hospitals that admitted patients with acute stroke had

access to MRI facilities, MRI was rarely performed either at all or sufficiently quickly to be of value in the acute management of stroke.⁵⁰ A European survey suggests that similar problems exist in the rest of the EU and, in fact, placed the UK at the top of the table for comprehensive stroke centres.⁵¹

4.4.5 Applicability of findings to clinical practice and policy

We reviewed the diagnostic accuracy of MRI compared with CT for acute ischaemic stroke, and the accuracy of MRI for early detection of haemorrhagic stroke. There is some evidence that MRI is more accurate than CT for the detection of mild ischaemic strokes. However, the use of MRI in the management of acute patients needs to take into consideration practicality and cost-effectiveness. In many countries CT is quicker to perform, inexpensive, applicable to a higher proportion of acutely ill stroke patients, and more readily available in most emergency care settings. MRI is contraindicated in patients with pacemakers and some metal implants. In acutely ill stroke patients it may be difficult to monitor their condition while they are being MRI scanned (which increases the risk of any developing respiratory difficulty or cardiovascular compromise being detected during the scan and so may have adverse effects for the patient). If the patient is confused or restless as a result of the stroke, the patient may not be able to co-operate for the longer scan times required for MRI. Furthermore, in clinical practice CT is the most used imaging technique for the diagnosis of acute intracerebral haemorrhage (and therefore for selecting patients for thrombolytic therapy). The role of MRI as the first choice modality for patients presenting with stroke symptoms requires further investigations.

4.5 Authors' conclusions

4.5.1 Implications for practice

It is likely that, in the future, both CT and MRI techniques will be more widely available in many countries. Pending further evidence, both techniques should be used in a complementary way with CT for the majority of strokes pre-thrombolysis and MRI for milder strokes, according to local, specific clinical needs.

4.5.2 Implications for research

Future research should focus on a robust and objective cost-effectiveness comparison of CT and MRI with particular attention to the evaluation of patients in broader and unselected patient populations more relevant to routine clinical practice in non-specialist stroke centres. In particular, further studies are needed to provide clear evidence that MRI can be used as the imaging modality of first choice for patients with suspected acute stroke in routine practice, and that patients without evidence of acute intracerebral haemorrhage on MRI really do not have acute intracranial bleeding (and hence can be safely considered for thrombolytic treatments).

4.5.3 Implications for the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy

This review was used to pilot the methods for Cochrane diagnostic test accuracy reviews. It was thus assembled, written and analysed on the basis of guidance in the draft version of the Handbook for Systematic Reviews of Diagnostic Test Accuracy. The Screening and Diagnostic Tests Working Group has used these pilot data to inform the development of the Handbook and to test the new proposed methodology. However, as the complete version of the Handbook has not been fully published yet, the impact of this thesis on the final version of the Handbook cannot be judged yet.

4.5.4 Summary of this review's findings

Chapter 5 assesses two possible approaches to producing a brief convenient summary of the complex evidence presented in this systematic review of diagnostic test accuracy. The Summary of Findings table in Appendix 4 was developed to summarise the review results of the seven included studies on ischaemic stroke.

This review has been conducted as a pilot Cochrane systematic review of diagnostic test accuracy and published in the Cochrane Library:

Brazzelli M, Sandercock PAG, Chappell FM, Celani MG, Righetti E, Arestis N, Wardlaw JM, Deeks JJ. Magnetic resonance imaging versus computed tomography for detection of acute vascular lesions in patients presenting with stroke symptoms. Cochrane Database of Systematic Reviews 2009, Issue 4. Art. No.: CD007424. DOI: 10.1002/14651858.CD007424.pub2

A short version of the review has also been published in the journal Stroke:

Brazzelli M, Sandercock PAG, Chappell FM, Celani MG, Righetti E, Arestis N, Wardlaw JM, Deeks JJ. MRI versus CT for detection of acute vascular lesions in patients presenting with stroke symptoms. Stroke 2010; 41: e427-e428

References

1. Warlow C, Sudlow C, Dennis M, Wardlaw J, Sandercock P. Stroke. *Lancet* 2003;362:1211-24
2. Feigin VL, Lawes CMM, Bennett DA, Anderson CS. Stroke epidemiology: a review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. *Lancet Neurology* 2003;2:43-53
3. Bamford J, Sandercock P, Dennis M, Burn J, Warlow C. Classification and natural history of clinically identifiable subtypes of cerebral infarction. *Lancet* 1991;337:1521-6
4. Coull AJ, Lovett JK, Rothwell PM. Population based study of early risk of stroke after transient ischaemic attack or minor stroke: implications for public education and organisation of services. *BMJ* 2004;328(7435):326
5. NINDS rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *New England Journal of Medicine* 1995;333:1581-7
6. Furlan A, Higashida R, Wechsler L. Intra-arterial prourokinase for acute ischemic stroke. The PROACT II study: a randomised controlled trial. PROlyse in Acute Cerebral Thromboembolism. *JAMA* 1999;34:2003-11

7. Wardlaw JM, Keir SL, Seymour J, Lewis S, Sandercock PA, Dennis MS, et al. What is the best imaging strategy for acute stroke? *Health Technology Assessment* 2004;8(1):1-180
8. Hacke W, Kaste M, Fieschi C, van Kummer R, Davalos A, Meier D, et al. Randomised double-blind placebo-controlled trial of thrombolytic therapy with intravenous alteplase in acute ischaemic stroke (ECASS II). *Lancet* 1998;352:1245-51
9. Lansberg MG, Norbash AL, Marks MP, Tong DC, Moseley ME, Albers GW. Advantages of adding diffusion-weighted magnetic resonance imaging to conventional magnetic resonance imaging for evaluating acute stroke. *Archives of Neurology* 2000;57:1311-6
10. Linfante I, Llinas RH, Caplan LR, Warach S. MRI features of intracerebral hemorrhage within 2 hours from symptom onset. *Stroke* 1999;30:2263-7
11. Lin DD, Filippi CG, Steever AB, Zimmerman RD. Detection of intracranial hemorrhage: comparison between gradient-echo images and b_0 images obtained from diffusion-weighted echo-planar sequences. *American Journal of Neuroradiology* 2001;22:1275-81

12. Fiebach JB, Schellinger PD, Gass A, Kucinski T, Siebler M, Villringer A, et al. Stroke magnetic resonance imaging is accurate in hyperacute intracerebral hemorrhage: a multicenter study on the validity of stroke imaging. *Stroke* 2004;35:502-6

13. Kidwell CS, Chalela JA, Saver JL, Starkman S, Hill MD, Demchuk AM, et al. Comparison of MRI and CT for detection of acute intracerebral hemorrhage. *JAMA* 2004;292:1823-30

14. Schellinger PD, Bryan RN, Caplan LR, Detre JA, Edelman RR, Jaigobin C, Kidwell CS, Mohr JP, Sloan M, Sorensen AG, Warach S. Evidence-based guideline: The role of diffusion and perfusion MRI for the diagnosis of acute ischemic stroke. Report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. *Neurology*, 2010; 75: 177-85

15. Astin M, Brazzelli M, Fraser C, Counsell C, Needham G, Grimshaw J. Developing a sensitive search strategy in MEDLINE to retrieve studies on assessment of the diagnostic performance of imaging techniques. *Radiology* 2008;247:365-73

16. Reference Manager Professional Edition [Computer program]. Version 10. ISI ResearchSoft, 1984-2001

17. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology* 2003;3:25

18. Macaskill P. Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. *Journal of Clinical Epidemiology* 2004;57:925-32

19. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JAC. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2006;1:1-21

20. Sorensen AG, Buonanno FS, Gonzalez RG, Schwamm LH, Lev MH, Huang-Hellinger FR, et al. Hyperacute stroke: evaluation with combined multisection diffusion-weighted and hemodynamically weighted echo-planar MR imaging. *Radiology* 1996;199:391-401

21. Barber PA, Darby DG, Desmond PM, Gerraty RP, Yang Q, Li T, et al. Identification of major ischemic change. Diffusion-weighted imaging versus computed tomography. *Stroke* 1999;30:2059-65

22. Bozzao A, Floris R, Giuliani V, Baviera ME, Montanaro M, Salvatore C, Simonetti G. Clinical efficacy of diffusion weighted MRI on acute cerebral ischemia [Efficacia clinica della risonanza magnetica con sequenze pesate in diffusione nella valutazione dell'ischemia cerebrale acuta]. *La Radiologia Medica* 1999;98:144-50
23. Gonzalez RG, Schaefer PW, Buonanno FS, Schwamm LH, Budzik RF, Rordorf G, Wang B, Sorensen AG, Koroshetz WJ. Diffusion-weighted MR Imaging: Diagnostic accuracy in patients imaged within 6 hours of stroke symptom onset. *Radiology*, 1999; 210: 155-62
24. Urbach H, Flacke S, Keller E, Textor J, Berlis A, Hartmann A, et al. Detectability and detection rate of acute cerebral hemisphere infarcts on CT and diffusion-weighted MRI. *Neuroradiology* 2000;42:722-7
25. Saur D, Kucinski T, Grzyska U, Eckert B, Eggers C, Niesen W, et al. Sensitivity and interrater agreement of CT and diffusion weighted MR imaging in hyperacute stroke. *American Journal of Neuroradiology* 2003;24:878-85
26. Oppenheim C, Touze E, Hernalsteen D, Peeters A, Lamy C, Mas JL, Meder JF, Cosnard G. Comparison of five MR sequences for the detection of acute intracranial hemorrhage. *Cerebrovascular Diseases* 2005;20:388-94

27. Chalela JA, Kidwell CS, Nentwich LN, Luby M, Butman JA, Demchuk AM, et al. Magnetic resonance imaging and computed tomography in emergency assessment of patients with suspected acute stroke: a prospective comparison. *Lancet* 2007;369:293-8
28. Fiebach J, Jansen O, Schellinger P, Knauth M, Hartmann M, Heiland S, et al. Comparison of CT with diffusion-weighted MRI in patients with hyperacute stroke. *Neuroradiology* 2001;43(8):628-32
29. Fiebach JB, Schellinger PD, Jansen O, Meyer M, Wilde P, Bender J, et al. CT and diffusion-weighted MR imaging in randomized order. *Stroke* 2002;33:2206-10
30. Jaillard A, Hommel M, Baird AE, Linfante I, Llinas RH, Caplan LR, et al. Significance of early CT signs in acute stroke: A CT scan-diffusion MRI study. *Cerebrovascular Diseases* 2002;. 13(1):47-56
31. Kucinski T, Vaterlein O, Glauche V, Fiehler J, Klotz E, Eckert B, et al. Correlation of apparent diffusion coefficient and computed tomography density in acute ischemic stroke. *Stroke* 2002;33(7):1786-91

32. Mullins ME, Schaefer PW, Sorensen AG, Halpern EF, Ay H, He J, et al. CT and conventional and diffusion-weighted MR imaging in acute stroke: study in 691 patients at presentation to the emergency department. *Radiology* 2002;224(2):353-60

33. Keir SL, Wardlaw JM, Bastin ME, Dennis MS. In which patients is diffusion-weighted magnetic resonance imaging most useful in routine stroke care? *Journal of Neuroimaging* 2004;14(2):118-22

34. Hand PJ, Wardlaw JM, Rowat AM, Haisma JA, Lindley RI, Dennis MS. Magnetic resonance brain imaging in patients with acute stroke: feasibility and patient-related difficulties. *Journal of Neurology, Neurosurgery, and Psychiatry* 2005;76:1525-7

35. Lövblad KO, Laubach HJ, Baird AE, Curtin F, Schlaug G, Edelman RR, et al. Clinical experience with diffusion-weighted MR in patients with acute stroke. *American Journal of Neuroradiology* 1998;19:1061-6

36. Oppenheim C, Stanescu R, Dormont D, Crozier S, Marro B, Samson Y, et al. False-negative diffusion-weighted MR findings in acute ischemic stroke. *American Journal of Neuroradiology* 2000;21:1434-40

37. Kidwell CS, Alger JR, Di Salle F, Starkman S, Villablanca P, Bentson J, et al. Diffusion MRI in patients with transient ischemic attacks. *Stroke* 1999;30:1174-80
38. Libman RB, Wirkowski E, Alvir J, Rao TH. Conditions that mimic stroke in the emergency department. Implications for acute stroke trials. *Archives of Neurology* 1995;52:1119-22
39. Scott PA, Silbergleit R. Misdiagnosis of stroke in tissue plasminogen activator-treated patients: characteristics and outcomes. *Annals of Emergency Medicine* 2003;42:611-8
40. Hand PJ, Kwan J, Lindley RI, Dennis MS, Wardlaw JM. Disitnguish between stroke and mimic at the bedside: the Brain Attack Study. *Stroke* 2006;37:769-75
41. Weingarten K, Filippi C, Zimmermann RD, Deck MD. Detection of hemorrhage in acute cerebral infarction. Evaluation with spin-echo and gradient-echo MRI. *Clinical Imaging* 1994;18:43-55

42. Patel MR, Edelman RR, Warach S. Detection of hyperacute primary intraparenchymal hemorrhage by magnetic resonance imaging. *Stroke* 1996;27:2321-4
43. Atlas SW, Thulborn KR. MR detection of hyperacute parenchymal hemorrhage of the brain. *American Journal of Neuroradiology* 1998;19:1471-7
44. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061-6
45. Rutjes AWS, Reitsma JB, Di Nisio F, Smidt N, van Rijn JC, Bossuyt MMP. Evidence of bias and variation in diagnostic accuracy studies. *Canadian Medical Association Journal* 2006;174(4). [DOI: 10.1503/cmaj.050090]
46. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. STAndards for Reporting of Diagnostic accuracy steering group. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003;326:41-4

47. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *Journal of Clinical Epidemiology* 1995;48:119-30
48. Singer OC, Sitzer M, du Mesnil de Rochemont R, Neumann-Haefelin T. Practical limitations of acute stroke MRI due to patient-related problems. *Neurology* 2004;62:1848-9
49. Barber PA, Hill MD, Eliasziw M, Demchuk AM, Pexman JHW, Hudon ME, et al. Imaging of the brain in acute ischaemic stroke: Comparison of computed tomography and magnetic resonance diffusion-weighted imaging. *Journal of Neurology, Neurosurgery, and Psychiatry* 2005;76(11):1528-33
50. Kane I, Whiteley WN, Sandercock PAG, Wardlaw JM. Availability of CT and MR for assessing patients with acute stroke. *Cerebrovascular Diseases* 2008;25:375-7
51. Leys D, Ringelstein EB, Kaste M, Hacke W, for the Executive Committee of the European Stroke Initiative. Facilities available in European hospitals treating stroke patients. *Stroke* 2007;38:2985-91

5. To assess the usefulness of two forms of summaries of a systematic review of diagnostic accuracy

5.1 Introduction

The correct use of one or more diagnostic tests and the interpretation of their results are crucial steps in determining the need for a particular medical intervention. The choice of the wrong diagnostic test or inaccurate interpretation of its results may lead to incorrect diagnosis and inappropriate medical care. A correct diagnosis is essential for providing effective treatments and establishing proper patients' management. With each new technological advancement in diagnostic techniques, clinicians are constantly challenged to master information from studies that evaluate the accuracy and applicability of new and existing diagnostic tests, and frequently have to choose one of several possible alternative tests or diagnostic strategies.

In particular, many clinicians struggle to apply information on quantitative measures of diagnostic test accuracy correctly. The term 'diagnostic accuracy' is used in the medical literature to indicate the ability of a test to categorise patients accurately according to the presence or absence of a specified disease. To establish accuracy, the results of a test are compared with the results of a reference standard, in other words, the accepted or best available method for finding out whether or not patients have the target disease. Test accuracy is often expressed in terms of sensitivity and specificity, predictive values, and likelihood ratios. Even though clinicians are

familiar with the definitions of sensitivity and specificity of a test, only a minority of them seem to consider these measures of accuracy before ordering a test in clinical practice or prove to apply them correctly.¹⁻² Recommended methods which require formal calculations to estimate the probability of disease (e.g. Bayesian and likelihood ratio transformations), are very rarely used.¹ This difficulty in applying knowledge of measures of diagnostic test accuracy raises the question of which is the most effective way to communicate the results of a formal assessment of a particular test to clinicians. This is now even more relevant in view of recent developments in the science of evaluating the performance of diagnostic tests. Firstly, there is now a degree of consensus on how individual studies of diagnostic accuracy should be reported (Standards for Reporting of Studies of Diagnostic Accuracy (STARD) initiative). Secondly, the Cochrane Collaboration has recently established a framework for the conduct and reporting of systematic reviews of test accuracy. The Cochrane Collaboration decided to include “diagnostic accuracy” within the scope of Cochrane reviews in March 2003 and the first diagnostic accuracy review conducted according to the current recommended methodology was published in the Cochrane Database of Systematic Reviews in October 2008.⁹ This development has occurred at the same time as a number of methodological papers on the methodology of systematic reviews of diagnostic test accuracy.³⁻⁸

Systematic reviews of diagnostic accuracy provide clinicians with a convenient summary of the best available evidence on the performance of a diagnostic test and on its potential to be used alongside alternative tests. However, if the findings of such systematic reviews are to be applied correctly in clinical practice this requires the clinicians to have a good understanding of: the measures used to assess test

accuracy; the methodological quality of the included primary studies; the context in which the test is performed. Furthermore, clinicians should also consider the implications of any false negatives (cases missed by the diagnostic test) or false positive (cases wrongly identified by the diagnostic test) results when making decisions about the use of a specific diagnostic test.

Busy clinicians often tend to rely on the information provided in the 'Abstract' of a systematic review rather than read its methods, results, and conclusions in detail. The abstract, as a concise summary of the review methods and findings, provides, however, only limited information and decisions based on that alone may be inaccurate.¹⁰

To provide a more informative summary of the data from a systematic review that illustrates the clinical impact of the findings when applied in different (high and low risk) populations, the Cochrane Collaboration has recently decided to include a Summary of Findings (SoF) table within the structure of Cochrane reviews. SoF tables have been proposed as a way to increase usability of reviews and help clinicians to make better informed clinical decisions. Review authors of systematic reviews of randomised controlled trials may create their SoF tables using the recommendations contained in the (Cochrane Handbook for Systematic Reviews of Interventions).¹¹ The template of standard Cochrane SoF tables includes six elements and a specific format: i) a list of all important outcomes, both desirable and undesirable; ii) a measure of the typical burden of these outcomes (e.g. illustrative risk, or illustrative mean, on control intervention); iii) absolute and relative

magnitude of effect (if appropriate); iv) numbers of participants and number of studies addressing the listed outcomes; v) a rating of the overall quality of evidence for each outcome (using the GRADE approach);¹²⁻¹³ iv) space for comments.¹¹ An additional software (GRADEprofile) is available to reviews authors for the preparation of SoF tables (website of GRADEpro). At present, however, there is not empirical evidence on how clinicians interpret SoF table for reviews of randomised controlled trials. The template for preparing SoF tables for systematic reviews of diagnostic accuracy has not yet been established and review authors are responsible for creating their own SoF tables.

The Cochrane systematic review of diagnostic accuracy,¹⁴ which I presented in Chapter 4, had served as a pilot review for the development of Cochrane diagnostic accuracy reviews. It was undertaken in collaboration with the Screening and Diagnostic Tests Methods Group. Performing this review enabled us to develop and test the methods for Cochrane systematic reviews of diagnostic test accuracy. We therefore had the opportunity to develop and test the design of a SoF table for diagnostic test accuracy studies. In the SoF table we included information on the characteristics and main results of the review, together with the major limitations identified and implications for clinical practice. To some extent, the information shown in the SoF table duplicates that of the Abstract. In the proposed SoF table, the information is presented in both tabular and graphical format and additional details are provided to enable the reader to put the review results in context and judge their applicability in clinical practice.

At present, there is very little evidence on the best way of presenting summary results of diagnostic accuracy reviews to clinicians and on how clinicians interpret results of such reviews. For this reason I decided to carry out a study to evaluate two alternative strategies for presenting the results of systematic reviews of diagnostic test accuracy to clinicians.

The aim of this study was to gather clinicians' opinion on two different ways of presenting the summary results of a systematic review of diagnostic test accuracy and to assess their ability to interpret information contained in both the SoF table and the Abstract of my pilot systematic review.

5.1.1 Objectives

To determine whether two different forms of summary of a systematic review of diagnostic test accuracy may influence clinicians' ability to interpret its results, and to assess their preferences about the format of the summary.

5.2 Methods

5.2.1 Study Design

At a series of small group meetings of clinicians, I presented two forms of summary of my pilot systematic review assessing the accuracy of MRI and CT for the detection of acute ischaemic lesions in patients presenting with stroke symptoms:¹⁴ a conventional Abstract and our proposed format of a SoF table. I presented each participant individually with the two forms of summary in alternate order (some

participants received the Abstract first and then the SoF table whilst other participants received the SoF table first and then the Abstract). For each format I asked the clinicians to complete a short questionnaire (see Appendix 5).

5.2.2 Participants

I recruited groups of neurologists, radiologists, and neuroradiologists (at different seniority in their professional career) at the Western General Hospital in Edinburgh (UK) at a series of meetings held between June and November 2009. I also administered the questionnaire to neurologists, radiologists, and general medicine physicians in Perugia (Italy) at a series of seminars organised by the Cochrane Neurological Network. I selected these clinicians as being reasonably representative of the types of medical practitioners most likely to need to read and correctly interpret Cochrane systematic reviews of test accuracy relevant to stroke.

5.2.3 Development of Abstract and Summary of Findings table

I initially developed and tested the SoF table by seeking the views of a small group of experienced neurologists and radiologists who provided information on the aspects they deemed important to include in a summary of findings table. In particular, they stressed the need for: the review question to be specified in an unambiguous and precise way; more information concerning the main limitations of the studies included in the review; a clear formulation of the implications of findings for clinical practice; the results to be shown by a graphical as well as a numerical representation. I re-designed the SoF table in the light of this initial feedback and

further tested it in a group of clinicians outside the study cohort with the aim to ensure consistency, ease of use, and better presentation of results. I structured the content of the SoF table in four main sections. In the title of the SoF I specified the clinical question in terms of the tests under investigation, the population studied and the clinical condition assessed; in the first part of the table, I provided general information on the number of studies included in the systematic review, geographical location of the studies, clinical setting, total number of participants, type of diagnostic (index test and comparator test) and reference standard; in the second part of the table, I listed all major limitations of included studies and provide an overall assessment of their methodological quality; in the third part of the table, I displayed the main results of the review using both a numerical summary of estimates and a graphical representation (forest plot); finally in the fourth part of the table, I presented some information on the applicability of results into clinical practice and on costs (see Box 5.1). The Abstract was prepared in accordance with the Cochrane Collaboration guidelines.¹¹ It was limited to 400 words and summarised the key methods, results, and conclusions of our pilot review. I used a simple and precise terminology, and I included only well known abbreviations, which referred to the type of imaging tests. The content of the Abstract was structured as follows: Background; Objectives; Search methods, Selection criteria, Data collection and analysis, Results, and Authors' conclusions (see Box 5.2). Compared to the Abstract, the SoF table contained more details on the limitations of the included studies and applicability of results in clinical practice. Moreover, it included a numerical as well as a graphical summary of main results.

5.2.4 Development of a questionnaire to assess clinicians' demographics, understanding of results, and preferences

I developed a short questionnaire and then tested it in a group of 21 clinicians outwith the study cohort (see Appendix 5). The first part allowed me to collect clinicians' demographic information: speciality (e.g. General Radiology, Neurology, Neuroradiology, Stroke Medicine, Geriatric Medicine); postgraduate qualifications (e.g. MD, PhD); years of experience in clinical practice; time spent in the previous year managing stroke patients; research involvement (as a percentage of Full Time Equivalency); and gender.

The second part of the questionnaire consisted of:

- A visual analogue scale (i.e. Likert scale) to determine the clinicians' confidence in interpreting measures of test accuracy;
- Six multiple choice questions to evaluate the clinicians' ability to interpret the main findings and implications for clinical practice of my systematic review of diagnostic accuracy presented either in the Abstract or in the SoF table. Each of the six questions was structured in five possible answers. The wording of the five answers was consistent across questions. For each question a neutral response was included amongst the five possible answers. The five possible answers were present in a different order in each question. The six questions aimed at assessing whether: i) MRI was more accurate than CT for the early detection of ischaemic lesions, ii) MRI was more accurate than CT for the early detection of mild ischaemic stroke, iii) MRI was more accurate than CT for the early detection of non-stroke lesions, iv) MRI was more sensitive but not more specific than CT for the early detection of

ischaemic lesions, v) either MRI or CT could be used for the early detection of ischaemic lesions, vi) more diagnostic studies comparing MRI with CT in acute stroke patients were needed.

For each question only one answer was considered correct. The correct answers were developed in discussion with two clinical experts (including a very experienced neuroradiologist), who contributed to the pilot systematic review, on the basis of the results and conclusions of the review.

The second part of the questionnaire was presented after each summary document (Abstract and SoF table).

The third part of the questionnaire sought the clinicians' preference for the format of presentation:

- Two visual analogue scales (i.e. Likert scales) to assess their opinion on whether information in the Abstract and SoF table were clearly presented;
- Three open questions to determine which 'summary' (Abstract or SoF table) they favoured and whether there was any additional information they would have liked to have included in it;
- A multiple choice question to assess their opinion on whether the two summaries contained useful information
- Two additional questions to assess their preference regarding terms used to describe results. Tests results were provided in two possible versions: a simple numerical description (e.g. "The test has a sensitivity of 99%) and a

verbal description (e.g. “Out of 100 patients with a diagnosis of acute stroke imaged with the test, 1 will not show a true lesion).

Box 5.1. Summary of findings table

Review question: *Comparison of diffusion-weighted magnetic resonance imaging with conventional computer tomography for the early detection of ischaemic brain lesions in patients suspected of stroke*

Patient population: Adults suspected of acute stroke

Setting: Hospital departments

Geographical location: Studies were conducted in Europe (3 studies), the USA (3 studies), and in Australia (one study)





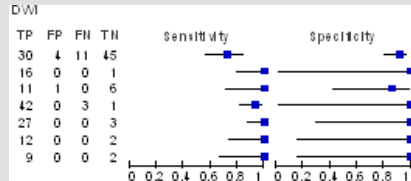
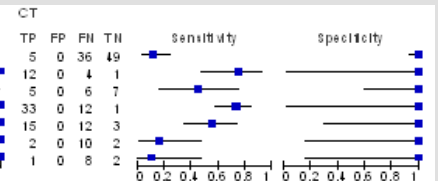
Index test: Diffusion-weighted magnetic resonance imaging (DWI) performed within 12 hours of stroke onset

Alternative test: Computer tomography (CT) performed within 12 hours of stroke onset

Reference standard: Clinical assessment and imaging follow-up (CT or MRI) in all patients

Included studies: Seven comparative studies that evaluated DWI and CT in the same patients

Total number of patients assessed: 226

LIMITATIONS OF INCLUDED STUDIES	OVERALL QUALITY ASSESSMENT																																																																								
<ul style="list-style-type: none"> ▪ Limited number of included studies (7 studies); small sample sizes; presence of incorporation bias ▪ DWI and CT were evaluated in highly selected patient samples (patients with high probability of stroke), which therefore are not representative of the typical population of patients presenting with 'suspected acute stroke' to an emergency department (poor generalisability of results) ▪ The stroke vascular territory was not reported in the majority of included studies although it is likely that they enrolled patients with typical anterior circulation stroke ▪ Only a minority of the studied patients had severe strokes (in whom DWI might be contraindicated) ▪ The high proportion of mild strokes and reclassification of TIA cases with a positive DWI lesion as strokes might have inflated the DWI sensitivity estimate ▪ In most of the studies stroke mimics were not included ▪ In all but one study, CT was performed before DWI (reducing the sensitivity of CT to detect ischaemia) 	<p style="text-align: center;">OVERALL QUALITY ASSESSMENT</p> <div style="text-align: center;">  </div> <div style="text-align: center; margin-top: 20px;">    </div>																																																																								
<table border="0" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th colspan="4" style="text-align: left; border-bottom: 1px solid black;">DWI</th> <th colspan="4" style="text-align: left; border-bottom: 1px solid black;">CT</th> </tr> <tr> <th>TP</th><th>FP</th><th>FN</th><th>TN</th> <th>TP</th><th>FP</th><th>FN</th><th>TN</th> </tr> </thead> <tbody> <tr> <td>30</td><td>4</td><td>11</td><td>45</td> <td>5</td><td>0</td><td>36</td><td>49</td> </tr> <tr> <td>16</td><td>0</td><td>0</td><td>1</td> <td>12</td><td>0</td><td>4</td><td>1</td> </tr> <tr> <td>11</td><td>1</td><td>0</td><td>6</td> <td>5</td><td>0</td><td>6</td><td>7</td> </tr> <tr> <td>42</td><td>0</td><td>3</td><td>1</td> <td>33</td><td>0</td><td>12</td><td>1</td> </tr> <tr> <td>27</td><td>0</td><td>0</td><td>3</td> <td>15</td><td>0</td><td>12</td><td>3</td> </tr> <tr> <td>12</td><td>0</td><td>0</td><td>2</td> <td>2</td><td>0</td><td>10</td><td>2</td> </tr> <tr> <td>9</td><td>0</td><td>0</td><td>2</td> <td>1</td><td>0</td><td>8</td><td>2</td> </tr> </tbody> </table> <div style="display: flex; justify-content: space-around; margin-top: 10px;"> <div style="text-align: center;"> <p>Sensitivity</p>  </div> <div style="text-align: center;"> <p>Specificity</p>  </div> </div>	DWI				CT				TP	FP	FN	TN	TP	FP	FN	TN	30	4	11	45	5	0	36	49	16	0	0	1	12	0	4	1	11	1	0	6	5	0	6	7	42	0	3	1	33	0	12	1	27	0	0	3	15	0	12	3	12	0	0	2	2	0	10	2	9	0	0	2	1	0	8	2	<p style="text-align: center;">Summary Effect (95% CI)</p> <p>DWI sensitivity 0.99 (0.23 to 1.00)</p> <p>DWI specificity 0.92 (0.83 to 0.97)</p> <p>CT sensitivity 0.39 (0.16 to 0.69)</p> <p>CT specificity 1.00 (0.94 to 1.00)</p>
DWI				CT																																																																					
TP	FP	FN	TN	TP	FP	FN	TN																																																																		
30	4	11	45	5	0	36	49																																																																		
16	0	0	1	12	0	4	1																																																																		
11	1	0	6	5	0	6	7																																																																		
42	0	3	1	33	0	12	1																																																																		
27	0	0	3	15	0	12	3																																																																		
12	0	0	2	2	0	10	2																																																																		
9	0	0	2	1	0	8	2																																																																		

CONCLUSIONS AND COMMENTS: In the small cohort of included studies, DWI is more sensitive than CT - but not more specific - for the early detection of ischaemic stroke. The small amount of data and the presence of methodological biases preclude any reliable calculation - from the sensitivity and specificity estimates of CT and DWI - of a positive/negative stroke diagnosis at different rates of stroke prevalence.

APPLICABILITY OF TESTS IN CLINICAL PRACTICE:

None of the studies addressed practicality. CT is known to be quicker to perform and more readily available in most emergency care settings than magnetic resonance imaging (MRI). MRI is contraindicated in patients with pacemakers and some metal implants. In acutely ill stroke patients, it may be difficult to monitor the patient's clinical condition while being MR scanned. If the patient is confused or restless as a result of the stroke, the patient may not be able to cooperate for the longer scan times of MRI.

COSTS: None of the studies included a cost-effectiveness evaluation. MRI is known to be more expensive than CT

Box 5.2: Abstract of the pilot review

Review question: *Comparison of diffusion-weighted magnetic resonance imaging with conventional computer tomography for the early detection of ischaemic brain lesions in patients suspected of stroke*

ABSTRACT

Background

In clinical practice, computed tomography (CT) is extensively used in the management of acute stroke, especially for the rapid exclusion of intracerebral haemorrhage. Magnetic resonance imaging (MRI) has been claimed to be more sensitive for the diagnosis of acute ischaemic stroke, even though its sensitivity for the early detection of intracerebral haemorrhage is still debated.

Objectives

To compare the diagnostic accuracy of diffusion-weighted MRI (DWI) and CT for acute ischaemic stroke.

Search Methods

We searched MEDLINE and EMBASE for articles published from January 1995 to March 2009 and we perused bibliographies of relevant studies for additional references.

Selection Criteria

We selected studies that compared DWI and CT for detection of ischaemic stroke in the same patients; had imaging performed within 12 hours of stroke onset; and presented sufficient data to allow construction of contingency tables.

Data Collection and Analysis

Three authors independently extracted data on study characteristics and measures of accuracy. Data were assessed by random-effects and fixed-effect meta-analyses.

Results

Seven studies with a total of 226 participants, met our inclusion criteria. The spectrum of patients was relatively narrow in all studies, sample sizes were small, there was substantial incorporation bias, and blinding procedures were often incomplete. Amongst the patients subsequently confirmed to have acute ischaemic stroke (161/226), the summary estimates for DWI were: sensitivity 0.99 (95% CI 0.23 to 1.00), specificity 0.92 (95% CI 0.83 to 0.97). The summary estimates for CT were: sensitivity 0.39 (95% CI 0.16 to 0.69), specificity 1.00 (95% CI 0.94 to 1.00). We were not able to assess practicality and cost-effectiveness.

Authors' conclusions

DWI appears to be more sensitive than CT for the early detection of ischaemic stroke in highly selected patients. However, the variability in the quality of included studies and the presence of spectrum and incorporation biases render the reliability and generalisability of observed results questionable. Further well-designed studies, without methodological biases, in more representative patient samples, with practicality and cost estimates, are now needed to determine which patients should undergo MRI and which CT in suspected acute stroke.

5.2.5 Planned Analyses

I analysed data from the questionnaires which were fully completed in their second part and fully or partly completed in their first and third parts. I entered the data into an Excel spreadsheet.

I summarised the demographic characteristics of participants in a narrative way. I summarised the clinicians' responses to both the multiple choice questions and the open questions quantitatively (mean scores with 95% CIs or proportions). For the second part of the questionnaire, I calculated, for all participants, the average number of correct answers to the six multiple choice questions presented after the Abstract and the SoF table as well as the mean difference between the number of correct answers to each summary document. I also calculated the distribution of correct answers to each of the six multiple choice questions after each summary document. I checked for the possible existence of an order effect (i.e. whether the number of correct answers varied according to the order of presentation of the two summary documents). I presented results as mean values with either 95% confidence intervals or standard deviation as the normality assumption for the data proved to be valid. I tested within-person differences between Abstract and SoF table results by means of one sample *t*-test and the differences between Abstract and SoF results by paired *t*-test. I also assessed whether the proportions of correct answers varied according to clinicians' speciality or according to their degree of confidence in understanding measures of diagnostic accuracy. For all analyses I set statistical significance at $p=0.05$. I used MINITAB (Minitab® 15.1.20.0.) for the statistical analyses.

5.3 Results

5.3.1 Response Rate

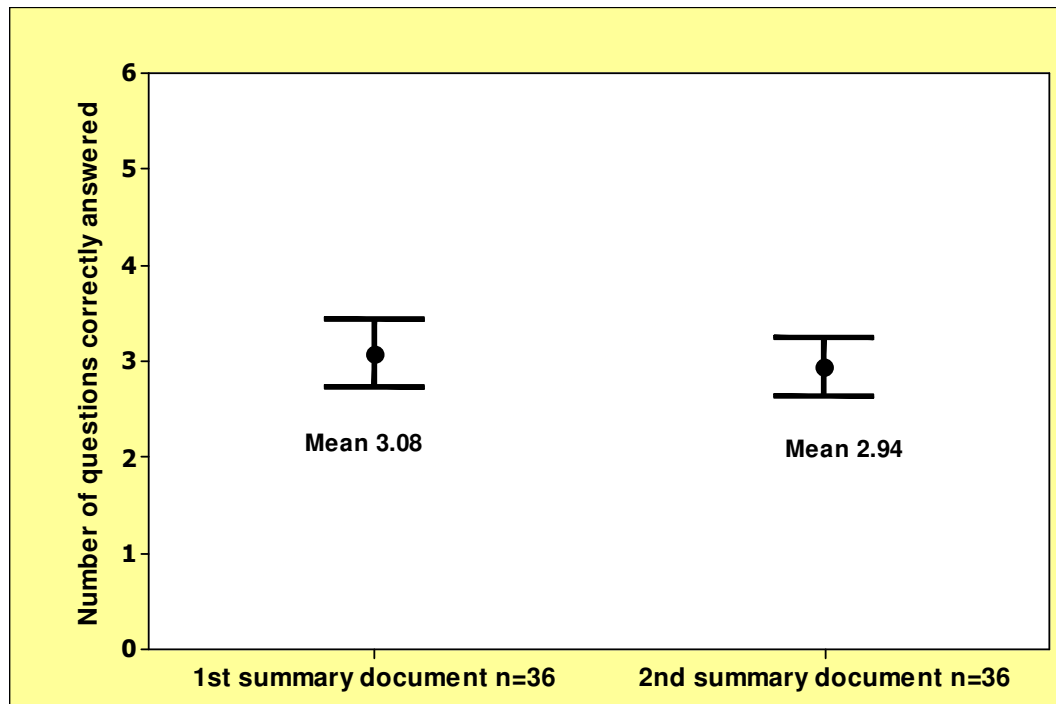
Overall 65 clinicians were presented with the two forms of summary and the short questionnaire. Six questionnaires were subsequently excluded because they were returned blank; five were excluded because the six multiple-choice questions were not fully answered; and 18 were not returned. Thirty-six questionnaires (response rate of 55% 36/65) were included in this study.

5.3.2 Demographics of participants

Thirty-six participants (14 males and 22 females) from two countries (UK and Italy) completed the short questionnaire. Twenty-four participants (67%) were from Italy and twelve (33%) from UK. Thirteen participants (36%) classified themselves as neurologists, twelve (34%) as radiologists or neuroradiologists, and eight (22%) as emergency medicine physicians. Three participants (8%) were from “other” medical specialities (e.g. neuropsychiatry). Overall the participants spent a median of 10% of their clinical time providing stroke patients care; and a median of 25% of time in research activities. About two-third of the participants (25/35) were experienced clinicians with ten or more years of clinical practice whilst 10 out of 35 participants were less experienced clinicians or novices with nine or less years of clinical practice. One participant did not provide this information. On average the years of clinical practice were similar between neurologists (mean 14.6 years) and radiologists (mean 12.7 years). Overall the participants were not particularly familiar

with the seven papers on ischaemic stroke included in our pilot review. The mean number of papers known by the participants was two (median = 1).

Figure 5.1. Mean number of participants' correct answers to the six multiple-choice questions for the 1st summary document and the 2nd summary document. (Interval plots show mean values and 95% CIs for the mean)



5.3.3 Summary documents: Abstract versus SoF table

All 36 participants completed the six multiple-choice questions of the second part of the questionnaire, which aimed at evaluating the clinicians' ability to interpret the main findings of our systematic review presented in summary format (25 participants completed the Abstract first and 11 participants completed the SoF table first). As participants were presented with both documents (Abstract and SoF) in alternate order and asked to reply to the six multiple-choice questions after each document, I initially assessed whether the order of presentation of the two documents might have affected the number of correct answers. Figure 5.1 shows the interval plots (with

95% CIs) of the correct answers to the multiple-choice questions after reading the first summary document versus the correct answers after reading the second summary document.

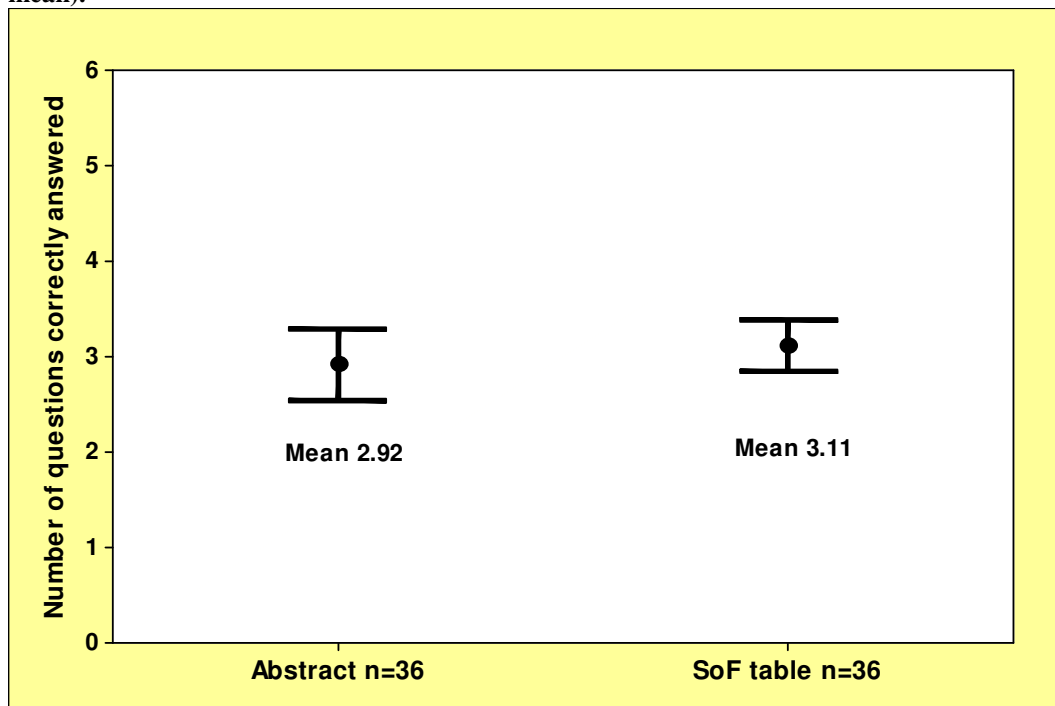
No statistically significant differences were detected between the two groups ($p=0.377$), indicating that a substantial order effect was not present. I therefore decided to compare all participants' answers to the Abstract and all participants' answers to the SoF table independently of their order of presentation. Table 5.1 shows the differences between participants' correct answers to the Abstract and SoF table. Most participants (17/36, 47%) provided the same number of correct answers to both documents. Twelve participants (33%) had a better performance at the SoF table compared with the Abstract and, on the contrary, seven participants (20%) had a better performance at the Abstract compare with the SoF table. The mean difference in score between Abstract and SoF table was not statistically significant (Abstract - SoF table mean was -0.194 CIs -0.506 to 0.117; one sample t -test $p=0.213$). Figure 5.2 illustrates that, overall, the number of correct answers for both documents was low with no significant differences between the number of correct answers to the Abstract (mean 2.92 - 95% CIs 2.54 to 3.29) and the number of correct answers to the SoF table (mean 3.1 - 95% CIs 2.92 to 3.38).

Table 5.1. Difference in scores between Abstract and SoF table results for all 36 participants

Difference in scores	Number of participants (%)
- 2	3 (8)
- 1	9 (25)
0	17 (47)
1	6 (17)
2	1 (3)

Note: Negative scores indicate a better performance at the SoF table; “0” indicates that the number of participants’ correct answers’ was the same for the Abstract and the SoF Table; positive scores indicate a better performance at the Abstract.

Figure 5.2. Mean number of participants’ correct answers to the six multiple-choice questions for the Abstract and the SoF table. (Interval plots show mean values and 95% CIs for the mean).



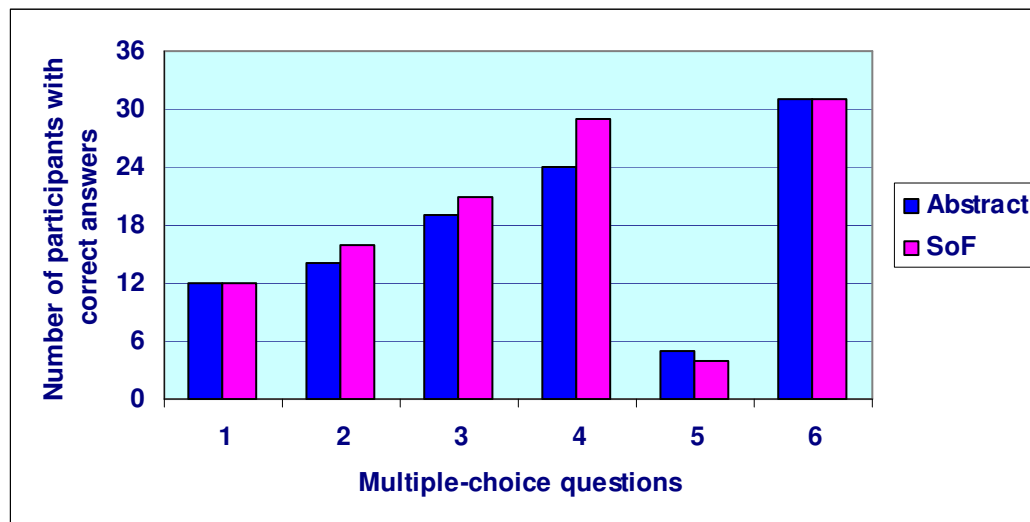
The distribution of the correct answers for the six multiple-choice questions for both summary documents amongst the 36 participants is shown in Figure 5.3. For each question the number of correct answers for the Abstract and the SoF table was similar even though the information contained in the two documents was slightly different. The findings of questions 2 and 3 were quite unexpected as these questions

were related specifically to the use of CT and MRI for detection of mild strokes and stroke mimics. As the information on mild strokes and stroke mimics was reported to some extent in the SoF table but not in the Abstract, I expected to detect a difference in the number of participants' correct answers between the two summary documents. The low rate of correct answers to question 5, which was related to the use in clinical practice of the two imaging tests, CT and MRI, for detection of early ischaemic lesions, was probably due a certain level of ambiguity of this specific question. Most of participants opted for a neutral answer (i.e. "evidence provided is not enough to inform clinical practice") instead of choosing correctly that "either CT or MRI should be used for the early detection of ischaemic lesions".

For both summary documents the average number of correct answers provided by more experienced clinicians (with 10 or more years of clinical practice) was similar to that provided by novices (with 9 or less years of clinical practice) - see Table 5.2. Similarly, the participants' degree of confidence in understanding measures of diagnostic accuracy did not result in better answers. Table 5.3 shows the mean number of correct answers to the Abstract and the SoF table according to the participants' degree of confidence in understanding measures of diagnostic accuracy (see Figure 5.4). For the SoF table - but not for the Abstract - the number of correct answers increased according to the participants' degree of confidence. Given the small number of participants in each group for the sake of the analyses I have collapsed Likert scale points 0 and 1, and Likert scale points 2 and 3. There was not significant difference ($p = 0.302$) between participants who declared themselves to have confidence in understanding measures of diagnostic accuracy (score 2 and 3 at

the Likert scale - mean 3.22 SD 0.74) and those who felt less confident (score 0 and 1 at the Likert scale - mean 2.92 SD 0.86).

Figure 5.3. Distribution of the participants' correct answers to the six multiple-choice questions



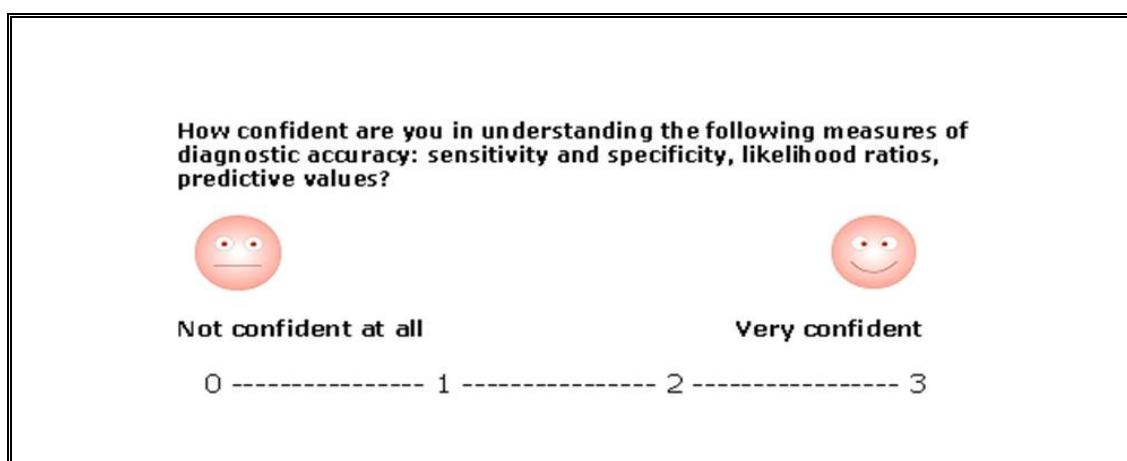
Note: Height of bars corresponds to the number of participants who answered that question correctly after reading either the Abstract or the SoF table

Table 5.2. Mean number of correct answers at the six multiple-choice questions for the Abstract and the SoF according to clinicians' years of clinical practice

	Mean score (SD)	Comparison between Experts and Novices for Abstract and SoF table results (<i>P</i>)
Abstract Experts (no. 25)	2.92 (1.13)	p = 0.956
Abstract Novices (no. 10)	2.90 (1.10)	
SoF Experts (no. 25)	3.08 (0.80)	p = 0.681
SoF Novices (no. 10)	3.20 (0.79)	

Note: Experts = 10 or more years of clinical practice; Novices = 9 or less years of clinical practice. SD: standard deviation. This analysis is based on a total of 35 participants who provided information on years of clinical practice.

Figure 5.4. Likert scale assessing the participants' degree of confidence in understanding measures of diagnostic accuracy



Note: Participants who scored 0 or 1 were considered 'less confident' and participants who scored 2 or 3 were considered 'confident'.

Table 5.3. Participants' mean number of correct answers at the Abstract and SoF table according to their degree of confidence in understanding measures of diagnostic accuracy

Likert score (0-3)	Number of participants	Abstract mean score (SD) to the six multiple-choice questions	SoF mean score (SD) to the six multiple-choice questions
0	3	2.00 (1.00)	2.33 (0.58)
1	10	3.20 (1.03)	3.10 (0.88)
2	14	2.86 (1.29)	3.14 (0.77)
3	9	3.00 (0.87)	3.33 (0.71)

SD: standard deviation

5.3.4 Clinicians' preference: Abstract versus SoF table

Most of the participants regarded the information reported in both the Abstract and the SoF table as clearly presented (69% versus 75%). Figure 5.5 shows that a slightly higher percentage of participants considered the SoF table more informative than the Abstract (55% versus 39%). This preference was clearly confirmed by the answers to

the multiple-choice question that aimed at assessing whether the clinicians judged the information contained in the SoF table more useful than that contained in the Abstract (see Figure 5.6). The participants did not show a clear preference as regards to the way diagnostic results should be phrased. Twenty out of the 36 participants favoured a more verbal description of results, whilst 14 opted for a more simple numerical description of results. Some participants provided ‘free’ comments on the way the Abstract and the SoF table could have been improved. For the Abstract some participants suggested adding more information on: i) limitations and biases of the included studies; ii) spectrum of patients; iii) applicability of results; and including a graphical representation of results. For the SoF table the comments included: i) explanation of findings in bullet points; ii) a legend for the smiley faces related to the grading of evidence; iii) a less ‘busy’ format; iv) no forest plot of results of included studies, as they were too difficult to interpret.

Figure 5.5. Participants’ answers on which summary document they considered more informative.

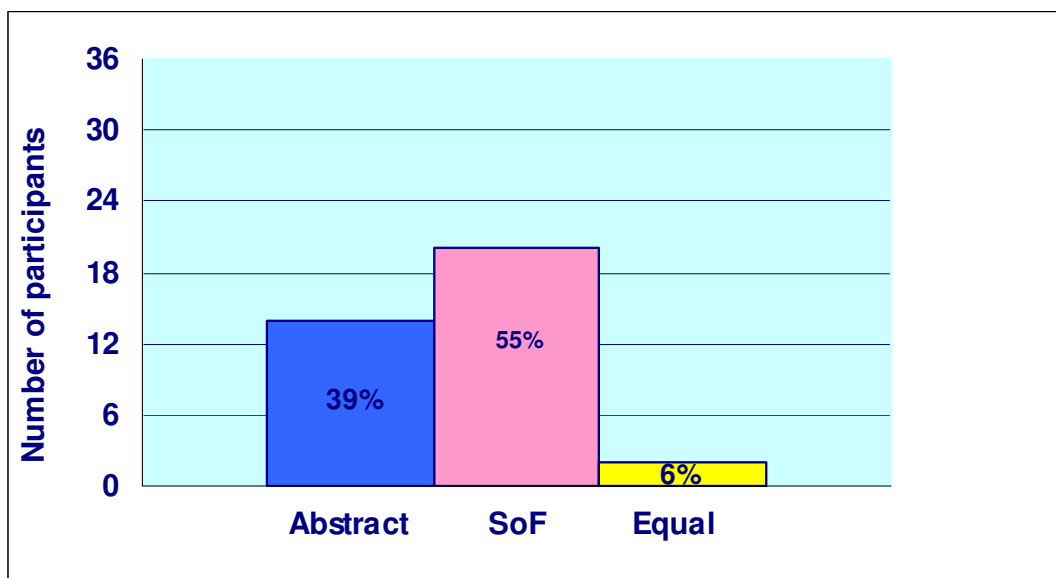
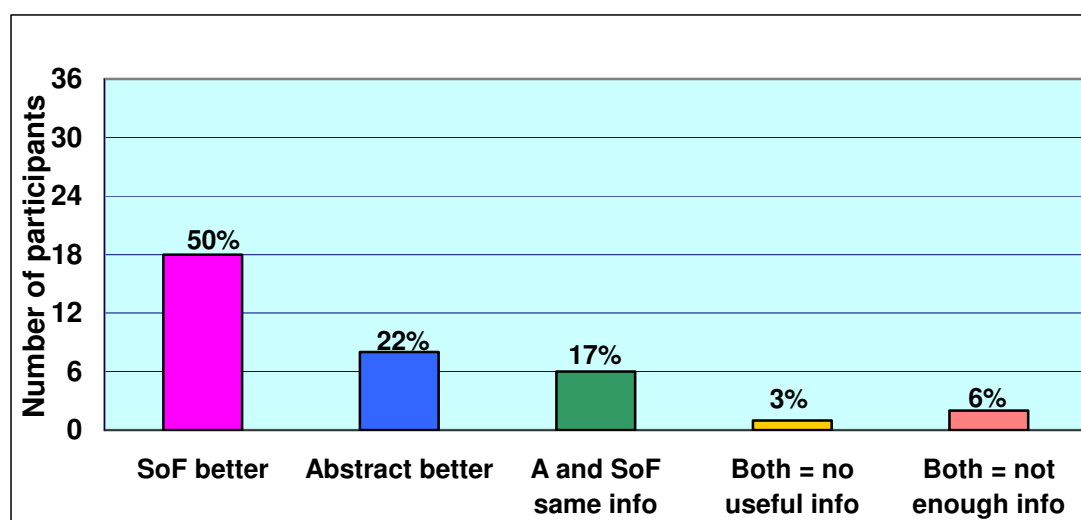


Figure 5.6. Participants' answers to the multiple-choice question which compared the usefulness of the information contained in the Abstract and in the SoF table.



5.4 Discussion

This is the first study that has tried to assess i) the ability of clinicians to interpret the findings of a Cochrane systematic review of diagnostic accuracy presented in two summary documents: a conventional Abstract and a Summary of Findings table (SoF), as well as ii) the clinicians' preference for the format of presentation of the two summary documents. The SoF table is a new tool for summarising information that the Cochrane Collaboration now encourages authors to include in systematic reviews published in the Cochrane Library. For systematic reviews of interventions, the general template for a SoF table is reasonably well developed but difficult to implement since it requires the use of a separate and complex software package (GradePro).¹¹ Moreover, the general template of SoF tables for systematic reviews of interventions has not yet been tested amongst clinicians and there are as yet no empirical data on whether the clinicians find it easy to interpret the information contained in the SoF tables and whether this information has any impact on their

clinical decisions (Phil Wiffen personal communication, 15th Annual Meeting of the UK & Ireland Based Contributors to the Cochrane Collaboration, March 2010). The format of SoF tables for systematic reviews of diagnostic accuracy within the Cochrane Collaboration has not yet been agreed and at present GRADEPro is not designed to handle this type of reviews. In preparing my Cochrane systematic review of diagnostic test accuracy,¹⁴ I developed, in collaboration with lead clinicians, my own SoF table using a simple tabular format (as described in the Methods section). This study is an attempt to determine which form of summary (conventional Abstract or SoF table) best captures and communicates the findings and implications of a systematic review of diagnostic accuracy; which is easier to understand, and displays the best layout.

The findings of this study indicate that the number of correct answers was low for both summary documents (the means were 2.92 for the Abstract and 3.11 for the SoF table respectively) and the mean difference between the participants' correct answers to the Abstract and SoF table was not statistically significant. There are a number of possible reasons which might explain the participants' low rate of correct answers.

- i) The summary documents might have lacked clarity, failing to communicate the relevant information in a transparent and easy way. However, both documents were vetted by experienced clinicians.
- ii) It is possible that the six multiple choice questions used to evaluate the ability of clinicians to interpret the information contained in the two summary documents failed to present concise interpretations of the review findings or were somehow

ambiguous. However, I took several steps to ensure that the questions were clear, unambiguous and formulated and structured to reflect the main findings of my pilot review. Firstly, the questions were developed in discussion with two experts, and revised in the light of subsequent discussions. They were then tested in a group of knowledgeable clinicians outside the study cohort. Of course, such precautions cannot ensure complete clarity and a certain level of ambiguity among questions is, however, a common problem when designing a questionnaire. It is always difficult to anticipate the participants' attitude towards even the most carefully worded questions. The correct answers to the six multiple choice questions were decided, in discussion with the same two experts, on the basis of the main results, conclusions, and implications of my pilot test accuracy review. Even though this choice involved an inevitable degree of arbitrariness, I considered it the best possible way to gather information on how clinicians interpret findings of systematic reviews of test accuracy presented in summary documents.

iii) Notwithstanding the care taken, participants might have found some questions open to interpretation and opted for neutral answers. This could explain, for example, the very low response rate to question 5 in both summary documents (about 56% percent of participants opted for the neutral answer). However, this did not apply to other questions and in particular to question 1, where instead a neutral answer was considered correct.

iv) Clinicians could have skimmed through Abstracts and SoF tables rather than read their contents meticulously, and therefore providing quick, incorrect answers – especially in a testing situation.

v) Clinicians could have answered according to their accrued clinical knowledge and pre-existing preference for a particular type of imaging (CT or MRI) irrespective of the information contained in the summary documents. Interestingly, the mean number of correct answers was similar between Abstract and SoF table and also the distribution of the participants' correct answers to each of the six multiple-choice questions was similar for both documents (see Figures 5.2 and 5.3). This outcome was unexpected as the two summary documents contain different information: in the SoF table the limitations of the included studies and information on the applicability of results in clinical practice are more clearly reported than in the Abstract (for example it is possible to derive information on mild strokes and stroke mimics). In other words I had reasons to expect a higher number of correct answers after presentation of the SoF table. The lack of difference in the answers to the Abstract and to the SoF, despite the dissimilarities between the two documents, might support the hypothesis that clinicians tend to skim quickly through summary documents and provide biased answers based on their pre-existing preferences and clinical experience.

There was no consistent pattern in any of the clinicians' demographic variables to predict their response to either document. The number of correct answers for both documents was similar between experienced clinicians, with more years of clinical

practice, and novices (see Table 5.2). Similarly, for both documents the clinicians' confidence in understanding measures of diagnostic accuracy did not result in a better interpretation of the review findings (see Table 5.3).

Overall, the clinicians preferred the SoF table format and found the information contained in the table more useful than that contained in the Abstract. The summary of findings table differed from the abstract mainly due to the information on the limitations of the included studies and applicability of tests results in clinical practice. This information is likely to be particularly useful to clinicians for deciding about the validity and generalizability of findings and about the use of imaging tests in clinical practice. This assumption is also confirmed by the opinion of the experienced clinicians who participated in the pilot phase of this study and underlined the importance of reporting information on the main limitations of the included studies and on applicability of imaging tests in clinical practice in the summary of findings table. Furthermore, it is worth noting that when participants were asked to comment on which information they thought necessary to add to the abstract and SoF table, they stated specifically that the abstract should contain more explicit details on the limitations of included studies and applicability of results.

The clinicians' preference for the format of the SoF table, however, did not translate into a better performance at the multiple choice questions after its presentation.

5.4.1 Limitations of the study

Only about half of the clinicians I recruited (55%), returned a completed questionnaire. This participation rate might indicate a moderately sized and

potentially selected study sample. This modest response rate may also reflect the difficulty of the task as well as the limited resources available. If future research is to be done in this field, response rate might well improve if, for example, the testing sessions provided more interactive one-to-one support for participating clinicians, or if more advanced ways for testing participants are implemented (e.g. electronic/internet randomised survey). I did not have the resources to determine whether non-responders were naturally different from responders. However, if response is a measure of willingness to participate and invest energy in attempting to understand diagnostic test accuracy reviews, it is plausible that non-responders were less enthusiastic or confident about the task and likely to have a lesser level of understanding of the correct answers and therefore of diagnostic data. The number of females and males and the years of clinical practice were not evenly distributed among participants. However, I am not aware of any evidence in the literature indicating that these participants' characteristics are associated with a better knowledge of diagnostic test accuracy concepts and a better application of these concepts in clinical practice. Both summary documents were administered to clinicians in alternate order in the same testing session. In theory, to avoid a possible 'contamination' effect in the answers provided by the clinicians to each document, I should have adopted a randomised or a cross-over design with a washout period. A randomised study would have, however, required a much larger number of clinicians ideally stratified by clinical speciality and years of clinical practice. A cross-over design in which clinicians receive one summary document in one testing session and the other summary document in a different session after a washout period was somehow difficult to organise as clinicians are not particularly keen to fill in

questionnaires and I was concerned about the possible attrition rate in the second part of the cross-over. Moreover, presentation of both summary documents in the same testing session was justified by the fact that one of the objectives of this study was to ascertain participants' preference for the format and content of the two documents, and therefore allow them to scrutinize both documents at the same time.

5.4.2 Implications for research and for the Cochrane Collaboration

Diagnostic results are not easy to interpret. Researchers and health professionals need to be aware that summary information from systematic reviews of diagnostic accuracy, presented either in a traditional abstract format or in a simple tabular format (summary of findings table), pose some problems to clinicians who may draw simplistic or erroneous conclusions. This survey involved mainly neurologists and neuroradiologists. Overall, the main relevant finding is a strong qualitative impression that these clinicians found the task of interpreting results of systematic reviews of diagnostic test accuracy difficult, especially if under pressure of time. It is likely that this would be the case for clinicians working in other medical disciplines, but this does require further research. Further research is needed to decide what is the best way to communicate the findings of systematic reviews of test accuracy to clinicians and what is the best format for summarising such findings. In particular future studies might consider the following points:

- Simplify the format of the Summary of Findings table further;
- Develop alternative graphical approaches to presenting data and test results which are most easily understood;

- Pilot the questions to be used with questionnaire more extensively to reduce ambiguity in wording;
- Consider alternative methods for testing clinicians and health professionals (e.g. head-to-head interviews; electronic survey);
- Present the summary documents to a wider range of health professionals with different grades of seniority.

References

1. Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *Am J Med* 1998; 104: 374-80
2. Steurer J, Fischer JE, Bachmann LM, Koller M, ter Riet G. Communicating accuracy of tests to general practitioners: a controlled study. *BMJ* 2002; 324:824-6
3. Irwig L, Tosteson AN, Gatsonis CA, Lau J, Colditz G, Chalmers TC et al. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med* 1994; 120:667-76
4. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytical methods for diagnostic test accuracy. *J Clin Epidemiol* 1995; 48: 119-30
5. Deeks JJ. Systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001; 323: 157-62
6. Leeflang MMG, Deeks JJ, Gatsonis CA, Bossuyt PMM, on behalf of the Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. *Annals of Internal Medicine* 2008; 149: 889-97

7. Bossuyt PMM, Reitsma JB, Bruns DE, Gatsonins CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HCW for the STARD Group. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Radiology* 2003; 226: 24-8
8. Smidt N, Rutjes AWS, van der Windt DAWM, Ostelo RWJG, Bossuyt PMM, Reitsma JB, Bouter LM, de Vet HCW. The quality of diagnostic accuracy studies since the STARD statement. Has it improved? *Neurology* 2006; 67: 792-7
9. Leeftang MM, Debets-Ossenkopp YJ, Visser CE, Sholten RJ, Hooft L, Bijlmer HA, et al. Galactomannan detection for invasive aspergillosis in immunocompromized patients. *Cochrane Database of Systematic Reviews* 2008, Issue 4. Art. No.: CD007394. DOI: 10.1002/14651858.CD007394
10. Pitkin RM, Branagan MA, Burmeister LF. Accuracy of data in abstracts of published research articles. *JAMA* 1999; 281:1110-1
11. Schunemann HJ, Oxman AD, Higgins JPT, Vist GE, Glasziou P, Guyatt GH on behalf of the Cochrane Applicability and Recommendations Methods Group and the Cochrane Statistical Methods Group. Presenting results and 'Summary of Findings' tables. In: Higgins JPT and Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester (UK): John Wiley & Sons, 2008

12. GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004; 328: 1-8

13. GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; 336: 924

14. Brazzelli M, Sandercock PAG, Chappell FM, Celani MG, Righetti E, Arestis N, Wardlaw JM, Deeks JJ. Magnetic resonance imaging versus computed tomography for detection of acute vascular lesions in patients presenting with stroke symptoms. *Cochrane Database of Systematic Reviews* 2009, Issue 4. Art. No.: CD007424. DOI: 10.1002/14651858.CD007424.pub2.

6. General Discussion

This Chapter seeks to draw together the implications of this methodological thesis for all stakeholders in future diagnostic test accuracy reviews. By stakeholders I mean authors, journal editors and publishers, peer reviewers, and the wide variety of health professionals who will need to apply the results of systematic reviews of test accuracy to their clinical practice.

6.1 Methodological relevance of this thesis

The methods for performing systematic reviews of diagnostic test accuracy studies have only recently been introduced and are still undergoing development. This makes performing such reviews rather challenging. In the research work carried out to prepare this thesis, I have addressed four specific methodological areas related to assembling systematic reviews of diagnostic test accuracy in stroke medicine:

- i) The quality of reporting of studies of diagnostic test accuracy,
- ii) The extent and direction of publication bias in studies of diagnostic test accuracy,
- iii) The optimal methods for preparing reviews of test accuracy, (by undertaking a pilot systematic review according to the draft recommendations of the Cochrane Diagnostic Test Accuracy Working Group)

- iv) The best way to summarise findings of Cochrane reviews of diagnostic test accuracy for clinicians and policy makers.

6.1.1 Evidence of reporting bias

The purpose of systematic reviews of diagnostic accuracy is to inform decisions about the use of diagnostic tests in clinical practice. To fulfill this end, systematic reviews of diagnostic accuracy need to provide valid and reliable results. The major threats to the validity of a systematic review of test accuracy arise from the quality of the individual studies included in the review. It is a recognized problem that the quality of reporting of diagnostic accuracy studies is less than satisfactory. For randomised controlled trials, the criteria for deciding what constitutes a high quality study are well established. By contrast, in studies of diagnostic accuracy, authors have greater freedom to make discretionary choices¹ (such as the decision about what constitutes true positive and true negative test results, about the referral pattern of patients, or about which characteristics patients should possess for being enrolled in the study). Therefore, the quality of reporting is complementary to the evaluation of study methodology and conduct. There is some evidence in the literature that poor methodological quality of diagnostic studies may result in an overestimation of test accuracy²⁻⁴ which, in theory, could lead to an association between poor quality of reporting and overestimation of test performance. Despite the STARD initiative and the publication of the STARD guidelines for reporting results of studies of diagnostic accuracy, I found that the design and conduct of studies of diagnostic accuracy varied considerably in the stroke literature, indicating that the editorial control on

what constitutes a high-quality diagnostic study is perhaps weak. Lack of methodological rigour in the conduct of diagnostic accuracy studies is a cause of concern, as it may jeopardize the effective application of diagnostic tests in clinical practice. This is particularly worrying if we consider that current recommendations on the use of imaging tests for the diagnosis of stroke may be based on limited and poorly reported evidence. Reporting of study characteristics and results is important for several reasons, since it helps to:

- i) make the study design and conduct transparent and reproducible;
- ii) explore potential sources of heterogeneity across studies in meta-analyses;
- iii) improve our understanding of the relationship between quality of reporting and methodological quality;
- iv) allow replication of study findings and promote further research.

The findings of this thesis indicate that more efforts are needed to improve the accuracy of reporting and the methodological quality of diagnostic research. Implementation of available guidelines such as the STARD and the QUADAS checklists should facilitate this process. A collaborative involvement of journal editors, publishers, peer reviewers, professional organizations, and Cochrane Groups could promote adherence to current guidelines and ensure transparency in diagnostic research. Reviews authors should receive strict and precise guidance on which criteria individual studies of diagnostic test accuracy should comply with. Efforts to ensure study reports are as complete as possible should be considered an ethical principle in medical research. In particular, journal editors and publishers should i)

explicitly require authors to adopt the current available guidelines when designing and reporting their study and, more importantly, ii) should not accept manuscripts reporting diagnostic studies unless they clearly comply with current guidelines.

6.1.2 Evidence of publication bias

Another methodological aspect I have considered in this thesis is the occurrence of publication bias in diagnostic accuracy reviews. There is little empirical evidence in the literature on the frequency and characteristics of publication bias in systematic reviews of studies of diagnostic test accuracy. I carried out a study to capture the extent of this problem by evaluating the proportion of abstracts of diagnostic studies presented at two major international stroke conferences which were later published in full. This study was the first to focus on the time interval between abstract submission and full publication and on the study characteristics associated with publication. I did not find evidence of substantial bias in the publication processes which occurred after abstract acceptance. However, it would be interesting to ascertain whether publication biases are more likely to occur at the level of abstract submission or selection. Apart from inter-observer agreement between test readers, I did not find other methodological aspects or study characteristics clearly associated to full publication of diagnostic studies in stroke medicine. In particular, I did not detect a 'positive findings' bias across the diagnostic abstracts identified. It is worth observing that the majority of identified abstracts generally reported high estimates of diagnostic accuracy. It is plausible that authors of studies with less favorable results would be less willing to submit their work or that reviewers who selected

abstracts for presentation at international stroke meetings would be more inclined to accept those with high estimates of accuracy. In diagnostic research it is probably more likely that investigators who do not get positive test results would not pursue a publication - the test is not good, it is clinically useless, hence it is not worth further work or does not attract any sponsors. Research on publication bias of randomised studies is facilitated by the existing ethical approval system and the registration constraints related to the conduct of clinical trials in human subjects. Many journals would not now publish reports of trials that were not registered on a trial registry at inception. Similar constraints are not yet in place for studies of diagnostic test accuracy. I believe there is a case for the creation of a future prospective register for studies of diagnostic test accuracy, which would allow better evaluation of publication bias as well as help to enforce adherence to current guidelines with regard to quality of reporting and methodological standards - although it may not be suitable for studies which are undertaken using a retrospective design. Hemingway and colleagues have recently suggested that all human research studies should have a study protocol detailing the methods of data collection and analysis, and that these protocols should be registered in advance.⁵ Registration of protocols prior to the conduct of studies of diagnostic test accuracy could be an effective way to improve quality and reduce the risk of potential methodological biases such as spectrum bias and publication bias. Research on the extent and direction of publication bias for diagnostic test accuracy is still at a very early stage. Further empirical evidence is needed on the existence, magnitude and impact of publication bias in diagnostic research. While waiting for further evidence, the potential existence of publication

bias should always be considered when preparing systematic reviews of diagnostic test accuracy and all relevant sources of literature should be thoroughly searched.

6.1.3 Evidence of methodological biases in systematic reviews of test accuracy

Variability between studies included in systematic reviews of diagnostic test accuracy may be larger than that observed between studies included in systematic reviews of interventions. In particular, I found a huge variability between studies included in the Cochrane pilot review reported in Chapter 4 in the selection of the study population and the methods used for the study. The most important sources of bias I detected were spectrum bias, selection bias, and incorporation bias. Most of the studies included patients with a high pre-test probability of stroke or an already established diagnosis of stroke (especially for studies with retrospective data collection where patients were selected on the basis of their final diagnosis and if they had received the imaging tests under investigation - MRI and CT). Moreover, in some studies patients were preselected on the basis of the absence of haemorrhagic brain lesions and the presence of mild symptoms. These spectrum and selection biases are likely to have inflated the sensitivity estimates and therefore affected the overall reliability of the review findings. In some studies the results of the imaging tests under investigation were 'incorporated' in the final diagnosis and hence they may have further increased the estimates of accuracy. I was surprised by the generally low quality of the studies included in my systematic review. Brain imaging is essential to distinguish between the different pathological causes of stroke and to

determine the brain territory affected by the vascular lesion, since the management of an individual patient will depend heavily upon accurate diagnosis of these aspects of his stroke. The continuous advances in diagnostic imaging technology can often make reliable large-scale studies assessing a particular version of a specific imaging test difficult to complete before it is implemented in clinical practice (or undergoes further development). However, assessing the accuracy and cost-effectiveness of these imaging technologies to determine their applicability in clinical practice is fundamental, in the light of the current constraints on health care resources and the need to optimize medical interventions. To achieve this goal diagnostic accuracy and cost-effectiveness studies should be conducted according to current guidelines, and evaluate broader and unselected patient populations. As already stated, I am convinced that journal editors and peer reviewers, as well as professional organizations have a great responsibility to set quality standards of diagnostic research in stroke medicine. Nowadays, there are many general problems affecting the quality of research. As Sørensen pointed out in a recent editorial: “among the many factors that influence research quality throughout the spectrum of biomedical studies are the consistency and quality of training, the vagaries of research funding, the independence of investigators, and the adequacy of peer review. Also pervasive is a system of academic promotion influenced more by the number of publications than by quality, which increases the demand for output of whatever quality.” The results of my study clearly endorse his view that a “continuing education efforts focused on journal editors might improve the quality of published research faster than any other intervention”.⁶

6.1.4 Complex issues in systematic reviews of diagnostic test

accuracy

Systematic reviews of diagnostic test accuracy are more complex, challenging, and time consuming than systematic reviews of randomised clinical trial. In conducting the Cochrane pilot review reported in Chapter 4, using the methods developed by the Cochrane Screening and Diagnostic tests Working Group, I had to consider three important methodological aspects: the searching for evidence; the assessment of individual studies of test accuracy, and the synthesis of results. Firstly, to develop a search strategy to identify relevant studies in the literature was not easy. At present, in the major electronic bibliographic databases there are no unequivocal indexing terms to identify studies of diagnostic accuracy with high sensitivity and specificity. I developed and refined the search strategies following the recommendations contained in the draft Chapter 7 “Searching for Studies” of the Handbook that the Screening and Diagnostic Working Group is currently preparing (available online at <http://srdata.cochrane.org/handbook-dta-reviews>). I used both database subject headings and text words to search major electronic databases for the imaging tests under investigation and for ‘stroke’ as clinical condition. Although the use of a search filter is usually discouraged because it could reduce the overall sensitivity of the search,⁷ I deemed it necessary to include a methodological search filter to identify diagnostic accuracy studies of imaging in the stroke literature. A search which had combined the keyword ‘stroke’ with terms for the imaging tests under investigation would have retrieved an unmanageably large number of citations (i.e. it would have high sensitivity but unacceptably low specificity). This is probably the case for many other clinical conditions assessed by imaging. Therefore, I opted for

the inclusion of a recent validated filter with optimal sensitivity and specificity for the retrieval of imaging studies.⁸ This methodological filter became available in the literature in 2008 and therefore was not included amongst those considered by Leeflang and colleagues in their study assessing the usefulness of methodological filters in search strategies developed to identify diagnostic studies to be included in systematic reviews, which was published in 2006.⁷ In the light of my experience in preparing the pilot systematic review for this thesis, I would suggest the use of such a filter for the preparation of further diagnostic test accuracy reviews which focus specifically on imaging studies.

Secondly, the critical assessment of individual studies selected for inclusion in the pilot review was difficult due to the recognised problem of the poor reporting and poor methodological quality of diagnostic studies in the literature, as I have already discussed in previous sections.

Thirdly, summarizing and synthesizing results of studies of diagnostic accuracy can be challenging. There are a range of approaches to the meta-analysis of studies of diagnostic test accuracy but all involve complex statistical modelling. So the process of selecting the most appropriate approach and then undertaking the statistical modelling is far more complex and difficult to perform and interpret than that of systematic reviews of randomised controlled trials. The decision about whether or not it is legitimate to pool data from studies of diagnostic accuracy is again complicated by the poor reporting and methodological standards of the diagnostic literature. In my pilot review I included only comparative studies where both imaging tests were performed on the same stroke patients against a reference

standard of clinical diagnosis and imaging follow-up. The recommended method for combining data from comparative studies, which takes into account the underlying relationship between sensitivity and specificity is the SROC curve. The hierarchical and bivariate methods are considered the most robust for fitting SROC curves.⁹⁻¹⁰ However, these models rely heavily on statistical modelling (and on modelling assumptions that may be difficult to verify) and require sophisticated programmes to be fitted (they cannot be fitted using the Cochrane software RevMan 5). The use and interpretation of both the hierarchical and bivariate SROC models in systematic reviews of diagnostic accuracy is far from straightforward and can be problematic.¹¹ These models do not work well when the number of studies is small and when there is little heterogeneity amongst estimates of accuracy across studies. In my pilot review, I was not able to compare the diagnostic performance of the imaging tests under investigation using these models, because there was too little variation in the true positive rates and false positive rates across the different studies. Experience with the hierarchical SROC methods is still indeed very limited and warrants further evaluation. In particular, further investigations to assess the use of these methods in clinical practice are needed before statisticians and researchers can really understand their strengths and weaknesses.¹² Furthermore, it would be interesting to verify how these methods are used and interpreted by clinicians. SROC curves are not very popular amongst clinicians and health professionals because their underlying concepts and interpretation are quite difficult.¹³

6.1.5 The challenge of summarising findings of systematic reviews of test accuracy for clinicians

Another challenge that systematic reviews of diagnostic test accuracy present is to understand how well clinicians and health professionals interpret their findings and whether these findings have appropriate impact on their clinical decision-making. Knottnerus maintains that, in clinical practice decision making problems often do not depend so much on the lack of research findings but on the lack of a good summary of those findings.¹⁴ For this thesis I studied how clinicians interpreted the information contained in two forms of summary of a Cochrane systematic review of diagnostic test accuracy - a traditional verbal Abstract and a newly developed Summary of Findings table. I did this by means of a survey questionnaire developed *ad hoc* for the purpose of the study. Under these conditions, the clinicians who participated did not do well in interpreting summary results of the systematic review of MRI and CT for the early diagnosis of stroke reported in Chapter 4. The overall number of correct answers was low and they failed to detect the main differences between the two summary documents. My interpretation to explain the observed low rate of correct answers is that clinicians tend to provide answers based on their pre-existing preference and clinical experience rather than on the real contents of the summary documents. This may pose some concerns, if we consider that busy clinicians often rely on information presented in a summary format rather than going through the whole content of a systematic review. Clinicians appeared also to favour the format and the type of information contained in the summary of findings table rather than those of the traditional abstract. I believe that abstracts are often too succinct to picture the characteristics and complexity of systematic reviews and in

particular of systematic reviews of diagnostic test accuracy. Summary of findings tables, which contain information on the main limitations of the included studies as well as applicability of results, may be a more useful summary of review findings. However, it is not yet clear how best to format this information and how it should be displayed and presented to clinicians and health professionals. This deserves further investigation to improve the translation of findings of systematic reviews of test accuracy into better clinical practice and enhanced patients care.

6.2 Implications of this thesis for research

- The quality of reporting of studies of diagnostic accuracy of imaging for the diagnosis of stroke is poor.
- Investigation of publication bias in systematic reviews of diagnostic test accuracy requires further research. It would be useful to assess whether there is evidence of publication bias of diagnostic accuracy studies in other clinical specialities and whether there is evidence of bias in abstract submission or acceptance. Further research on the methods for detecting and dealing with publication bias is clearly needed.
- Important methodological biases are present in diagnostic accuracy studies of imaging. In particular, spectrum and selection biases and incorporation biases may affect validity and generalisability of findings of systematic reviews of test accuracy. There is the need for assessing further the quality of medical imaging studies in the literature.

- Summary information (abstract or summary of findings table) may be erroneously interpreted by clinicians who may process this information according to their pre-existing knowledge and experience. Further investigations on the way to present and communicate summary results of systematic reviews of diagnostic accuracy are needed.

6.3 Implications of this thesis for authors and consumers of diagnostic test accuracy reviews

- Diagnostic test accuracy reviews are methodologically challenging and time consuming
- Searching for diagnostic evidence in the literature is not easy and consultation with an information scientist is recommended. For specific topics (e.g. imaging studies) it is worth considering the use of a validated search filter (to avoid an unmanageable number of references).
- The quality of reporting and the methodological quality of imaging test accuracy studies in the literature is poor and this could compromise the validity and generalisability of systematic review findings.
- The statistical models for summarising diagnostic results are difficult to interpret. Review authors should seek the advice of a statistician with relevant expertise to in the complex models required for the statistical synthesis of diagnostic data. Further evaluation of the hierarchical models required for meta-analysis in clinical settings is needed. Further evidence is needed on how clinicians interpret these sophisticated statistical models.

- Further development is needed to improve presentation and interpretation of systematic reviews of test accuracy. Clinicians may struggle to interpret diagnostic results presented in summary format. Further investigation is needed to study the best way to present and display diagnostic information to clinicians and health professionals.

6.4 Implications of this thesis for journal editors, publishers, professional organizations, peer reviewers

- As the technologies for new diagnostic tests become ever more complex, it is important to establish high quality methodological standards. Journal editors, publishers, professional organizations, and peer reviewers should consider the quality of reporting and the quality of study design in diagnostic research of great importance.
- They should require authors to adhere to current specific guidelines such as the QUADAS and the STARD checklists and they should reject manuscripts which do not comply with these guidelines.
- Journal editors should also consider inviting methodological experts in diagnostic research among the members of their editorial teams.
- To reduce occurrence of publication bias and ensure transparency all primary diagnostic test accuracy studies should be prospectively registered in publicly accessible and easy searchable databases. This register would inform the ‘diagnostic community’ about the most up-to-date ongoing studies. The International Committee of Medical Journal Editors, once the registry is

established, could require that diagnostic accuracy studies be registered at starting as *conditio sine qua non* for being considered for publication.

6.5 Implications of this thesis for the Cochrane Collaboration

The results of this thesis are, in a sense, disappointing, since it did not prove possible with the resources available, to establish clear guidance on several important methodological points. The conduct of the pilot diagnostic test accuracy review reported in Chapter 4 proved very time consuming (the review required an estimated 2000 hours of work). Similarly, the work for each methodological project included in this thesis involved a considerable amount of time (estimated time 1300 hours per project). In other words, this review of diagnostic test accuracy appears to have been unusually resource intensive (based on my own experience of conducting systematic reviews of interventions), though this is necessarily a qualitative conclusion, since there is little data available on the resources that were required for the pilot Cochrane Diagnostic test accuracy reviews conducted in other areas. The work of identifying studies to be included in the review, assessing their methodological quality and obtaining agreement on the data to be extracted was considerable, and compounded by the generally incomplete reporting of studies in publications.

The overall implications of this work are the follows:

Cochrane Groups capacity

Cochrane groups planning to undertake diagnostic test accuracy reviews will need to ensure they have:

- sufficient methodological and statistical expertise to support the conduct of the review;
- sufficient time and capacity to provide editorial support, to help authors deal with:
 - i) complex review methods (e.g. development of comprehensive search strategies; models for statistical analyses)
 - ii) issues for which there is not methodological consensus yet (e.g. summary of findings tables)

Cochrane Groups

The editorial base teams have finite resources, so when prioritising and selecting the topics which are to be registered for diagnostic test accuracy reviews, they should bear in mind the following constraints:

As a result of the complexity and statistical demands of diagnostic test accuracy reviews, each Cochrane Review Group is likely to be able to support only a limited number of such reviews.

Reviews should therefore be restricted to those high priority topics where:

- the diagnosis it is likely to lead to significant changes in patient management and hence
- influence clinical outcome.

Cochrane groups planning to recruit authors to undertake diagnostic test accuracy reviews will need to ensure any potential authors are competent not only in the content area of the medical disorder and the test under evaluation, but also in the methods of systematic reviews and the statistics of diagnostic test accuracy. Again, speaking from my own perspective as an individual with: prior professional experience of the medical disorder (stroke); experience of systematic review methods and statistics, and having very high-level support on medical imaging, the process of overcoming some of the challenges was significant. Whilst the development of the RevMan software and the section of Handbook on diagnostic test accuracy reviews will facilitate the process of such reviews, my qualitative conclusion is that review groups will need to enquire closely about the expertise of potential authors before accepting them. Naïve reviewers might well be easily disheartened at the scale of the task, and inexperienced reviewers might generate unreasonably high demand on the resources of the editorial base team!

Impact of the work of this thesis on the Cochrane Stroke Group

The experience with my pilot diagnostic accuracy review and its related methodological work has led to the Cochrane Stroke Group formulating its editorial policy on diagnostic test accuracy reviews. This policy has already been of benefit to the Stroke Group to enable it to accept one high priority diagnostic test accuracy review and reject one low priority review (Hazel Fraser personal communication, May 2010).

6.6 Implications of this thesis for clinical practice

Although the focus of this thesis was on methodological aspects related to systematic reviews of diagnostic test accuracy, I feel the findings I have found have also some clinical value.

In Chapter 4, I have demonstrated that comparative studies of diagnostic test accuracy assessing MRI and CT for the diagnosis of stroke are hampered by important methodological biases (spectrum and selection bias; incorporation bias). I have presented some evidence that MRI is more sensitive, but not more specific, than CT for the early detection of mild stroke. I found very little evidence on the diagnostic performance of MRI compared to CT for the early detection of haemorrhagic stroke. There is a need for further comparative studies on the cost-effectiveness of MRI and CT for the early diagnosis of stroke. In particular, further studies - in unselected patient populations - are needed to provide clear evidence on whether MRI can be used as the imaging modality of first choice for the majority of stroke patients in routine practice. In Chapter 5, I have showed that the way findings of systematic reviews of diagnostic test accuracy are currently presented may be difficult for clinicians to interpret and implement.

6.7 Future developments

The methodological problems I encountered and underlined in this thesis could be priority topics for specific methodological projects to inform diagnostic research methods (e.g. development of summary of findings tables; assessment of publication bias; using quality assessment of included studies to adjust estimates of diagnostic test accuracy reviews).

References

1. Begg CB. Systematic reviews of diagnostic accuracy studies require study by study examination: first for heterogeneity, and then for sources of heterogeneity. *Journal of Clinical Epidemiology* 2005; 58: 865-6 commentary
2. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JHP, Bossuyt PMM. Empirical evidence of design related bias in studies of diagnostic tests. *JAMA* 1999; 282: 1061-6
3. Rutjes AWS, Reitsma JB, Di Nisio F, Smidt N, van Rijn JC, Bossuyt PMM. Evidence of bias and variation in diagnostic accuracy studies. *Canadian Medical Association Journal* 2006; 174(4) [DOI:10.1503/cmaj.050090]
4. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Annals of Internal Medicine* 2004; 140: 189-202)
5. Hemingway H, Riley RD, Altman DG. Ten steps towards improving prognosis research. *BMJ* 2009; 339: b4184
6. Sørensen HT, Rothman KJ. The prognosis for research. *BMJ* 2010; 340: c703

7. Leeflang MMG, Scholten RJPM, Rutjes AWS, Reitsma JB, Bossuyt PMM.
Use of methodological search filters to identify diagnostic accuracy studies
can lead to the omission of relevant studies. *Journal of Clinical Epidemiology*
2006; 59: 234-40

8. Astin M, Brazzelli M, Fraser C, Counsell C, Needham G, Grimshaw J.
Developing a sensitive search strategy in MEDLINE to retrieve studies on
assessment of the diagnostic performance of imaging techniques. *Radiology*
2008; 247:365-73

9. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis
of diagnostic test accuracy evaluation. *Statistics in Medicine* 2001; 20: 2865-
84

10. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman
AH. Bivariate analysis of sensitivity and specificity produces informative
summary measures in diagnostic reviews. *Journal of Clinical Epidemiology*
2005; 58: 982-90

11. Chappell FM, Raab GM, Wardlaw JM. When are summary ROC curves
appropriate for diagnostic meta-analyses. *Statistics in Medicine* 2009; 28:
2653-68

12. Begg CB. Meta-analysis methods for diagnostic accuracy. *Journal of Clinical Epidemiology* 2008; 61: 1081-2

13. Carrington Reid M, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *Am J Med* 1998; 104: 374-80

14. Knottnerus JA, Van Weel C. General introduction: evaluation of diagnostic procedures. In Knottnerus JA (editor), *The Evidence Base of Clinical Diagnosis*. BMJ Books 2002

Summary

Diagnostic tests are used in clinical practice to help the diagnosis making process and hence improve patients care and management. In stroke medicine effective patient management depends on a rapid and precise diagnosis. Brain imaging with CT and MRI is necessary to identify the exact vascular territory of the brain affected by the lesion and to determine the pathological type of stroke. The demand for evidence on which imaging test is the more accurate for the rapid diagnosis of acute stroke has increased because appropriate use of the recently developed emergency treatments for acute ischaemic stroke (e.g. thrombolysis, neuro-interventional treatment) depend heavily on accurate diagnosis. Systematic reviews of diagnostic test accuracy may therefore help clinicians and health professionals to decide about the best use of diagnostic tests in clinical practice and to choose between alternative tests. The Cochrane Collaboration has recently included systematic reviews of diagnostic test accuracy within its remit. However, to prepare Cochrane systematic reviews of diagnostic test accuracy is still quite challenging because the methods for such reviews are still developing. The research work undertaken for this thesis addresses some relevant methodological issues and contributes to inform the development of the methods for Cochrane systematic reviews of test accuracy.

Chapter 1 considers the importance of using imaging tests for the diagnosis of stroke, introduces some basic concepts of diagnostic test evaluation, and addresses some of the methodological challenges in preparing systematic reviews of diagnostic test accuracy.

The validity and reliability of findings of systematic reviews of diagnostic test accuracy depend on the availability and the quality of the included studies. Methodological characteristics of included studies may be, however, difficult to ascertain if the quality of reporting of such studies is poor.

The aim of the study reported in *Chapter 2* was to evaluate the extent to which studies on magnetic resonance imaging for the diagnosis of stroke published between 1999 and 2008 complied with the current STAndards for the Reporting of Diagnostic accuracy studies (STARD) criteria and to explore whether the introduction of the STARD statement in 2003 has contributed to a better quality of reporting. I identified 18 studies published between 1999 and 2003 and 7 studies published between 2004 and 2008. The findings of this study were limited by the small number of studies available in the literature. However, they were comparable to those of other fields of medicine. There was a wide variation in the quality of reporting of individual STARD items. The mean number of reported STARD items in diagnostic articles published between 2004 and 2008 was not significantly different to that of diagnostic articles published between 1999 and 2003. In particular, diagnostic articles pre- and post-STARD lacked several important items: a description of the study population and demographic characteristics of the patient population; cross tabulation of results of the index test by the results of the reference standard; a description of how indeterminate results were handled; a description of diagnoses in patients without the target condition; and a description of any adverse events related to performing the index test or the reference standard. Overall, in stroke medicine diagnostic accuracy articles on brain imaging were not uniformly reported and there is still ample room

for further improvement. I therefore recommend a broader use of the STARD criteria by authors, journal editors, and professional bodies to improve the standards of diagnostic research.

Chapter 3 focuses on a review of the empirical evidence of publication bias in studies of diagnostic test accuracy in stroke medicine. I evaluated the proportion of abstracts on diagnostic tests presented at international stroke meetings that were later published in full. I investigated which study features were associated with full publication. Of the 160 identified diagnostic abstracts, 76% were subsequently published in full. Sixty-two percent were published in full within 24 months of presentation. The only factor that predicted full publication was whether or not inter-observer agreement between test readers had been assessed. No other studies features (including the clinical utility of results, the country of origin of the corresponding author, the multi-centre status, or Youden's Index) was associated with subsequent full publication. I found no clear evidence of bias in the processes of publication that occur after abstract acceptance. I was, however, unable to assess bias in the process of abstract submission or acceptance. Overall, diagnostic abstracts fail to report relevant methodological aspects.

In *Chapter 4* I systematically compared the diagnostic accuracy of diffusion-weighted magnetic resonance imaging and computed tomography (CT) for acute ischaemic stroke, and estimated the diagnostic accuracy of magnetic resonance imaging (MRI) for acute haemorrhagic stroke. MRI is increasingly used for the diagnosis of acute ischaemic stroke but its sensitivity for the early detection of

intracerebral haemorrhage has been debated. CT is extensively used in the clinical management of acute stroke, especially for the rapid exclusion of intracerebral haemorrhage. Only eight studies met my inclusion criteria. Seven studies contributed to the assessment of ischaemic stroke and two studies to the assessment of haemorrhagic stroke. The spectrum of patients was relatively narrow in all studies, sample sizes were small, there was substantial incorporation bias, and blinding procedures were often incomplete. Among the patients subsequently confirmed to have acute ischaemic stroke (161/226), the summary estimates for DWI were: sensitivity 0.99 (95% CI 0.23 to 1.00), specificity 0.92 (95% CI 0.83 to 0.97). The summary estimates for CT were: sensitivity 0.39 (95% CI 0.16 to 0.69), specificity 1.00 (95% CI 0.94 to 1.00). The two studies on haemorrhagic stroke reported high estimates for diffusion-weighted and gradient-echo sequences but had inconsistent reference standards. I did not calculate overall estimates for these two studies. I was not able to assess practicality or cost-effectiveness issues. MRI had a better sensitivity than CT for the early detection of ischaemic stroke in highly selected patients. However, the variability in the quality of included studies and the presence of methodological biases rendered the reliability and generalisability of observed results questionable. Further well-designed studies without methodological biases, in more representative patient samples, with practicality and cost estimates are required to determine which patients should undergo MRI and which CT in suspected acute stroke.

Chapter 5 addressed the issue on how to summarise findings of systematic reviews of diagnostic test accuracy. At present, there is very little evidence on the best way of presenting summary results of diagnostic accuracy reviews to clinicians and on how clinicians interpret results of such reviews. I carried out a study to evaluate the usefulness to clinicians of two forms of summaries (a traditional abstract and a new developed summary of findings table) of a systematic review of diagnostic test accuracy. I developed a questionnaire to ascertain the clinicians' ability to interpret the information contained in either documents and to assess their preference about the format of the two summaries. Overall, the number of correct answers to both documents was low with no significant differences between the correct answers at the abstract and the correct answers at the summary of findings table. In particular, clinicians did not differentiate their answers according to the information contained in the two documents. They showed the tendency to provide the same answers for both documents even though the information contained in the two summaries was dissimilar. One possible explanation is that clinicians tended to provide biased answers due to their pre-existing preferences and clinical knowledge. I did not observe any difference in the rate of correct answers between clinicians with more years of clinical practice compared to clinicians with less years of clinical practice. Similarly, the clinicians declared confidence in understanding measures of diagnostic accuracy did not result in a better interpretation of the systematic review findings. Further investigations are needed to determine the way to improve summaries of findings of systematic reviews of diagnostic test accuracy.

Chapter 6 discusses the main findings reported in this thesis and their implication for practice and research.

In conclusion, methodological issues concerning the validity and reliability of findings of studies included in systematic reviews of diagnostic accuracy remain of fundamental importance. More empirical evidence is needed to address potential biases such as reporting bias and publication bias. To allow dissemination of diagnostic reviews findings in clinical practice better ways of communicating main characteristics and key results of systematic reviews of diagnostic accuracy should be considered.

In the current literature, the quality of reporting and methodological quality of imaging studies for the diagnosis of stroke is less than satisfactory and leaves room for improvement. This is worrying, especially if current health imaging policies are in fact based on poor quality evidence and hence scarce health resources may not being deployed as effectively as they could be.

My contribution to this thesis

I was awarded a Research Fellowship from the Chief Scientist Office (CSO) to study the methodology of systematic reviews of diagnostic test accuracy.

I conducted the study reported in Chapter 2, which assessed the quality of reporting of studies on MRI for the early diagnosis of stroke. In particular, I developed the search strategy to identify relevant diagnostic accuracy studies in the literature; I personally assessed the quality of reporting of all included studies; and performed the analyses.

For the study reported in Chapter 3 on publication bias of diagnostic accuracy studies in stroke medicine, I reviewed all conference proceedings of two major international stroke meetings published as abstracts between 1995 and 2004. I extracted information on study characteristics from all identified abstracts. I then searched MEDLINE and EMBASE to identify how many abstracts were published in full in the literature. For abstracts for which a full-text publication could not be located I sent a questionnaire to the corresponding author. I entered data in Excel and SAS and performed the statistical analyses.

I performed the pilot systematic review that formed the basis of Chapter 4. I wrote the initial protocol; I developed the search strategies for identifying relevant diagnostic studies of CT and MRI for the early diagnosis of stroke; I assessed the

quality of all identified studies; I extracted the data; I entered the data in RevMan 5 and performed the statistical analyses.

I designed the study reported in Chapter 5, which assessed two forms of summary of the systematic review findings presented in Chapter 4. I developed the layout and content of a summary of findings table to present in a simple tabular format the main results of systematic reviews of diagnostic test accuracy. I tested the table in a small group of experienced clinicians. I developed the questionnaire to evaluate clinicians' ability to interpret the main findings of a diagnostic test accuracy review. I recruited clinicians (mainly neurologists, radiologists, and neuroradiologists) in the UK (Edinburgh) and in Italy (Perugia), and organised the data collection. I entered the data in Excel and Minitab. I analysed results of all completed questionnaires and interpreted them.

Work from Chapter 3 and 4 formed the basis of oral presentations at the XIV Cochrane Colloquium in Dublin (Ireland) in 2006 and at the Methods for Evaluating Medical Tests - First International Symposium in Birmingham (UK) in July 2008. Work from Chapter 5 has been presented in research seminar meetings at the Stroke Research Group as part of the activities of the Centre for Clinical Brain Sciences in Edinburgh (UK). The study presented in Chapter 3 was published in full in the *Journal of Clinical Epidemiology*. The systematic review presented in Chapter 4 was published in full in the Cochrane Library and in brief in the journal *Stroke*.

I entirely composed this thesis while I was in place at the Division of Clinical Neurosciences, University of Edinburgh. My position was funded by a CSO

Research Fellowship. This thesis has not been submitted for candidature for any other degree, postgraduate diploma or professional qualification.

Acknowledgements

I was able to complete the work for this thesis thanks to a Research Fellowship awarded by the Chief Scientist Office. Many thanks to Jennifer Waterton, Elaine Moir, and Karen Ford for being vigilant on the progress of my work and for their overall support.

I would like to express my sincere gratitude to Peter Sandercock who has provided supervision throughout my entire work and dispensed invaluable comments, positive insight, and kind support. I am also indebted to Steff Lewis and Jon Deeks for their statistical advice and for guiding my work. I am much grateful to Joanna Wardlaw for her precise comments and fruitful discussions. A special thank goes to Francesca Chappell for her help and encouragement, but specially for putting up with my occasional ranting over the past few years and cheering me up by talking to me in Italian!

I am indebted to my Italian colleagues, particularly to Maria Grazia Celani and Enrico Righetti but also to Stefano Ricci, Teresa Cantisani, and Alfonso Ciccone for their contribution and support, and for helping me with the recruitment of participants for my questionnaire survey.

The Cochrane Collaboration and in particular the work of the Diagnostic Test Accuracy Working Group has provided inspiration and stimulus for the completion of this thesis. I was lucky to be able to participate in meetings and workshops

organised by the members of the Diagnostic Test Accuracy Working Group and to discuss issues related to systematic reviews of diagnostic accuracy with major experts in the field. I particularly thank Patrick Bossuyt for his useful suggestions.

I am also grateful to the members of the Cochrane Stroke Group in particular to Brenda Thomas for her useful advice on search strategies and to Hazel Fraser for her assistance in submitting my Cochrane diagnostic accuracy review and for answering all my queries about the Cochrane Information Management System. Many thanks also to Rachel Burrow for her kind assistance with secretarial matters.

My gratitude also goes to Norman Waugh, who firstly introduced me to systematic reviewing and to the need to question assumptions and critically appraise evidence, and to Carl Counsell with whom, when I was still in Aberdeen, I initially started to work on the use of imaging tests for the diagnosis of stroke.

Last but not least, I am deeply grateful to Sergio, Sofia, Isabella, and Marta for standing by me and for tolerating my grumpy mood over the past few months. Without their support and affection I would not have been able to complete this thesis.

Appendix 1 Details of the search strategies

DWI for the diagnosis of ischaemic stroke

MEDLINE (Ovid) search strategy - January 1999 to January 2009

1. cerebrovascular disorders/ or basal ganglia cerebrovascular disease/ or exp brain ischemia/ or carotid artery diseases/ or carotid artery thrombosis/ or cerebrovascular accident/ or exp brain infarction/ or exp hypoxia-ischemia, brain/ or intracranial arterial diseases/ or cerebral arterial diseases/ or exp "intracranial embolism and thrombosis"/

2. ((brain or cerebr\$ or cerebell\$ or vertebrobasil\$ or hemispher\$ or intracran\$ or intracerebral or infratentorial or supratentorial or middle cerebr\$ or mca\$ or anterior circulation) adj5 (isch?emi\$ or infarct\$ or thrombo\$ or emboli\$ or occlus\$ or hypoxi\$)).tw.

3. (isch?emi\$ adj6 (stroke\$ or apoplex\$ or cerebral vasc\$ or cerebrovasc\$ or cva or attack\$)).tw.

4. 1 or 2 or 3

5. exp Magnetic Resonance Imaging/

6. (magnetic resonance or MR or NMR or diffusion weighted or T2-weighted).tw.

7. (MR or NMR).tw.

8. (MRI or DWI).tw.

9. 5 or 6 or 7 or 8

10. 4 and 9

11. ("1999\$" or "200\$").yr.

12. 10 and 11

13. exp "Sensitivity and Specificity"/

14. false negative reactions/ or false positive reactions/

15. Du.fs

16. (sensitiv\$ distinguish\$ or differentiat\$ or enhancement or identif\$ or detect\$ or diagnos\$ or accurat\$).tw.

17. (predictive adj4 value\$).tw.

18. or/13-17

19. 12 and 18

20. limit 19 to human

DWI for the diagnosis of ischaemic stroke

EMBASE (Ovid) search strategy - January 1999 to January 2009

1. cerebrovascular disease/ or cerebral artery disease/ or cerebrovascular accident/ or stroke/ or vertebrobasilar insufficiency/ or carotid artery disease/ or exp carotid artery obstruction/ or exp brain infarction/ or exp brain ischemia/ or exp occlusive cerebrovascular disease/

2. ((brain or cerebr\$ or cerebell\$ or vertebrobasil\$ or hemispher\$ or intracran\$ or intracerebral or infratentorial or supratentorial or middle cerebr\$ or mca\$ or anterior circulation) adj5 (isch?emi\$ or infarct\$ or thrombo\$ or emboli\$ or occlus\$ or hypoxi\$)).tw.

3. (isch?emi\$ adj6 (stroke\$ or apoplex\$ or cerebral vasc\$ or cerebrovasc\$ or cva or attack\$)).tw.

4. 1 or 2 or 3

5. exp nuclear magnetic resonance imaging/

6. ((magnetic resonance or MR or NMR or diffusion weighted or T2-weighted) adj2 imag\$).tw.

7. ((MR or NMR) adj2 tomography).tw.

8. (MRI or DWI).tw.

9. 5 or 6 or 7 or 8

10. 4 and 9

15. ("1999\$" or "200\$").yr.

16. 10 and 15

17. "sensitivity and specificity"/

18. diagnostic accuracy/

19. (sensitiv\$ or distinguish\$ or differentiat\$ or enhancement or identif\$ or detect\$ or diagnos\$ or accura\$).tw.

20. (predictive adj4 value\$).tw.

21 or/17-20

22. 16 and 21

23. limit 22 to human

Appendix 2 Details of the search strategies

CT and MRI for the diagnosis of ischaemic and haemorrhagic stroke

MEDLINE (Ovid) search strategy - January 1995 to March 2009

1. cerebrovascular disorders/ or basal ganglia cerebrovascular disease/ or exp brain ischemia/ or carotid artery diseases/ or carotid artery thrombosis/ or cerebrovascular accident/ or exp brain infarction/ or exp hypoxia-ischemia, brain/ or intracranial arterial diseases/ or cerebral arterial diseases/ or exp "intracranial embolism and thrombosis"/
2. ((brain or cerebr\$ or cerebell\$ or vertebrobasil\$ or hemispher\$ or intracran\$ or intracerebral or infratentorial or supratentorial or middle cerebr\$ or mca\$ or anterior circulation) adj5 (isch?emi\$ or infarct\$ or thrombo\$ or emboli\$ or occlus\$ or hypoxi\$)).tw.
3. (isch?emi\$ adj6 (stroke\$ or apoplex\$ or cerebral vasc\$ or cerebrovasc\$ or cva or attack\$)).tw.
- 4.1 or 2 or 3
- 5.exp Magnetic Resonance Imaging/
6. ((magnetic resonance or MR or NMR or diffusion weighted or T2-weighted) adj2 imag\$).tw.
7. ((MR or NMR) adj2 tomograph\$).tw.
8. (MRI or DWI).tw.
9. 5 or 6 or 7 or 8
- 10.exp Tomography, X-Ray Computed/
11. (CT or CAT).tw.
12. (comput\$ adj3 tomograph\$).tw.
13. 10 or 11 or 12
14. 4 and 9 and 13
15. ("1995\$" or "1996\$" or "1997\$" or "1998\$" or "1999\$" or "200\$").ed.
16. 14 and 15 (*most sensitive search - ischaemic stroke*)

17. cerebrovascular disorders/di or basal ganglia cerebrovascular disease/di or exp brain ischemia/di or carotid artery diseases/di or carotid artery thrombosis/di or cerebrovascular accident/di or exp brain infarction/di or exp hypoxia-ischemia, brain/di or intracranial arterial diseases/di or cerebral arterial diseases/di or exp "intracranial embolism and thrombosis"/di
18. exp *Magnetic Resonance Imaging/
19. exp *Tomography, X-Ray Computed/
20. 17 and 18 and 19 and 15 (**SET DOWNLOADED 1**)
21. exp "Sensitivity and Specificity"/
22. false negative reactions/ or false positive reactions/ or diagnostic errors/
23. (sensitivity\$ or specificity\$ or distinguish\$ or differentiate\$ or enhancement\$ or identify\$ or detect\$ or diagnosis\$ or accuracy\$.tw.
24. (predictive adj4 value\$.tw.
25. (false adj (positive\$ or negative\$)).tw.
26. (receiver operating adj (characteristic\$ or curve\$ or analysis)).tw.
27. (ROC or SROC).tw.
28. comparative study/
29. (compared\$ or comparison\$ or correlation\$ or versus).tw.
30. or/21-29
31. (16 and 30) not 20 (*ischaemia + MRI + CT + years + diagnostic filter - set downloaded*) (**SET DOWNLOADED 2**)
32. exp basal ganglia hemorrhage/ or exp intracranial hemorrhages/
33. ((brain\$ or cerebr\$ or cerebell\$ or intracerebral\$ or intracran\$ or parenchymal\$ or intraventricular\$ or infratentorial\$ or supratentorial\$ or basal gangli\$ or putaminal\$ or putamen\$ or posterior fossa) adj10 (haemorrhage\$ or hemorrhage\$ or haematoma\$ or hematoma\$ or bleed\$)).tw.
34. 32 or 33
35. 9 and 34 and 15 (*MRI + haemorrhage + years*) (**most sensitive search - haemorrhagic stroke**)

36. exp basal ganglia hemorrhage/di or exp intracranial hemorrhages/di
37. 36 and 18 and 15 (*haemorrhage diagnosis/di + exp MRI + years*)
38. or/21-27 (*diagnostic filter*)
39. 35 and 38 (*MRI + haemorrhage + years + diagnostic filter*)
40. 37 not (20 or 31) (*haemorrhage diagnosis/di + exp MRI + years - downloaded sets 1 and 2*) (**SET DOWNLOADED 3**)
41. 39 not (20 or 31 or 37) (*MRI (not exp MRI) + haemorrhage (not haemorrhage diagnosis/di) + years + diagnostic filter - downloaded sets 1 and 2*) (**SET DOWNLOADED 4**)

The above search strategy has been designed to cover both MRI and CT for the detection of ischaemic stroke and MRI for the detection of haemorrhagic stroke. In theory two separate search strategies could have been designed but we preferred to combine the searches to avoid looking at duplicate references.

Summary of the MEDLINE search strategy

- Line 16 of the search strategy identifies all records related to ischaemic stroke (broad terms) and the use of both MRI and CT for the period 1995 – 2009;
- Line 20 of the search strategy identifies records that focus specifically on MRI and CT for the diagnosis of ischaemic stroke (stroke terms searched with the subheading /di) for the period 1995 – 2009;
- Line 31 of the search strategy employs a diagnostic filter to identify records related to ischaemic stroke (broad terms) and the use of both MRI and CT for the period 1995 – 2009;
- Line 35 of the search strategy identifies all records related to haemorrhagic stroke (broad terms) and the use of MRI for the period 1995 – 2009;
- Line 37 of the search strategy identifies records that focus specifically on MRI for the diagnosis of haemorrhagic stroke (stroke terms searched with the subheading /di) for the period 1995 – 2009;
- Line 39 of the search strategy employs a diagnostic filter to identify records related to haemorrhagic stroke (broad terms) and the use of MRI for the period 1995 – 2009.

The most sensitive search for ischaemic stroke would have been to assess all the references at line 16 (approximately 2800 hits from 1995). However, as CT and MRI are routinely used in clinical practice and the terms occur very frequently in abstracts, we tried to limit the search specifically to imaging diagnostic studies in two ways: (a) by using the subheading diagnosis (/di) on the stroke MeSH terms and the focused imaging MeSH terms, and (b) by developing a search filter for diagnostic studies to increase precision. In order to test this approach we intended to scan the remaining references from line 16 to see if any relevant papers were missed and not identified by (a) or (b). However, due to the limited resources available this approach proved unfeasible.

The above comments apply to the search section on haemorrhagic stroke with line 35 being the most sensitive search.

CT and MRI for the diagnosis of ischaemic and haemorrhagic stroke

EMBASE (Ovid) search strategy - January 1995 to March 2009

Adapted from the MEDLINE search strategy

1. cerebrovascular disease/ or cerebral artery disease/ or cerebrovascular accident/ or stroke/ or vertebrobasilar insufficiency/ or carotid artery disease/ or exp carotid artery obstruction/ or exp brain infarction/ or exp brain ischemia/ or exp occlusive cerebrovascular disease/
2. ((brain or cerebr\$ or cerebell\$ or vertebrobasil\$ or hemispher\$ or intracran\$ or intracerebral or infratentorial or supratentorial or middle cerebr\$ or mca\$ or anterior circulation) adj5 (isch?emi\$ or infarct\$ or thrombo\$ or emboli\$ or occlus\$ or hypoxi\$)).tw.
3. (isch?emi\$ adj6 (stroke\$ or apoplex\$ or cerebral vasc\$ or cerebrovasc\$ or cva or attack\$)).tw.
4. 1 or 2 or 3
5. exp nuclear magnetic resonance imaging/
6. ((magnetic resonance or MR or NMR or diffusion weighted or T2-weighted) adj2 imag\$).tw.
7. ((MR or NMR) adj2 tomography).tw.
8. (MRI or DWI).tw.
9. 5 or 6 or 7 or 8
10. exp computer assisted tomography/
11. (CT or CAT).tw.
12. (comput\$ adj3 tomograph\$).tw.
13. 10 or 11 or 12
14. 4 and 9 and 13
15. ("1995\$" or "1996\$" or "1997\$" or "1998\$" or "1999\$" or "2000\$").em.
16. 14 and 15 (*most sensitive search - ischaemic stroke*)

17. cerebrovascular disease/di or cerebral artery disease/di or cerebrovascular accident/di or stroke/di or vertebrobasilar insufficiency/di or carotid artery disease/di or exp carotid artery obstruction/di or exp brain infarction/di or exp brain ischemia/di or exp occlusive cerebrovascular disease/di
18. exp *nuclear magnetic resonance imaging/
19. exp *computer assisted tomography/
20. 17 and 18 and 19 and 15 (**SET DOWNLOADED 1**)
21. "sensitivity and specificity"/
22. laboratory diagnosis/
23. prediction/
24. "prediction and forecasting"/
25. receiver operating characteristic/ or roc curve/
26. diagnostic accuracy/
27. diagnostic value/
28. reliability/
29. (sensitiv\$ or specificity or distinguish\$ or differentiat\$ or enhancement or identif\$ or detect\$ or diagnos\$ or accur\$).tw.
30. (predictive adj4 value\$).tw.
31. (false adj (positive\$ or negative\$)).tw.
32. (receiver operat\$ adj (characteristic\$ or curve or analysis)).tw.
33. (ROC or SROC).tw.
34. comparative study/
35. exp controlled study/
36. intermethod comparison/
37. correlation analysis/
38. (compared or comparison or correlat\$ or versus).tw.

39. or/21-38

40. (16 and 39) not 20 (*ischaemia + MRI + CT + years + diagnostic filter - set downloaded 1*) (**SET DOWNLOADED 2**)

41. basal ganglion hemorrhage/ or brain hemorrhage/ or brain ventricle hemorrhage/ or cerebellum hemorrhage/

42. ((brain\$ or cerebr\$ or cerebell\$ or intracerebral or intracran\$ or parenchymal or intraventricular or infratentorial or supratentorial or basal gangli\$ or putaminal or putamen or posterior fossa) adj10 (haemorrhage\$ or hemorrhage\$ or haematoma\$ or hematoma\$ or bleed\$)).tw.

43. 41 or 42

44. 9 and 15 and 43 (*MRI + years + haemorrhage*) (**most sensitive search - haemorrhagic stroke**)

45. basal ganglion hemorrhage/di or brain hemorrhage/di or brain ventricle hemorrhage/di or cerebellum hemorrhage/di

46. 45 and 18 and 15 (*haemorrhage diagnosis/di + exp MRI + years*)

47. 46 not (20 or 40) (*haemorrhage diagnosis/di + exp MRI + years - downloaded sets 1 and 2*) (**SET DOWNLOADED 3**)

48. or/21-33 (*diagnostic filter*)

49. 44 and 48 (*MRI + years + haemorrhage + diagnostic filter*)

50. 49 not (20 or 40 or 47) (*MRI + years + haemorrhage + diagnostic filter - downloaded sets 1,2 and 3*) (**SET DOWNLOADED 4**)

Appendix 3 Study Design and Data Extraction Form

Reviewer initials:

Date information extracted:

Type of publication:

(Full-text paper, abstract)

Study identifier:

(First author and year of publication)

Number of studies included in this paper:

(if more than one, complete separate extraction forms for each, and add letters A, B, C etc to the study identifier)

Language of report: _____

Paper no/RefManager ID of other studies with which this may link:

Aim of the study: _____

Did the study focus on ischaemic stroke **haemorrhagic stroke** **or both?**

Geographical location of study centre(s): _____

Single **or multi-centre** **study?**

Setting:

Tertiary
Secondary
Primary

Academic Unit: Yes **No**

Study Design

Systematic Review

Randomised controlled study

Prospective

Cross-sectional study

Retrospective

Case-control study

Other _____

Characteristics of the participants

Description of how the patient population was assembled (e.g. inpatients/outpatients)

Recruitment period

Inclusion criteria (clinical and imaging)

Exclusion criteria (clinical and imaging)

Criteria used to assess the clinical manifestations of stroke symptoms (e.g. NIHSS, CNS)

Selection by stroke type (e.g. restricted to MCA stroke, PACS)

Were participants with haemorrhagic stroke symptoms *included* or *excluded* at entrance?

Was haemorrhagic stroke confirmed by imaging?
No. patients eligible (e.g. all consecutive patients who met inclusion criteria) CT MRI Total
No. patients enrolled (e.g. all consecutive patients who entered the study) CT MRI Total
No. patients assessed (e.g. all patients for whom data are reported) CT MRI Total
No. of patients included in the 2x2 results tables CT MRI
No. of patients who underwent the reference standard CT MRI
No. of patients who underwent a <i>different</i> reference standard CT MRI

Age (mean, range)			
Gender	M	F	Total
Ethnicity			
No. of patients with previous history of stroke			
No. of patients who could not or did not undertake the test(s) or whose images were unusable despite completing tests			
Reasons			
Co-morbidity			
Concomitant interventions (any intervention given to all participants in addition to MRI/CT)			
Discharge diagnosis			
<i>Where TIA cases counted as true positive (i.e. stroke) or true negative cases (i.e. not stroke)?</i>			

Characteristics of the diagnostic assessment
Index test
Comparator
Reference standard
<p>For studies assessing more than one imaging test what was the order of tests?</p> <p>Was the order of tests strictly randomised?</p>
Manufacturer and model of CT scanner
Manufacturer and model of MRI scanner
MRI sequences
Were ADC maps reviewed for MR images?
<p>Slice thickness</p> <p>CT</p> <p>MRI</p>
Contrast medium:
<p>Time of imaging after symptom onset</p> <p>CT (mean, median, range)</p> <p>MRI (mean, median, range)</p>

Characteristics of the diagnostic assessment

For studies assessing more than one imaging test, time between index test and comparator(s):

Time between index test(s) and reference standard:

Who did assess and interpret imaging results (i.e. qualification and experience of the assessor)?

Was there a neuroradiologist amongst the study authors?

Were assessors of imaging results blinded to patients' clinical symptoms?

For each patient were scans from different imaging tests assessed separately?

Was inter-rater agreement assessed?

What was the definition of a positive test result?

MRI

CT

Adverse events and patient acceptability

Diagnostic results			
	MRI	CT	Comments on reference standard
True positives			
False positives			
True negatives			
False negatives			
Sensitivity			
Specificity			
Positive predictive value			
Negative predictive value			
Positive likelihood ratio			
Negative likelihood ratio			
Diagnostic odds ratio			
Diagnostic accuracy			
2x2 Table			

Other comments

APPENDIX 4 Modified QUADAS Checklist

Extractor initials:

Date information extracted:

Study identifier: _____

(First author + year of publication)

	Item	Yes	No	Unclear
1.	Was the spectrum of patients representative of the patients who will receive the test in practice? (e.g. female and male patients of all ages presenting with light, mild, or severe stroke symptoms, with or without previous history of stroke, scanned within a few hours of onset)			
2.	Were selection criteria clearly described?			
3.	Is the reference standard likely to correctly classify the target condition? (e.g. expert clinical assessment coupled with clinical and imaging follow-up)			
4.	Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?			
5.	Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?			
6.	Did patients receive the same reference standard regardless of the index test result?			
7.	Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?			
8.	Was the execution of the index test described in sufficient detail to permit replication of the test?			
9.	Was the execution of the reference standard described in sufficient detail to permit its replication?			
10.	Were the index test results interpreted without knowledge of the results of the reference standard?			
11.	Were the reference standard results interpreted without knowledge of the results of the index test?			
12.	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?			
13.	Were uninterpretable/ intermediate test results reported?			
14.	Were withdrawals from the study explained?			
15.	Was the study prospective?			
16.	<i>For retrospective studies</i> Were the original scans re-examined by study investigators?			
17.	Is the technology of the index test(s) likely to be changed since the study was carried out?			
18.	Was the expertise of the clinician(s) assessing results of the diagnostic tests reported?			
19.	<i>For studies with a direct comparison of diagnostic tests:</i> Was the sequence of the diagnostic tests (index test + comparator) determined at random?			
20.	Were the scans read blind to clinical data?			

Appendix 5 Characteristics and references of excluded studies

Study ID	Reasons for exclusion
<i>Allkemper 2004</i>	No direct comparison MRI with CT.
<i>Arenillas 2002</i>	Focus on 'early neurological deterioration' in patients with proven MCA and ICA occlusion. Beyond the scope of this review.
<i>Arnould 2004</i>	Focus on haemorrhagic transformation in hyperacute ischaemic stroke. Beyond the scope of this review.
<i>Ba-Ssalamaha 2000</i>	Heterogeneous sample. Vascular lesions in only 9 patients. No suitable diagnostic accuracy data.
<i>Barber 2005</i>	Comparison of CT and DWI in acute ischaemic stroke using the Alberta Stroke Programme Early Computed Tomography Score (ASPECTS criteria). No suitable diagnostic accuracy data.
<i>Bartylla 1997</i>	German study. No direct comparison of CT with DWI. Only DWI and T2WI assessed.
<i>Brant-Zawadzki 1996</i>	Focus on FLAIR images - not on DWI. No suitable imaging test.
<i>Buckley 2003</i>	Audit of MRI as first-line neuroimaging for stroke patients. Patients scanned within 48 hours. Not suitable diagnostic data.
<i>Chung 2002</i>	Four single cases of ischaemic stroke assessed by DWI.
<i>Chung 2003</i>	Five single cases of haemorrhagic stroke assessed by DWI .
<i>Dorenbeck 2005</i>	Assessment of ADC values obtained using DWI. No suitable diagnostic accuracy data.
<i>Dylewski 2000</i>	Use of MRI in acute intracerebral haemorrhage. Heterogeneous etiologies of haemorrhage. No suitable patient population.
<i>Eastwood 2003</i>	Correlation of dynamic CT perfusion imaging and MR diffusion and perfusion imaging in acute stroke. No direct comparison of MRI with non-contrast CT.

Characteristics of excluded studies - *continued*

Study ID	Reasons for exclusion
<i>Ebisu 1997</i>	No suitable diagnostic accuracy data.
<i>Egelhof 1998</i>	German study. MRI to detect acute ischaemic cerebral infarcts. Imaging performed within 48 hours of stroke onset (and within 24 hours only in a subgroup of patients).
<i>Eliasziw 2005</i>	Letter/comment with no suitable diagnostic accuracy data.
<i>Etgen 2004</i>	Study looking at stroke in one anatomical region (brainstem infarcts). Only DWI assessed. 62% of patients were scanned outside 24 hours.
<i>Fazekas 1996</i>	Frequency and type of TIA-related infarcts shown by MRI. Beyond the scope of this review.
<i>Fiebach 2001</i>	No enough data to allow construction of a 2x2 contingency table.
<i>Fiebach 2002</i>	Only sensitivity and specificity estimates reported. No enough data to construct a 2x2 contingency table
<i>Fiebach 2004</i>	No enough data to allow construction of a 2x2 contingency table.
<i>Fitzek 1998</i>	Comparison of CT with DWI for detection of acute ischaemic stroke. Imaging performed within 11 days after stroke onset.
<i>Flacke 1998</i>	German study. Assessment of diffusion-weighted and perfusion imaging in addition to FLAIR-TSE and T2W-GraSE and MR angiography for the diagnosis of acute stroke. No direct comparison of MRI with CT.
<i>Flacke 2000</i>	MCA susceptibility sign compare with hyperdense MCA sign on CT. No suitable test comparison.
<i>Girot 2003</i>	Focus on inter- and intra-observer reproducibility. No suitable diagnostic accuracy data.
<i>Greer 2004</i>	Evaluation of DWI versus CT for the detection of haemorrhage after thrombolysis. Beyond the scope of this review.
<i>Griffiths 2000</i>	Assessment of patients with neurological symptoms and signs (not limited to stroke patients). Imaging performed within 18 hours of stroke onset.

Characteristics of excluded studies - *continued*

Study ID	Reasons for exclusion
<i>Hacke 2000</i>	Letter/comment with no suitable diagnostic accuracy data.
<i>Haraguchi 2000</i>	Japanese study. No suitable diagnostic accuracy data.
<i>Heidenreich 2008</i>	MRI in addition to CT for the diagnosis of hyperacute stroke. MRI protocol included T2-W, DWI, PWI, and MRA. No direct comparison of DWI with CT.
<i>Hermier 2001</i>	DWI for the detection of post-ischaemic haemorrhage. Beyond the scope of this review
<i>Jager 2000</i>	Narrative review of the literature. No suitable diagnostic accuracy data.
<i>Jaillard 2002</i>	Early CT signs in acute stroke. Beyond the scope of this review.
<i>Kamal 2003</i>	Tissue response of the brain to intracranial haemorrhage as shown by DWI. Beyond the scope of this review.
<i>Keir 2000</i>	Systematic review of diffusion and perfusion imaging in acute ischemic stroke. No suitable diagnostic accuracy data.
<i>Kidwell 2008</i>	Discussion paper on neuroimaging for the diagnosis of intracranial haemorrhage. No suitable diagnostic accuracy data
<i>Kimura 1999</i>	Duration of symptoms in TIA. Beyond the scope of this review
<i>Kloska 2004</i>	No direct comparison of CT with DWI. Only CT assessed.
<i>Koennecke 2001</i>	Not suitable diagnostic data. Only positive cases on DWI analysed.
<i>Krasnianski 2001</i>	German study. MRI findings in patients with brainstem infarctions. Imaging performed within 7 days of stroke onset. No suitable time of imaging.
<i>Krasnianski 2002</i>	Brainstem infarctions in patients with normal MRI. Beyond the scope of this review.
<i>Köhrmann 2007</i>	Discussion paper on acute stroke imaging for thrombolytic therapy. No suitable diagnostic accuracy data
<i>Laloux 1995</i>	Mean interval of MRI: 11 days. No suitable time of imaging.

Characteristics of excluded studies - *continued*

Study ID	Reasons for exclusion
<i>Lam 2003</i>	CT and DWI for the detection of haemorrhagic stroke. Imaging performed within 40 hours of stroke onset. No enough data to allow construction of a 2X2 contingency table.
<i>Lam 2005</i>	Use of B0 echo planar imaging (EPI) for the detection of intracerebral bleeds. Imaging performed within 48 hours. Beyond the scope of this review.
<i>Lansberg 2000</i>	Comparison of DWI with CT for the detection of ischaemic stroke. No enough data to allow construction of a 2x2 contingency table.
<i>Lansberg 2000a</i>	Conventional MRI versus DWI. No direct comparison of DWI with CT.
<i>Lee 2000</i>	No direct comparison of MRI with CT.
<i>Lee 2001</i>	Assessment of the Yonsei Stroke Registry. No suitable diagnostic accuracy data.
<i>Lev 2000</i>	Focus on CTA. No suitable test comparison.
<i>Lin 2001</i>	Only a non-random subset of patients underwent MRI and CT. Imaging performed within 4 days after stroke. No suitable time of imaging.
<i>Linfante 1999</i>	Description of five cases with intracerebral haemorrhage.
<i>Linfante 2001</i>	DWI in acute posterior circulation stroke. No suitable diagnostic accuracy data.
<i>Linfante 2004</i>	Letter/comment with no suitable diagnostic accuracy data
<i>Lövblad 1998</i>	No direct comparison of CT with DWI. Only DWI assessed. Imaging performed within 24 of stroke onset.
<i>Lövblad 1998a</i>	Comparison of diffusion-weighted spin-echo with diffusion-weighted HASTE sequences in ischaemic stroke. No data on CT.
<i>Marx 2004</i>	German study. DWI in vertebrobasilar ischaemia. Only posterior circulation strokes included. Imaging performed within 24 hours of stroke onset. No suitable time of imaging

Characteristics of excluded studies - *continued*

Study ID	Reasons for exclusion
<i>Masdeu 2006</i>	Guideline on neuroimaging in acute stroke. No suitable diagnostic accuracy data.
<i>Mayer 2000</i>	Focus on haemorrhagic transformation. Beyond the scope of this review.
<i>Melhem 1998</i>	Dual-echo gradient- and spin-echo and fast spin-echo MRI for haemorrhagic lesions. Heterogeneous patient population. Causes of haemorrhage included ischemia, trauma, vascular malformations, hypertension, and brain tumors.
<i>Mohr 1995</i>	T1-T2 versus CT. No suitable test comparison
<i>Mullins 2002</i>	Retrospective studies on CT and DWI for detection of acute stroke. Ischaemic and haemorrhagic cases were not reported separately. No direct comparison of CT and DWI.
<i>Mullins 2002a</i>	Retrospective studies on CT and DWI for detection of acute stroke. Ischaemic and haemorrhagic cases were not reported separately. No direct comparison of CT and DWI. Same data as in Mullins 2002.
<i>Na 1998</i>	Evaluation of MCA occlusion using triphasic helical CT. Beyond the scope of this review
<i>Nighoghossian 2001</i>	Focus on haemorrhagic transformations. Beyond the scope of this review.
<i>Olszycki 2007</i>	CT and MRI in patients with acute stroke. Imaging performed between 3 and 15 hours of stroke onset. No suitable time of imaging.
<i>Oppenheim 2000</i>	Assessment of DWI and FLAIR sequences for the diagnosis of ischaemic stroke. No data on CT.
<i>Patel 1996</i>	MRI for the detection of intraparenchymal haemorrhage. Description of five cases.
<i>Poniatowska 2007</i>	No direct comparison of DWI with CT. DWI performed only on negative cases. No suitable test comparison.
<i>Powers 2000</i>	Letter/comment with no suitable diagnostic accuracy data.

Characteristics of excluded studies - *continued*

Study ID	Reasons for exclusion
<i>Rajajee 2008</i>	Clinical and CT criteria versus MRI for the diagnosis of small deep infarcts. DW and MRA imaging used to exclude large-vessel stenosis or occlusion (reference standard). No suitable test comparison.
<i>Razumovsky 1999</i>	TCD, MRA, and MRI in acute cerebral ischemia. No data on DWI.
<i>Read 1998</i>	CT at admission and DWI for the diagnosis of ischaemic stroke. Delay between CT and DWI varied from 11 to 36 hours. No suitable time of imaging.
<i>Restrepo 2004</i>	Assessment of TIA with diffusion and perfusion MRI. No suitable test comparison.
<i>Rincon 2004</i>	Dynamic CT perfusion for acute ischemia. No suitable imaging test.
<i>Roberts 2001</i>	Focus on CT perfusion. No suitable imaging test.
<i>Rovira 2000</i>	No direct comparison of CT with DWI. Imaging performed within 48 hours of stroke onset. No suitable time of imaging.
<i>Rovira 2002</i>	DWI in acute TIA. Patients studied with MRI within 10 days. No suitable time of imaging.
<i>Schellinger 1999</i>	Selected sample (9 patients with ICH). Assessment of hematoma size on CT and MRI. Beyond the scope of this review.
<i>Schellinger 2000</i>	Practicality of MRI in acute ischemia. Beyond the scope of this review.
<i>Schellinger 2001</i>	PWI and DWI lesion volumes in hyperacute ischaemia. Beyond the scope of this review.
<i>Schramm 2002</i>	Focus on CTA versus MRA. Assessment of blood volumes. No suitable test comparison.
<i>Singer 1998</i>	No direct comparison of CT with DWI. Only DWI assessed. Mean time from stroke onset to imaging: 48.1 hours (range 7 hours - 4 days). No suitable test comparison.
<i>Smajlovic 2004</i>	DWI and CT in acute ischaemic stroke. DWI performed 48 hours after stroke onset. No suitable time of imaging.

Characteristics of excluded studies - *continued*

Study ID	Reasons for exclusion
<i>Stapf 2000</i>	No direct comparison of CT with DWI. Only CT assessed.
<i>Sunshine 2001</i>	No direct comparison of CT with DWI. Only DWI assessed.
<i>Sunshine 2004</i>	Discussion paper on the use of CT, MRI and MRA in the evaluation of acute stroke. No suitable diagnostic accuracy data.
<i>Tei 1997</i>	Japanese study. No direct comparison of DWI with CT.
<i>Toyoda 2001</i>	Use of FLAIR for detecting intra-arterial signal of ischaemia. Beyond the scope of this review.
<i>Van Everdingen 1998</i>	No direct comparison of CT with DWI. Only DWI assessed.
<i>Verro 2002</i>	Focus on CT angiography. Beyond the scope of this review
<i>Von Kummer 2000</i>	Letter/comment with no suitable diagnostic accuracy data
<i>von Kummer 2001</i>	No direct comparison of CT with DWI. Only CT assessed.
<i>Von Kummer 2002</i>	Letter/comment with no suitable diagnostic accuracy data.
<i>Wang 1997</i>	Chinese study. Not a diagnostic accuracy study.
<i>Warach 1995</i>	No direct comparison of CT with DWI. Only DWI assessed. Imaging performed within 48 of stroke onset.
<i>Warach 1996</i>	No direct comparison of CT with DWI. Only DWI assessed. Imaging performed within 48 of stroke onset. Same data as in Warach 1995.
<i>Wardlaw 2003</i>	Impact of delays in CT of the brain on the accuracy of stroke diagnosis. Beyond the scope of this review.
<i>Watanabe 2000</i>	Japanese study. No suitable test comparisons. No CT data.

Characteristics of excluded studies - *continued*

Study ID	Reasons for exclusion
<i>Weber 2003</i>	German study. No direct comparison of DWI with CT.
<i>Wintermark 2005</i>	Accuracy of dynamic perfusion CT. No suitable imaging test.
<i>Wycliffe 2004</i>	MRI for detection of haemorrhagic transformations. Beyond the scope of this review.
<i>Zivin 1997</i>	Letter/comment with no suitable diagnostic accuracy data.

References to excluded studies

Allkemper 2004

Allkemper T, Tombach B, Schwindt W, Kugel H, Schilling M, Debus O, et al. Acute and subacute intracerebral hemorrhages: comparison of MR imaging at 1.5 and 3.0 T - initial experience. *Radiology* 2004;232(3):874-81.

Arenillas 2002

Arenillas JF, Rovira A, Molina CA, Grive E, Montaner J, Alvarez-Sabin J, et al. Prediction of early neurological deterioration using diffusion- and perfusion-weighted imaging in hyperacute middle cerebral artery ischemic stroke. *Stroke* 2002;33(9):2197-203.

Arnould 2004

Arnould MC, Grandin CB, Peeters A, Cosnard G, Duprez TP. Comparison of CT and three MR sequences for detecting and categorizing early (48 hours) hemorrhagic transformation in hyperacute ischemic stroke. *American Journal of Neuroradiology* 2004;25(6):939-44.

Ba-Ssalamaha 2000

Ba-Ssalamaha A, Schick S, Heimberger K, Linnau KF, Schibany N, Prokesch R, et al. Ultrafast magnetic resonance imaging of the brain. *Magnetic Resonance Imaging* 2000;18(3):237-43.

Barber 2005

Barber PA, Hill MD, Eliasziw M, Demchuk AM, Pexman JH, Hudon ME, et al. Imaging of the brain in acute ischaemic stroke: comparison of computed tomography

and magnetic resonance diffusion-weighted imaging. *Journal of Neurology, Neurosurgery, and Psychiatry* 2005;76(11):1528-33.

Bartylla 1997

Bartylla K, Hagen T, Globel H, Jost V, Schneider G. [Diffusion-weighted magnetic resonance imaging in the diagnosis of cerebral infarct] [German]. *Radiologe* 1997;37(11):859-64.

Brant-Zawadzki 1996

Brant-Zawadzki M, Atkinson D, Detrick M, Bradley WG, Scidmore G. Fluid-attenuated inversion recovery (FLAIR) for assessment of cerebral infarction. Initial clinical experience in 50 patients. *Stroke* 1996;27(7):1187-91.

Buckley 2003

Buckley BT, Wainwright A, Meagher T, Briley D. Audit of a policy of magnetic resonance imaging with diffusion-weighted imaging as first-line neuroimaging for in-patients with clinically suspected acute stroke. *Clinical Radiology* 2003;58(3):234-7.

Chung 2002

Chung SP, Ha YR, Kim SW, Yoo IS. Diffusion-weighted MRI as a screening tool of stroke in the ED. *American Journal of Emergency Medicine* 2002;20(4):327-31.

Chung 2003

Chung SP, Ha YR, Kim SW, Yoo IS. Diffusion-weighted MRI of intracerebral hemorrhage clinically undifferentiated from ischemic stroke. *American Journal of Emergency Medicine* 2003;21(3):236-40.

Dorenbeck 2005

Dorenbeck U, Schlaier J, Bretschneider T, Schuierer G, Feuerbach S. Diffusion-weighted imaging with calculated apparent diffusion coefficient in intracranial hemorrhagic lesions. *Clinical Imaging* 2005;29(2):86-93.

Dylewski 2000

Dylewski DA, Demchuk AM, Morgenstern LB. Utility of magnetic resonance imaging in acute intracerebral hemorrhage. *Journal of Neuroimaging* 2000;10(2):78-83.

Eastwood 2003

Eastwood JD, Lev MH, Wintermark M, Fitzek C, Barboriak DP, DeLong DM, et al. Correlation of early dynamic CT perfusion imaging with whole-brain MR diffusion and perfusion imaging in acute hemispheric stroke. *American Journal of Neuroradiology* 2003;24(9):1869-75.

Ebisu 1997

Ebisu T, Tanaka C, Umeda M, Kitamura M, Fukunaga M, Aoki I, et al. Hemorrhagic and nonhemorrhagic stroke: diagnosis with diffusion-weighted and T2-weighted echo-planar MR imaging. *Radiology* 1997;203(3):823-8.

Egelhof 1998

Egelhof T, Essig M, Von Kummer R, Dorfler A, Winter R, Sartor K, et al. [Acute ischemic cerebral infarct: prospective serial observations by magnetic resonance imaging] [German]. *Rofo: Fortschritte auf dem Gebiete der Rontgenstrahlen und der Nuklearmedizin* 1998;168(3):222-7.

Eliasziw 2005

Eliasziw M, Paddock-Eliasziw L. Comparison of MRI and CT for detection of acute intracerebral hemorrhage. JAMA 2005;293(5):550-1.

Etgen 2004

Etgen T, Graf von Einsiedel, Rottinger M, Winbeck K, Conrad B, Sander D, et al. Detection of acute brainstem infarction by using DWI/MRI. European Neurology 2004;52(3):145-50.

Fazekas 1996

Fazekas F, Fazekas G, Schmidt R, Kapeller P, Offenbacher H. Magnetic resonance imaging correlates of transient cerebral ischemic attacks. Stroke 1996;27(4):607-11.

Fiebach 2001

Fiebach J, Jansen O, Schellinger P, Knauth M, Hartmann M, Heiland S, et al. Comparison of CT with diffusion-weighted MRI in patients with hyperacute stroke. Neuroradiology 2001;43(8):628-32.

Fiebach 2002

Fiebach JB, Schellinger PD, Jansen O, Meyer M, Wilde P, Bender J, et al. CT and diffusion-weighted MR imaging in randomized order: diffusion-weighted imaging results in higher accuracy and lower interrater variability in the diagnosis of hyperacute ischemic stroke. Stroke 2002;33(9):2206-10.

Fiebach 2004

Fiebach JB, Schellinger PD, Gass A, Kucinski T, Siebler M, Villringer A, et al. Stroke magnetic resonance imaging is accurate in hyperacute intracerebral

hemorrhage: a multicenter study on the validity of stroke imaging. *Stroke* 2004;35(2):502-6.

Fitzek 1998

Fitzek C, Tintera J, Muller-Forell W, Urban P, Thomke F, Fitzek S, et al. Differentiation of recent and old cerebral infarcts by diffusion-weighted MRI. *Neuroradiology* 1998;40(12):778-82.

Flacke 1998

Flacke S, Keller E, Hartmann A, Murtz P, Textor J, Urbach H, et al. [Improved diagnosis of early cerebral infarct by the combined use of diffusion and perfusion] [German]. *Rofo: Fortschritte auf dem Gebiete der Rontgenstrahlen und der Nuklearmedizin* 1998;168(5):493-501.

Flacke 2000

Flacke S, Urbach H, Keller E, Traber F, Hartmann A, et al. Middle cerebral artery (MCA) susceptibility sign at susceptibility-based perfusion MR imaging: clinical importance and comparison with hyperdense MCA sign at CT. *Radiology* 2000;215(2):476-82.

Giot 2003

Giot M, Leclerc X, Gauvrit JY, Verdelho A, Pruvo JP, Leys D, et al. Cerebral magnetic resonance imaging within 6 hours of stroke onset: inter- and intra-observer reproducibility. *Cerebrovascular Diseases* 2003;16(2):122-7.

Greer 2004

Greer DM, Koroshetz WJ, Cullen S, Gonzalez RG, Lev MH. Magnetic resonance imaging improves detection of intracerebral hemorrhage over computed tomography after intra-arterial thrombolysis. *Stroke* 2004;35(2):491-5.

Griffiths 2000

Griffiths PD, Wilkinson ID, Patel MC, Romanowski CA, Mitchell P, Graham A, et al. Acute neuromedical and neurosurgical admissions. Standard and ultrafast MR imaging of the brain compared with cranial CT. *Acta Radiologica* 2000;41(5):401-9.

Hacke 2000

Hacke W, Warach S. Diffusion-weighted MRI as an evolving standard of care in acute stroke. *Neurology* 2000;54(8):1548-9.

Haraguchi 2000

Haraguchi K, Takaya S, Sakamoto Y, Morimoto S, Tanooka A, Ishizaki T, et al. [Benefits of 0.5T MR diffusion-weighted images for super-acute-phase cerebral infarction] [Japanese]. *No Shinkei Geka - Neurological Surgery* 2000;28(7):615-21.

Heidenreich 2008

Heidenreich JO, Hsu D, Wang G, Jesberger JA, Tarr RW, Zaidat OO, Sunshine JL. Magnetic resonance imaging results can affect therapy decisions in hyperacute stroke care. *Acta Radiologica* 2008;5:550-7.

Hermier 2001

Hermier M, Nighoghossian N, Derex L, Berthezene Y, Blanc-Lasserre K, Trouillas P, et al. MRI of acute post-ischemic cerebral hemorrhage in stroke patients:

diagnosis with T2*-weighted gradient-echo sequences. *Neuroradiology* 2001;43(10):809-15.

Jager 2000

Jager HR. Diagnosis of stroke with advanced CT and MR imaging. *British Medical Bulletin* 2000;56(2):318-33.

Jaillard 2002

Jaillard A, Hommel M, Baird AE, Linfante I, Llinas RH, Caplan LR, et al. Significance of early CT signs in acute stroke. A CT scan-diffusion MRI study. *Cerebrovascular Diseases* 2002;13(1):47-56.

Kamal 2003

Kamal AK, Dyke JP, Katz JM, Liberato B, Filippi CG, Zimmerman RD, et al. Temporal evolution of diffusion after spontaneous supratentorial intracranial hemorrhage. *American Journal of Neuroradiology* 2003;24(5):895-901.

Keir 2000

Keir SL, Wardlaw JM. Systematic review of diffusion and perfusion imaging in acute ischemic stroke. *Stroke* 2000;31(11):2723-31.

Kidwell 2008

Kidwell CS, Wintermark M. Imaging of intracranial haemorrhage. *Lancet Neurology* 2008;7:256-67.

Kimura 1999

Kimura K, Minematsu K, Yasaka M, Wada K, Yamaguchi T. The duration of symptoms in transient ischemic attack. *Neurology* 1999;52(5):976-80.

Kloska 2004

Kloska SP, Nabavi DG, Gaus C, Nam EM, Klotz E, Ringelstein EB, et al. Acute stroke assessment with CT: do we need multimodal evaluation? *Radiology* 2004;233(1):79-86.

Koennecke 2001

Koennecke HC, Bernarding J, Braun J, Faulstich A, Hofmeister C, Nohr R, et al. Scattered brain infarct pattern on diffusion-weighted magnetic resonance imaging in patients with acute ischemic stroke. *Cerebrovascular Diseases* 2001;11(3):157-63.

Krasnianski 2001

Krasnianski M, Georgiadis D, Grehl H, Lindner A. Correlation of clinical and MR-tomographic findings in patients with infarctions of brainstem. *Fortschritte der Neurologie-Psychiatrie* 2001;69:236-41.

Krasnianski 2002

Krasnianski M, Lindner A, Zierz S. Brainstem infarctions with normal MRI. *European Journal of Medical Research* 2002;7(3):125-7.

Köhrmann 2007

Köhrmann M, Jüttler E, Huttner HB, Nowe T, Schellinger PD. Acute stroke imaging for thrombolytic therapy - an update. *Cerebrovascular Diseases* 2007;24:161-9.

Laloux 1995

Laloux P, Richelle F, De Coster P, Jamart J. HMPAO single-photon emission computed tomography in posterior circulation infarcts. *Journal of Neuroimaging* 1995;5(3):145-51.

Lam 2003

Lam WW, So NM, Wong KS, Rainer T. B0 images obtained from diffusion-weighted echo planar sequences for the detection of intracerebral bleeds. *Journal of Neuroimaging* 2003;13(2):99-105.

Lam 2005

Lam WW, Wong KS, Rainer TH, So NM. Assessment of hyperacute stroke like symptoms by diffusion-weighted images. *Clinical Imaging* 2005;29(1):6-9.

Lansberg 2000

Lansberg MG, Albers GW, Beaulieu C, Marks MP. Comparison of diffusion-weighted MRI and CT in acute stroke. *Neurology* 2000;54(8):1557-61.

Lansberg 2000a

Lansberg MG, Norbash AM, Marks MP, Tong DC, Moseley ME, Albers GW. Advantages of adding diffusion-weighted magnetic resonance imaging to conventional magnetic resonance imaging for evaluating acute stroke. *Archives of Neurology* 2000;57:1311-16.

Lee 2000

Lee LJ, Kidwell CS, Alger J, Starkman S, Saver JL. Impact on stroke subtype diagnosis of early diffusion-weighted magnetic resonance imaging and magnetic resonance angiography. *Stroke* 2000;31(5):1081-9.

Lee 2001

Lee BI, Nam HS, Heo JH, Kim DI, Yonsei Stroke Team. Yonsei Stroke Registry. Analysis of 1,000 patients with acute cerebral infarctions. *Cerebrovascular Diseases* 2001;12(3):145-51.

Lev 2000

Lev MH, Nichols SJ, Lev MH, Nichols SJ. Computed tomographic angiography and computed tomographic perfusion imaging of hyperacute stroke. *Topics in Magnetic Resonance Imaging* 2000;11(5):273-87.

Lin 2001

Lin DD, Filippi CG, Steever AB, Zimmerman RD. Detection of intracranial hemorrhage: comparison between gradient-echo images and b(0) images obtained from diffusion-weighted echo-planar sequences. *American Journal of Neuroradiology* 2001;22(7):1275-81.

Linfante 1999

Linfante I, Llinas RH, Caplan LR, Warach S. MRI features of intracerebral hemorrhage within 2 hours from symptom onset. *Stroke* 1999;30(11):2263-7.

Linfante 2001

Linfante I, Llinas RH, Schlaug G, Chaves C, Warach S, Caplan LR. Diffusion-weighted imaging and National Institutes of Health Stroke Scale in the acute phase of posterior-circulation stroke. *Archives of Neurology* 2001;58(4):621-8.

Linfante 2004

Linfante I. Editorial comment--can MRI reliably detect hyperacute intracerebral hemorrhage? Ask the medical student. *Stroke* 2004;35(2):506-7.

Lövblad 1998

Lövblad KO, Laubach HJ, Baird AE, Curtin F, Schlaug G, Edelman RR, Warach S. Clinical experience with diffusion-weighted MR in patients with acute stroke. *American Journal of Neuroradiology* 1998;19(6):1061-6.

Lövblad 1998a

Lövblad KO, Jakob PM, Chen Q, Baird AE, Schlaug G, Warach S, Edelman RR. Turbo spin-echo diffusion-weighted MR of ischemic stroke. *American Journal of Neuroradiology* 1998;19:201-8.

Marx 2004

Marx JJ, Thoemke F, Mika-Gruettner A, Fitzek S, Vucurevic G, Urban PP, et al. [Diffusion-weighted MRT in vertebrobasilar ischemia. Application, sensitivity, and prognostic value] [German]. *Nervenarzt* 2004;75(4):341-6.

Masdeu 2006

Masdeu JC, Irimia P, Asenbaum S, Bogosslavsky J, Brainin M, Chabriat H, et al. EFNS guideline on neuroimaging in acute stroke. Report of an EFNS task force. *European Journal of Neurology* 2006;13:1271-83.

Mayer 2000

Mayer TE, Schulte-Altendorneburg G, Droste DW, Bruckmann H. Serial CT and MRI of ischaemic cerebral infarcts: frequency and clinical impact of haemorrhagic transformation. *Neuroradiology* 2000;42(4):233-9.

Melhem 1998

Melhem ER, Patel RT, Whitehead RE, Bhatia RG, Rockwell DT, Jara H, et al. MR imaging of hemorrhagic brain lesions: a comparison of dual-echo gradient- and spin-echo and fast spin-echo techniques. *American Journal of Roentgenology* 1998;171(3):797-802.

Mohr 1995

Mohr JP, Biller J, Hilal SK, Yuh WTC, Tatemichi TK, Hedges S, et al. Magnetic resonance versus computed tomographic imaging in acute stroke. *Stroke* 1995;26(5):807-12.

Mullins 2002

Mullins ME, Schaefer PW, Sorensen AG, Halpern EF, Ay H, He J, et al. CT and conventional and diffusion-weighted MR imaging in acute stroke: study in 691 patients at presentation to the emergency department. *Radiology* 2002;224(2):353-60.

Mullins 2002a

Mullins ME, Lev MH, Schellingerhout D, Koroshetz WJ, Gonzalez RG. Influence of availability of clinical history on detection of early stroke using unenhanced CT and diffusion-weighted MR imaging. *American Journal of Roentgenology* 2002;179(1):223-8.

Na 1998

Na DG, Byun HS, Lee KH, Chung CS, Kim EY, Ro DW, et al. Acute occlusion of the middle cerebral artery: early evaluation with triphasic helical CT - preliminary results. *Radiology* 1998;207(1):113-22.

Nighoghossian 2001

Nighoghossian N, Hermier M, Berthezene Y, Wiart M, Derex L, Honnorat J, et al. Early diagnosis of hemorrhagic transformation: diffusion/perfusion-weighted MRI versus CT scan. *Cerebrovascular Diseases* 2001;11(3):151-6.

Olszycki 2007

Olszycki M, Grzelak P, Biernacki R, Majos A, Stefańczyk L. Comparison of the fluid attenuated inversion recovery and diffusion-weighted imaging in the early brain stroke. *Polski Merkurusz Lekarski* 2007;12(127):28-31.

Oppenheim 2000

Oppenheim C, Logak M, Dormont D, et al. Diagnosis of acute ischaemic stroke with fluid-attenuated inversion recovery and diffusion-weighted sequences. *Neuroradiology* 2000;42:602-7.

Patel 1996

Patel MR, Edelman RR, Warach S. Detection of hyperacute primary intraparenchymal hemorrhage by magnetic resonance imaging. *Stroke* 1996;27(12):2321-4.

Poniatowska 2007

Poniatowska R, Ryterski J, Boguslawska R, Sobczyk W, Kobayashi A. Early signs of acute middle cerebral artery ischemia in computerized tomography and diffusion weighted magnetic resonance. Polish Journal of Radiology 2007;72(2):65-70.

Powers 2000

Powers WJ. Testing a test: a report card for DWI in acute stroke. Neurology 2000;54(8):1549-51.

Rajajee 2008

Rajajee V, Kidwell C, Starkman S, Ovbiagele B, Alger J, Villablanca P, et al. Diagnosis of lacunar infarcts within 6 hours of onset by clinical and CT criteria versus MRI. Journal of Neuroimaging 2008;18:66-72.

Razumovsky 1999

Razumovsky AY, Gillard JH, Bryan RN, et al. TCD, MRA and MRI in acute cerebral ischemia. Acta Neurologica Scandinavica 1999;29(1):65-76.

Read 1998

Read SJ, Jackson GD, Abbott DF, Syngienotis A, Mitchell LA, Fitt GR, et al. Experience with diffusion-weighted imaging in an acute stroke unit. Cerebrovascular Diseases 1998;8(3):135-43.

Restrepo 2004

Restrepo L, Jacobs MA, Barker PB, Wityk RJ. Assessment of transient ischemic attack with diffusion- and perfusion-weighted imaging. American Journal of Neuroradiology 2004;25(10):1645-52.

Rincon 2004

Rincon F. Anticoagulation and thrombolysis for acute ischemic stroke and the role of diagnostic magnetic resonance imaging. *Archives of Neurology* 2004;61(5):801-2.

Roberts 2001

Roberts HC, Roberts TP, Smith WS, Lee TJ, Fischbein NJ, Dillon WP, et al. Multisection dynamic CT perfusion for acute cerebral ischemia: the "togglng-table" technique. *American Journal of Neuroradiology* 2001;22(6):1077-80.

Rovira 2000

Rovira A, Pedraza S, Molina C, Capellades J, Grive E. [Diffusion-weighted magnetic resonance in the diagnosis of acute subcortical infarcts] [Spanish]. *Revista de Neurologia* 2000;30(10):914-9.

Rovira 2002

Rovira A, Rovira-Gols A, Pedraza S, Grive E, Molina C, Alvarez-Sabin J, et al. Diffusion-weighted MR imaging in the acute phase of transient ischemic attacks. *American Journal of Neuroradiology* 2002;23(1):77-83.

Schellinger 1999

Schellinger PD, Jansen O, Fiebach JB, Hacke W, Sartor K. A standardized MRI stroke protocol: comparison with CT in hyperacute intracerebral hemorrhage. *Stroke* 1999;30(4):765-8.

Schellinger 2000

Schellinger PD, Jansen O, Fiebach JB, Pohlers O, Ryssel H, Heiland S, et al. Feasibility and practicality of MR imaging of stroke in the management of

hyperacute cerebral ischemia. American Journal of Neuroradiology 2000;21(7):1184-9.

Schellinger 2001

Schellinger PD, Fiebach J, Mohr A, Kollmar R, Schwarz S, Schabitz WR, et al. [Value of MRI in intracerebral and subarachnoid hemorrhage] [German]. Nervenarzt 2001;72(12):907-17.

Schramm 2002

Schramm P, Schellinger PD, Fiebach JB, Heiland S, Jansen O, Knauth M, et al. Comparison of CT and CT angiography source images with diffusion-weighted imaging in patients with acute stroke within 6 hours after onset. Stroke 2002;33(10):2426-32.

Singer 1998

Singer MB, Chong J, Lu D, Schonewille WJ, Tuhim S, Atlas SW. Diffusion-weighted MRI in acute subcortical infarction. Stroke 1998;29(1):133-6.

Smajlovic 2004

Smajlovic D, Sinanovic O. Sensitivity of the neuroimaging techniques in ischemic stroke. Medicinski Arhiv 2004;58(5):282-4.

Stapf 2000

Stapf C, Hofmeister C, Hartmann A, Marx P, Mast H. Predictive value of clinical lacunar syndromes for lacunar infarcts on magnetic resonance brain imaging. Acta Neurologica Scandinavica 2000;101(1):13-8.

Sunshine 2001

Sunshine JL, Bambakidis N, Tarr RW, Lanzieri CF, Zaidat OO, Suarez JI, et al. Benefits of perfusion MR imaging relative to diffusion MR imaging in the diagnosis and treatment of hyperacute stroke. *American Journal of Neuroradiology* 2001;22(5):915-21.

Sunshine 2004

Sunshine JL. CT, MR imaging, and MR angiography in the evaluation of patients with acute stroke. *Journal of Vascular and Interventional Radiology* 2004;15 Suppl(1 Pt 2):47-55.

Tei 1997

Tei H, Uchiyama S, Koshimizu K, Murakami H, Iwata M. [Accuracy of three-step diagnosis in discriminating subtypes of acute ischemic stroke] [Japanese]. *Rinsho Shinkeigaku - Clinical Neurology* 1997;37(1):21-5.

Toyoda 2001

Toyoda K, Ida M, Fukuda K. Fluid-attenuated inversion recovery intra-arterial signal: an early sign of hyperacute cerebral ischemia. *American Journal of Neuroradiology* 2001;22(6):1021-9.

van Everdingen 1998

van Everdingen KJ, van der Grond J, Kappelle LJ, Ramos LMP, Mali WPTM. Diffusion-weighted magnetic resonance imaging in acute stroke. *Stroke* 1998;29(9):1783-90.

Verro 2002

Verro P, Tanenbaum LN, Borden NM, Sen S, Eshkar N. CT angiography in acute ischemic stroke: preliminary results. *Stroke* 2002;33(1):276-8.

Von Kummer 2000

Von Kummer R, Gahn G. Comparison of diffusion-weighted MRI and CT in acute stroke. *Neurology* 2000;55(11):1760.

von Kummer 2001

von Kummer R, Bourquain H, Bastianello S, Bozzao L, Manelfe C, Meier D, et al. Early prediction of irreversible brain damage after ischemic stroke at CT. *Radiology* 2001;219(1):95-100.

Von Kummer 2002

Von Kummer R. MRI: the new gold standard for detecting brain hemorrhage? *Stroke* 2002;33(7):1748-9.

Wang 1997

Wang J, Yuan J, Tang J. [Clinical and image diagnostics of brain stem infarction] [Chinese]. *Chung-Hua Nei Ko Tsa Chih Chinese Journal of Internal Medicine* 1997;36(12):819-21.

Warach 1995

Warach S, Gaa J, Siewert B, Wielopolski P, Edelman R. Acute human stroke studied by whole brain echo planar diffusion-weighted magnetic resonance imaging. *Annals of Neurology* 1995;37:231-41.

Warach 1996

Warach S, Dashe JF, Edelman RR. Clinical outcome in ischemic stroke predicted by early diffusion-weighted and perfusion magnetic resonance imaging: a preliminary analysis. *Journal of Cerebral Blood Flow and Metabolism* 1996;16(1):53-9.

Wardlaw 2003

Wardlaw JM, Keir SL, Dennis MS. The impact of delays in computed tomography of the brain on the accuracy of diagnosis and subsequent management in patients with minor stroke. *Journal of Neurology, Neurosurgery, and Psychiatry* 2003;74(1):77-81.

Watanabe 2000

Watanabe T, Sugimoto K, Sato N, Matsuda W, Hattori A, Yanaka K, et al. [Clinical usefulness of diffusion-weighted magnetic resonance imaging in patients with ischemic cerebrovascular disease] [Japanese]. *No to Shinkei - Brain & Nerve* 2000;52(2):157-61.

Weber 2003

Weber C, Grzyska U, Lehner E, Adam G. Clinical relevance of cranial CT under emergency conditions - basic neuroradiologic investigations. *Rofo* 2003;175(5):654-62.

Wintermark 2005

Wintermark M, Fischbein NJ, Smith WS, Ko NU, Quist M, Dillon WP, et al. Accuracy of dynamic perfusion CT with deconvolution in detecting acute hemispheric stroke. *American Journal of Neuroradiology* 2005;26(1):104-12.

Wycliffe 2004

Wycliffe ND, Choe J, Holshouser B, Oyoyo UE, Haacke EM, Kido DK, et al. Reliability in detection of hemorrhage in acute stroke by a new three-dimensional gradient recalled echo susceptibility-weighted imaging technique compared to computed tomography: a retrospective study. *Journal of Magnetic Resonance Imaging* 2004;20(3):372-7.

Zivin 1997

Zivin JA. Diffusion-weighted MRI for diagnosis and treatment of ischemic stroke. *Annals of Neurology* 1997;41(5):567-8

Appendix 4 Summary of findings table

Results of studies on ischaemic stroke

Review question: *Comparison of diffusion-weighted magnetic resonance imaging with conventional computer tomography for the early detection of ischaemic brain lesions in patients suspected of stroke*

Patient population: adults suspected of acute stroke

Setting: hospital departments

Geographical location: studies were conducted in Europe (3 studies), the USA (3 studies), and in Australia (1 study)

Index test: diffusion-weighted magnetic resonance imaging (DWI) performed within 12 hours of stroke onset

Alternative test: computer tomography (CT) performed within 12 hours of stroke onset

Reference standard: clinical assessment and imaging follow up

Included studies: 7 comparative studies that evaluated DWI and CT in the same patients

Total number of patients assessed: 226

Limitations of included studies

- Limited number of included studies (7 studies); small sample sizes; presence of incorporation bias
- DWI and CT were evaluated in highly selected patient samples (patients with high probability of stroke), which therefore are not representative of the typical population of patients presenting with 'suspected acute stroke' to an emergency department (poor generalisability of results)
- The stroke vascular territory was not reported in the majority of included studies although it is likely that they enrolled patients with typical anterior circulation stroke
- Only a minority of the studied patients had severe strokes (in whom DWI might be contraindicated)
- The high proportion of mild strokes and reclassification of TIA cases with a positive DWI lesion as strokes might have inflated the DWI sensitivity estimate
- In most of the studies stroke mimics were not included
- In all but one study CT was performed before DWI (reducing the sensitivity of CT to detect ischaemia)

Summary of Findings Table - continued

CT results		DWI results		Summary effect (95% CI)
TP	73	TP	147	DWI sensitivity 0.99 (0.23 to 1.00)
FP	0	FP	5	DWI specificity 0.92 (0.83 to 0.97)
FN	88	FN	14	CT sensitivity 0.39 (0.16 to 0.69)
TN	65	TN	60	CT specificity 1.00 (0.94 to 1.00)
Total	226	Total	226	

Conclusions and comments

In the small cohort of included studies, DWI is more sensitive than CT - but not more specific - for the early detection of ischaemic stroke.

The small amount of data and the presence of methodological biases preclude any reliable calculation - from the sensitivity and specificity estimates of CT and DWI - of a positive or negative stroke diagnosis at different rates of stroke prevalence.

Applicability of tests in clinical practice

None of the studies addressed practicality. CT is known to be quicker to perform and more readily available in most emergency care settings than magnetic resonance imaging (MRI). MRI is contraindicated in patients with pacemakers and some metal implants. In acutely ill stroke patients it may be difficult to monitor the patient's condition while being MR scanned (and this increases the risk of any respiratory difficulty or cardiovascular compromise that develops during the scan which passes undetected and may have adverse effects for the patient). If the patient is confused or restless as a result of the stroke, the patient may not be able to co-operate for the longer scan times of MRI.

Costs

None of the studies included a cost-effectiveness evaluation. MRI is known to be more expensive than CT.

CI: confidence interval; CT: computed tomography; DWI: diffusion-weighted magnetic resonance imaging; FN: false negative; FP: false positive; MR/MRI: magnetic resonance imaging; TN: true negative; TP: true positive.

Appendix 5 Survey questionnaire

HOW TO SUMMARISE FINDINGS OF SYSTEMATIC REVIEWS OF DIAGNOSTIC ACCURACY

REVIEW TITLE:

Comparison of diffusion-weighted magnetic resonance imaging with conventional computer tomography for the early detection of ischaemic brain lesions in patients suspected of stroke

- 1. Please complete the ‘Demographic Information’ sheet**
- 2. Please read the first ‘summary’ document of the above review**
- 3. Please complete multiple-choice questions**
- 4. Please read the second ‘summary’ document of the above review**
- 5. Please complete multiple-choice questions**
- 6. Please complete third part of the questionnaire**

Thank you!

DEMOGRAPHIC INFORMATION

Please provide the following information:

Gender: Male Female

What is your clinical speciality?

General Radiology

Neurology

Neuroradiology

Stroke Medicine

Geriatric Medicine

Other (Please specify) _____

Please specify your postgraduate qualifications (e.g. MD, PhD)

Please specify how many years of clinical practice after MD/speciality training do you have _____ years

Please specify what proportion of your time (as a percentage of Full Time Equivalency) you spent routinely managing acute stroke patients in the last year _____%

Please specify your research involvement (as a percentage of Full Time Equivalency) _____%

Have you read the following published papers?

(Papers are listed in alphabetical order)

Barber PA, Darby DG, Desmond PM, Gerraty RP, Yang Q, Li T, et al. Identification of major ischemic change. Diffusion-weighted imaging versus computed tomography. Stroke 1999; 30: 2059-65

NO

YES

Bozzao A, Floris R, Giuliani V, Baviera ME, Montanaro M, Salvatore C, Simonetti G. Efficacia clinica della risonanza magnetica con sequenze pesate in diffusione nella valutazione dell'ischemia cerebrale acuta. La Radiologia Medica 1999; 98: 144-50

NO

YES

Chalela JA, Kidwell CS, Nentwich LN, Luby M, Butman JA, Demchuk AM, et al. Magnetic resonance imaging and computed tomography in emergency assessment of patients with suspected acute stroke: a prospective comparison. Lancet 2007; 369: 293-8

NO

YES

Gonzalez RG, Schaefer PW, Buonanno FS, Schwamm LH, Budzik RF, Rordorf G, et al. Diffusion-weighted MR imaging: diagnostic accuracy in patients imaged within 6 hours of stroke symptoms onset. Radiology 1999; 210: 155-62

NO

YES

Saur D, Kucinski T, Grzyska U, Eckert B, Eggers C, Niesen W, et al. Sensitivity and interrater agreement of CT and diffusion weighted MR imaging in hyperacute stroke. Am J Neuroradiol 2003; 24: 878-85

NO

YES

Sorensen AG, Buonanno FS, Gonzalez RG, Schwamm LH, Lev MH, Huang-Hellinger FR, et al. Hyperacute stroke: evaluation with combined multisection diffusion-weighted and hemodynamically weighted echo-planar MR imaging. Radiology 1996; 199: 391-401.

NO

YES

Urbach H, Flacke S, Keller E, Textor J, Berlis A, Hartmann A, et al. Detectability and detection rate of acute cerebral hemisphere infarcts on CT and diffusion-weighted MRI. Neuroradiology 2000; 42: 722-7

NO

YES

QUESTIONNAIRE – SECOND PART

Likert Visual Analogue Scale

How confident are you in understanding the following measures of diagnostic accuracy: sensitivity and specificity, likelihood ratios, predictive values? *Please tick the number in the scale below which better reflects your confidence level.*



Not confident at all



Very confident

0 ----- 1 ----- 2 ----- 3

QUESTIONNAIRE – SECOND PART

MULTIPLE-CHOICE QUESTIONS

The following questions are related to the information contained in the ‘Summary’ you have received. (Please tick only ONE answer for each question)

1. According to the information contained in the ‘summary’ you have received, would you maintain that in clinical practice:

- i. CT is more accurate than MRI for the early detection of ischaemic lesions
- ii. MRI is more accurate than CT for the early detection of ischaemic lesions
- iii. CT and MRI have similar accuracy for the early detection of ischaemic lesions
- iv. Neither CT nor MRI is accurate for the early detection of ischaemic lesions
- v. Evidence provided is not enough to inform clinical practice

2. According to the information contained in the ‘summary’ you have received, would you maintain that in clinical practice:

- i. CT and MRI have similar accuracy for the early detection of mild ischaemic strokes
- ii. CT is more accurate than MRI for the early detection of mild ischaemic strokes
- iii. Evidence provided is not enough to inform clinical practice
- iv. Neither CT nor MRI is accurate for the early detection of mild ischaemic strokes
- v. MRI is more accurate than CT for the early detection of mild ischaemic strokes

3. According to the information contained in the 'summary' you have received, would you maintain that in clinical practice:

- i. Evidence provided is not enough to inform clinical practice
- ii. Neither CT nor MRI is accurate for the early detection of non-stroke lesions
- iii. MRI is more accurate than CT for the early detection of non-stroke lesions
- iv. CT and MRI have similar accuracy for the early detection of non-stroke lesions
- v. CT is more accurate than MRI for the early detection of non-stroke lesions

4. According to the information contained in the 'summary' you have received, would you maintain that in clinical practice:

- i. MRI is more sensitive but not more specific than CT for the early detection of ischaemic lesions
- ii. Evidence provided is not enough to inform clinical practice
- iii. CT is more sensitive but not more specific than MRI for the early detection of ischaemic lesions
- iv. Neither CT nor MRI has a high sensitivity and specificity for the early detection of ischaemic lesions
- v. CT and MRI have similar sensitivity and specificity for the early detection of ischaemic lesions

5. According to the information contained in the 'summary' you have received, would you maintain that in clinical practice:

- i. Neither CT nor MRI should be used for the early detection of ischaemic lesions
- ii. Either CT or MRI should be used for the early detection of ischaemic lesions
- iii. Evidence provided is not enough to inform clinical practice
- iv. CT should be used instead of MRI for the early detection of ischaemic lesions
- v. MRI should be used instead of CT for the early detection of ischaemic lesions

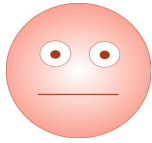
6. According to the information contained in the 'summary' you have received, would you maintain that in clinical practice:

- i. More diagnostic studies comparing CT versus MRI in acute stroke patients are needed
- ii. More diagnostic studies on CT in acute stroke patients are needed
- iii. More diagnostic studies on MRI in acute stroke patients are needed
- iv. No more diagnostic studies comparing CT versus MRI in acute stroke patients are needed
- v. I do not know

Questionnaire - Third Part

GENERAL QUESTIONS

1. The information in the Abstract is clearly presented?



I disagree



I agree

0 ----- 1 ----- 2 ----- 3

2. The information in the Summary of Findings Table is clearly presented?



I disagree



I agree

0 ----- 1 ----- 2 ----- 3

3. Did you find more informative the Abstract or the Summary of Findings Table?

4. If possible, what further information would you add to the Abstract?

5. If possible, what further information would you add to the Summary of Findings Table?

6. If you were asked to compare the information contained in the Abstract with that contained in the Summary of Findings table, would you say that: (*Please tick only one answer*)

- i.. Both the Abstract and the Summary of Findings table show the same information
- ii. Amount of information in either document is not enough to inform clinical practice
- iii. Information contained in the Summary of Findings table is more useful than that contained in the Abstract
- iv. Neither the Abstract nor the Summary of Findings table contained useful information
- v. Information contained in the Abstract is more useful than that contained in the Summary of Findings table

7. To interpret results, which of the following descriptions do you find easier to understand? (*Please circle your preference*)

- a) The test has a sensitivity of 99%
- or
- b) Out of 100 patients with a diagnosis of acute stroke imaged with the test, 1 will not show a true lesion

8. To interpret results, which of the following descriptions do you find easier to understand? (*Please circle your preference*)

- a) The test has a specificity of 92%
- or
- b) Out of 100 patients without a diagnosis of acute stroke imaged with the test, 8 will be wrongly identified as having a stroke

9. Do you have any further comments?

Appendix 6 Correct answers (underlined) for the multiple choice questions of the survey questionnaire

Multiple choice questions

1. According to the information contained in the ‘summary’ you have received, would you maintain that in clinical practice:

- i. CT is more accurate than MRI for the early detection of ischaemic lesions
- ii. MRI is more accurate than CT for the early detection of ischaemic lesions
- iii. CT and MRI have similar accuracy for the early detection of ischaemic lesions
- iv. Neither CT nor MRI is accurate for the early detection of ischaemic lesions
- v. Evidence provided is not enough to inform clinical practice

2. According to the information contained in the ‘summary’ you have received, would you maintain that in clinical practice:

- i. CT and MRI have similar accuracy for the early detection of mild ischaemic strokes
- ii. CT is more accurate than MRI for the early detection of mild ischaemic strokes
- iii. Evidence provided is not enough to inform clinical practice
- iv. Neither CT nor MRI is accurate for the early detection of mild ischaemic strokes
- v. MRI is more accurate than CT for the early detection of mild ischaemic strokes

3. According to the information contained in the ‘summary’ you have received, would you maintain that in clinical practice:

- i. Evidence provided is not enough to inform clinical practice
- ii. Neither CT nor MRI is accurate for the early detection of non-stroke lesions
- iii. MRI is more accurate than CT for the early detection of non-stroke lesions
- iv. CT and MRI have similar accuracy for the early detection of non-stroke lesions
- v. CT is more accurate than MRI for the early detection of non-stroke lesions

4. According to the information contained in the ‘summary’ you have received, would you maintain that in clinical practice:

- i. MRI is more sensitive but not more specific than CT for the early detection of ischaemic lesions
- ii. Evidence provided is not enough to inform clinical practice
- iii. CT is more sensitive but not more specific than MRI for the early detection of ischaemic lesions
- iv. Neither CT nor MRI has a high sensitivity and specificity for the early detection of ischaemic lesions
- v. CT and MRI have similar sensitivity and specificity for the early detection of ischaemic lesions

5. According to the information contained in the ‘summary’ you have received, would you maintain that in clinical practice:

- i. Neither CT nor MRI should be used for the early detection of ischaemic lesions
- ii. Either CT or MRI should be used for the early detection of ischaemic lesions
- iii. Evidence provided is not enough to inform clinical practice
- iv. CT should be used instead of MRI for the early detection of ischaemic lesions
- v. MRI should be used instead of CT for the early detection of ischaemic lesions

6. According to the information contained in the 'summary' you have received, would you maintain that in clinical practice:

- i. More diagnostic studies comparing CT versus MRI in acute stroke patients are needed
- ii. More diagnostic studies on CT in acute stroke patients are needed
- iii. More diagnostic studies on MRI in acute stroke patients are needed
- iv. No more diagnostic studies comparing CT versus MRI in acute stroke patients are needed
- v. I do not know