

Accuracy of magnetic resonance imaging for the diagnosis of multiple sclerosis: systematic review

Penny Whiting, Roger Harbord, Caroline Main, Jonathan J Deeks, Graziella Filippini, Matthias Egger, Jonathan A C Sterne

Abstract

Objective To determine the accuracy of magnetic resonance imaging criteria for the early diagnosis of multiple sclerosis in patients with suspected disease.

Design Systematic review.

Data sources 12 electronic databases, citation searches, and reference lists of included studies.

Review methods Studies on accuracy of diagnosis that compared magnetic resonance imaging, or diagnostic criteria incorporating such imaging, to a reference standard for the diagnosis of multiple sclerosis.

Results 29 studies (18 cohort studies, 11 other designs) were included. On average, studies of other designs (mainly diagnostic case-control studies) produced higher estimated diagnostic odds ratios than did cohort studies. Among 15 studies of higher methodological quality (cohort design, clinical follow-up as reference standard), those with longer follow-up produced higher estimates of specificity and lower estimates of sensitivity. Only two such studies followed patients for more than 10 years. Even in the presence of many lesions (> 10 or > 8), magnetic resonance imaging could not accurately rule in multiple sclerosis (likelihood ratio of a positive test result 3.0 and 2.0, respectively). Similarly, the absence of lesions was of limited utility in ruling out a diagnosis of multiple sclerosis (likelihood ratio of a negative test result 0.1 and 0.5).

Conclusions Many evaluations of the accuracy of magnetic resonance imaging for the early detection of multiple sclerosis have produced inflated estimates of test performance owing to methodological weaknesses. Use of magnetic resonance imaging to confirm multiple sclerosis on the basis of a single attack of neurological dysfunction may lead to over-diagnosis and over-treatment.

Introduction

Recent criteria state that a diagnosis of multiple sclerosis should be based on two attacks of neurological dysfunction occurring at different times and affecting different parts of the central nervous system.¹ Magnetic resonance imaging may assist in earlier diagnosis of the disease by enabling visualisation of clinically silent lesions in the brain. The McDonald 2001 criteria² allow

diagnosis of multiple sclerosis after one clinical attack if the patient also meets criteria for a positive result on a magnetic resonance imaging scan.

We carried out a systematic review to estimate the accuracy of different magnetic resonance imaging criteria for the early diagnosis of multiple sclerosis in patients presenting with suspected disease.

Methods

We searched 12 databases from inception until September or November 2004, undertook a citation search on the article reporting the McDonald 2001 criteria,² screened reference lists of included studies, and assessed studies included in the National Institute for Health and Clinical Excellence guidelines for multiple sclerosis.³

Studies had to compare magnetic resonance imaging (or criteria incorporating such imaging) to a reference standard for the diagnosis of multiple sclerosis and report sufficient data to construct a 2×2 table of test performance. For multiple publications of a study, we included the one with data for the longest follow-up. We also included separate publications that reported on different criteria for magnetic resonance imaging or separate results for relevant patient subgroups.

Studies were assessed for methodological quality using the QUADAS (quality assessment of diagnostic accuracy studies) tool (see bmj.com for scoring of items).⁴ We grouped studies according to patient spectrum: prospective cohort studies of patients with suspected multiple sclerosis, and studies of other designs.

Data analysis

From each 2×2 table we computed sensitivity, specificity, and likelihood ratios.⁵ We plotted results from all studies on a receiver operating characteristic plot. To compare accuracy of cohort and other studies we selected the median diagnostic odds ratio for each

Editorial by
Rothwell

MRC Health Services Research
Collaboration,
Department of
Social Medicine,
Bristol BS8 2PR
Penny Whiting
research fellow

Roger Harbord
research associate
Jonathan A C
Sterne
reader in medical
statistics and
epidemiology

Centre for Reviews
and Dissemination,
University of York
Caroline Main
research fellow

Centre for Statistics
in Medicine,
Wolfson College,
Oxford

Jonathan J Deeks
senior medical
statistician

Unit of
Neuroepidemiology,
Istituto Nazionale
Neurologico "Carlo
Besta," Milan, Italy
Graziella Filippini
head

Department of
Social and
Preventive
Medicine,
University of Bern,
Switzerland
Matthias Egger
professor of
epidemiology and
public health

Correspondence to:
P Whiting
penny.whiting@
bristol.ac.uk

BMJ 2006;332:875-8



Additional information and references w1-w43 are on bmj.com



This is the abridged version of an article that was posted on bmj.com on 24 March 2006: <http://bmj.com/cgi/doi/10.1136/bmj.38771.583796.7C>

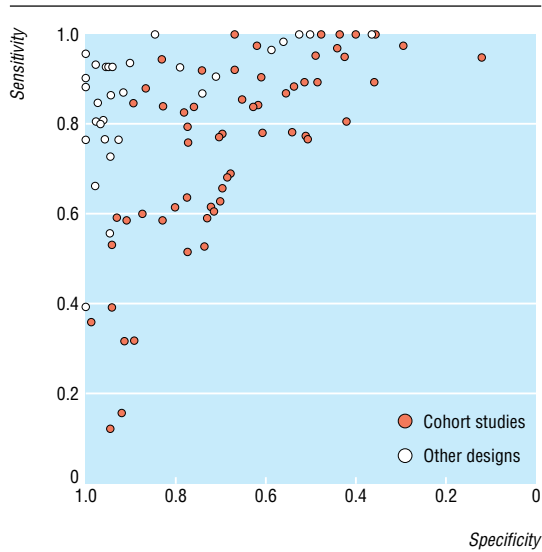


Fig 1 Receiver operating characteristic plots for cohort studies and for studies of other designs

study. We used random effects meta-analysis to obtain summary diagnostic odds ratios in each group and carried out a permutation test⁶ to obtain a P value for comparison.

To assess the effect of duration of follow-up on accuracy and threshold, we used the hierarchical summary receiver operating characteristic method.⁷ We drew separate receiver operating characteristic plots for studies that evaluated commonly reported magnetic resonance imaging criteria (Barkhof, Paty, and Fazekas), and the McDonald 2001 criteria.

Further analysis was restricted to cohort studies with at least 10 years' clinical follow-up. For each of these we produced separate receiver operating characteristic plots.

Results

Forty publications reporting the results of 29 studies were included (see *bmj.com* for flow diagram)^{w1-w43}, 18 cohort studies and 11 studies of other designs (see tables on *bmj.com*). The studies differed according to population, quality, magnetic resonance imaging protocol, and criteria used to define a positive test result. Publication dates ranged from 1986 to 2003. Over this time improvements occurred in magnetic resonance imaging (see table A on *bmj.com*).

Study quality was generally poor (see table B and figure 2 on *bmj.com*). Three cohort studies were susceptible to incorporation bias as magnetic resonance imaging contributed to the final diagnosis.¹⁰⁻¹³ All other cohort studies used clinical follow-up alone as the reference standard. Most used the Poser criteria,⁸ although some used the McDonald 1977 criteria.²⁻⁹ The McDonald 1977 criteria, based on clinical information alone, are not the same as the McDonald 2001 criteria, which incorporate magnetic resonance imaging.⁴

Cohort studies produced lower estimated sensitivity and specificity than studies of other designs (fig 1): pooled diagnostic odds ratio 9 (95% confidence interval 5 to 16) for cohort studies and 213 (85 to 535) for

studies of other designs ($P < 0.001$). Further analysis was restricted to the 15 cohort studies that used a diagnosis of clinically definite multiple sclerosis as the reference standard.

The average duration of follow-up ranged from seven months to 14 years. Sufficient data were available to investigate the effects of duration of follow-up for presence of one or more lesions and presence of one or more non-clinical lesions. Figure 2 shows that studies with longer follow-up produced higher estimated specificity and lower estimated sensitivity; evidence from the hierarchical summary receiver operating characteristic analysis ($p = 0.074$) supported this.

The longest average duration of follow-up was three years in studies assessing the Barkhof, Fazekas, and McDonald 2001 criteria, and six years in studies assessing the Paty criteria (see *bmj.com*). Both positive (< 5) and negative likelihood ratios for these criteria (range 0.2 to 0.5) suggested that they are of limited utility for ruling in or out the development of multiple sclerosis within three to six years. Positive likelihood ratios for the McDonald 2001 criteria (range 2.7 to 8.7) suggest greater potential for predicting multiple sclerosis within three years than the criteria based on magnetic resonance imaging alone, but negative likelihood ratios (0.1 in one study, 0.2 to 0.5 in three studies) suggest that they are of limited utility for ruling out the development of multiple sclerosis within three years.¹⁵⁻¹⁸

Only two studies, one from the United States¹⁹ and one from England,²⁰ followed patients for more than 10 years, long enough to be reasonably confident that almost all patients who ever would be had been diagnosed as having multiple sclerosis. The US study included 351 patients with optic neuritis; follow-up of more than 10 years was available for 302 (86%) of these. The English study included 135 patients with a range of presenting symptoms, of whom 71 (53%) were included in the final evaluation. Both evaluated thresholds based on the number of lesions present on magnetic resonance imaging of the brain (fig 3).

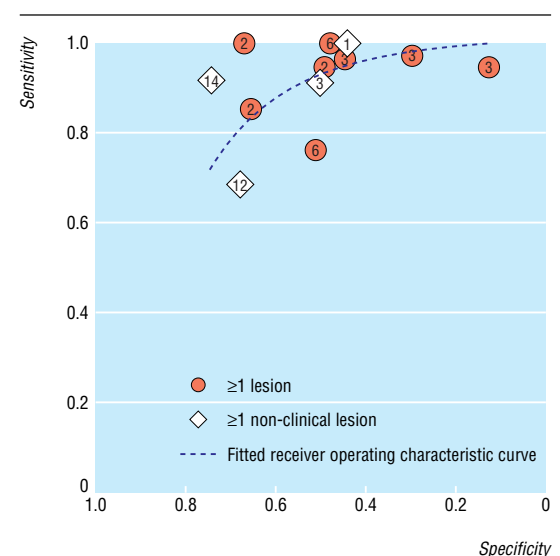


Fig 2 Receiver operating characteristic plots for studies included in hierarchical summary receiver operating characteristic analysis. Numbers are duration of follow-up in years

Estimated sensitivity was higher with fewer lesions but specificity was lower (fig 3). The English study produced higher estimates of sensitivity. Positive likelihood ratios for the presence of various numbers of lesions ranged from 2.0 to 3.4. Assuming a pretest probability of multiple sclerosis of 60% this is equivalent to a post-test probability of 75%-84%, suggesting that magnetic resonance imaging is of limited utility for ruling in multiple sclerosis at any threshold. Negative likelihood ratios ranged from 0.1 to 0.9 but were greater than 0.5 for all but one of the thresholds in the English study. This is equivalent to modifying a pretest probability of 60% to give a post-test probability of multiple sclerosis of 43%-57%, suggesting that magnetic resonance imaging is also of limited utility in ruling out a diagnosis of multiple sclerosis.

Discussion

Use of magnetic resonance imaging to confirm multiple sclerosis on the basis of a single attack of neurological dysfunction may lead to over-diagnosis and over-treatment. Many studies in our systematic review produced inflated estimates of test performance owing to methodological weaknesses.

Only two cohort studies included at least 10 years' follow-up. These suggested that the role of magnetic resonance imaging for ruling in or ruling out multiple sclerosis is limited. Studies lacking an appropriate patient spectrum overestimated both sensitivity and specificity. Studies of shorter clinical follow-up overestimated sensitivity and underestimated specificity. The Fazekas, Barkhof, and Paty criteria showed poor accuracy for predicting the development of multiple sclerosis within three to six years. The limited data on the McDonald 2001 criteria suggest that these have some potential to rule in the development of multiple sclerosis within three years. Neither the specific magnetic resonance imaging criteria nor the McDonald 2001 criteria were evaluated in studies with long term follow-up, and so it is not possible to determine their accuracy for the diagnosis of multiple sclerosis.

Considerable weaknesses existed in the primary studies included in the review. The only reference standard for the diagnosis of multiple sclerosis is long term clinical follow-up. Most studies followed patients for short periods or included an inappropriate patient spectrum, such as people with clinically definite multiple sclerosis and a control group of people known not

What is already known on this topic

Magnetic resonance imaging has been recommended in the diagnosis of multiple sclerosis

The diagnostic accuracy of such imaging has been assessed but a systematic review has not previously been carried out

What this study adds

Magnetic resonance imaging is of limited utility for both ruling in and ruling out multiple sclerosis

Studies with shorter follow-up tended to produce higher estimates of sensitivity and lower estimates of specificity compared with longer term studies

to have the disease. That such studies tend to exaggerate the accuracy²¹ of magnetic resonance imaging in the diagnosis of multiple sclerosis is to be expected; people with more advanced disease are more likely to have lesions on scans than those in the early stages of the disease.

The McDonald 2001 criteria incorporate the Barkhof criteria to define a positive magnetic resonance imaging scan.² The article reporting the McDonald 2001 criteria² refers to a small number of studies to justify its selection of the Barkhof criteria for this purpose. All these had methodological weaknesses. This paper was published before the two long term cohort studies from England and the United States.¹⁹⁻²⁰ Those two studies produced differing results, with the US study reporting lower estimates of sensitivity for similar thresholds for magnetic resonance imaging. These differences may reflect the smaller sample size of the English study or large proportion of dropouts in this study. Alternatively, magnetic resonance imaging may be more accurate in patients with brainstem or spinal cord symptoms than in patients with optic neuritis.

The main clinical question is whether magnetic resonance imaging should be included in the investigation of patients with multiple sclerosis. Such imaging is not simply ordered to increase the certainty of the diagnosis: other possible reasons include ruling out differential diagnoses such as brain tumours, providing a baseline for monitoring disease progression, patient request, and patient reassurance. Rather than the accuracy of magnetic resonance imaging alone in diagnosing multiple sclerosis, the issue of clinical relevance is, arguably, the added value of magnetic resonance imaging in diagnosing multiple sclerosis compared with the patient's history and clinical examination alone.²² None of the located studies tackled this issue. A further limitation of published studies is that they tend to dichotomise the results of magnetic resonance imaging into positive or negative scans. The use of a scale based on features present on a scan should be considered. This is probably consistent with how the results of magnetic resonance imaging are interpreted in practice.

In conclusion, magnetic resonance imaging is a relatively poor test for both ruling in and ruling out multiple sclerosis. In clinical practice a false positive

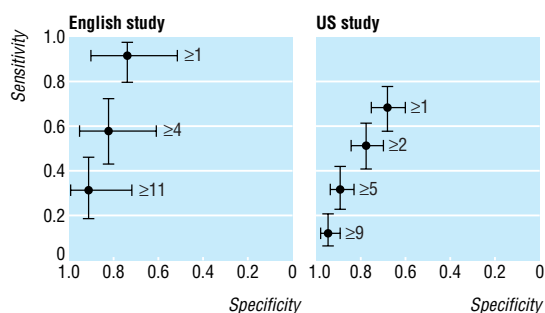


Fig 3 Sensitivity plotted against specificity (95% confidence intervals) for different thresholds (number of lesions shown next to plots) reported in English and US studies²⁻³

diagnosis of multiple sclerosis is potentially more dangerous than a false negative one because it implies unnecessary successive tests and treatments, or needless anxiety and psychological distress for the patient. Multiple sclerosis remains predominantly a clinical diagnosis.

Contributors: See *bmj.com*.

Funding: This work was supported by the Medical Research Council Health Services Research Collaboration. The authors' work was independent of the funders. JJD is funded by a senior research fellowship in evidence synthesis from the UK Department of Health.

Competing interests: None declared.

Ethical approval: Not required.

- 1 Poser CM, Brinar VV. Diagnostic criteria for multiple sclerosis: an historical review. *Clin Neurol Neurosurg* 2004;106:147-58.
- 2 McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the international panel on the diagnosis of multiple sclerosis. *Ann Neurol* 2001;50:121-7.
- 3 National Collaborating Centre for Chronic Conditions. *Multiple sclerosis. National clinical guideline for diagnosis and management in primary and secondary care*. London: Royal College of Physicians, 2004.
- 4 Whiting P, Rutjes AW, Dinnes J, Reitsma J, Bossuyt PM, Kleijnen J. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol Assess* 2004;8:1-234.
- 5 Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ* 2004;329:168-9.
- 6 Higgins JP, Thompson SG. Controlling the risk of spurious findings from meta-regression. *Stat Med* 2004;23:1663-82.
- 7 Rutter CA, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001;20:2865-84.
- 8 Macaskill P. Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. *J Clin Epidemiol* 2004;57:925-32.
- 9 Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, Ebers GC, et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann Neurol* 1983;13:227-31.

- 10 McDonald WI, Halliday AM. Diagnosis and classification of multiple sclerosis. *Br Med Bull* 1977;33:4-9.
- 11 Beer S, Rosler KM, Hess CW. Diagnostic value of paraclinical tests in multiple sclerosis: relative sensitivities and specificities for reclassification according to the Poser committee criteria. *J Neurol Neurosurg Psychiatry* 1995;59:152-9.
- 12 Miller DH, Ormerod IE, McDonald WI, MacManus DG, Kendall BE, Kingsley DP, et al. The early risk of multiple sclerosis after optic neuritis. *J Neurol Neurosurg Psychiatry* 1988;51:1569-71.
- 13 Mushlin AI, Detsky AS, Phelps CE, O'Connor PW, Kido DK, Kucharczyk W, et al. The accuracy of magnetic resonance imaging in patients with suspected multiple sclerosis. The Rochester-Toronto Magnetic Resonance Imaging Study Group. *JAMA* 1993;269:3146-51.
- 14 Barkhof F, Filippi M, Miller DH, Scheltens P, Campi A, Polman CH, et al. Comparison of MRI criteria at first presentation to predict conversion to clinically definite multiple sclerosis. *Brain* 1997;120:2059-69.
- 15 Dalton CM, Brex PA, Miszkiel KA, Hickman SJ, MacManus DG, Plant GT, et al. Application of the new McDonald criteria to patients with clinically isolated syndromes suggestive of multiple sclerosis. *Ann Neurol* 2002;52:47-53.
- 16 Dalton CM, Brex PA, Miszkiel KA, Fernando K, MacManus DG, Plant GT, et al. New T2 lesions enable an earlier diagnosis of multiple sclerosis in clinically isolated syndromes. *Ann Neurol* 2003;53:673-6.
- 17 Di Legge S, Piattella MC, Pantano P, Pestalozza IF, Nucciarelli W, Bozzao L, et al. The impact of revised McDonald criteria in predicting multiple sclerosis. *Neurology* 2002;58:A173.
- 18 Tintore M, Rovira A, Rio J, Nos C, Grive E, Sastre-Garriga J, et al. New diagnostic criteria for multiple sclerosis: application in first demyelinating episode. *Neurology* 2003;60:27-30.
- 19 Beck RW, Trobe JD, Moke PS, Gal RL, Xing D, Bhatti MT, et al. High- and low-risk profiles for the development of multiple sclerosis within 10 years after optic neuritis: experience of the optic neuritis treatment trial. *Arch Ophthalmol* 2003;121:944-9.
- 20 Brex PA, Ciccarelli O, O'Riordan JI, Sailer M, Thompson AJ, Miller DH. A longitudinal study of abnormalities on MRI and disability from multiple sclerosis. *New Engl J Med* 2002;346:158-64.
- 21 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. *Ann Intern Med* 2004;140:189-202.
- 22 Moons KG, Biesheuvel CJ, Grobbee DE. Test research versus diagnostic research. *Clin Chem* 2004;50:473-6.

(Accepted 2 February 2006)

doi 10.1136/bmj.38771.583796.7C

Clinical value of the metabolic syndrome for long term prediction of total and cardiovascular mortality: prospective, population based cohort study

Johan Sundström, Ulf Risérus, Liisa Byberg, Björn Zethelius, Hans Lithell, Lars Lind

Abstract

Objectives To find out if the presence of the metabolic syndrome increases the risk of subsequent total and cardiovascular mortality, taking into account established risk factors for cardiovascular disease.

Design Prospective cohort study.

Setting General population.

Participants A community based sample of 2322 men followed since 1970 for a maximum of 32.7 years, investigated at ages 50 and 70.

Main outcome measures The relations of the metabolic syndrome defined by the national cholesterol education programme (NCEP) of the US National Heart, Lung, and Blood Institute or criteria of the World Health Organization (WHO) to subsequent total and cardiovascular mortality.

Results When adding the metabolic syndrome to models with established risk factors for cardiovascular disease (smoking, diabetes, hypertension, and serum cholesterol) at age 50, presence of the metabolic

syndrome as defined in the NCEP significantly predicted total and cardiovascular mortality (Cox proportional hazard ratios 1.36, 95% confidence interval 1.17 to 1.58; and 1.59, 1.29 to 1.95, respectively). The metabolic syndrome added prognostic information to that of the established risk factors for cardiovascular disease (likelihood ratio tests, $P < 0.0001$ for both outcomes). Similar results were obtained in a subsample without diabetes or manifest cardiovascular disease.

Conclusions In a large, community based sample of middle aged men, the presence of the metabolic syndrome according to the definition of the NCEP gave long term prognostic information regarding

Department of Medical Sciences, Uppsala University Hospital; SE-751 85 Uppsala, Sweden
 Johan Sundström
associate professor of cardiovascular epidemiology
 Lars Lind
professor of cardiovascular medicine

continued over

BMJ 2006;332:878-82



A table is on *bmj.com*



This is the abridged version of an article that was posted on *bmj.com* on 20 March 2006: <http://bmj.com/cgi/doi/10.1136/bmj.38766.624097.1F>