Cite this article as: BMJ, doi:10.1136/bmj.38771.583796.7C (published 24 March 2006)

Research

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 than10 years. Even in the presence of many lesions (>10 or >8), magnetic resonance imaging could not accurately rule multiple sclerosis in (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

Diagnosis of multiple sclerosis is based on the principle of dissemination in both time and space. Recent criteria state that patients should experience two attacks of neurological dysfunction, such as optic neuritis, transverse myelitis, double vision, or numbness and tingling of the leg, occurring at different points in time and affecting different parts of the central nervous system—that is, signs or symptoms that cannot be attributable to a single lesion.[1] Many years may elapse between first and second attacks, and not all patients who experience a first attack develop multiple sclerosis. In a study of patients with optic neuritis, a common presenting symptom of multiple sclerosis, 38% developed the disease by 10 years; of these, 50% received their diagnosis more than three years after presentation and 28% more than five years after presentation.[2] In a study of patients presenting with clinically isolated syndromes (optic, spinal cord, or brain symptoms) 68% of patients had developed multiple sclerosis by 14 years, the proportions being similar for the different presenting symptoms.[3]

Magnetic resonance imaging may assist in earlier diagnosis of multiple sclerosis by enabling visualisation of lesions in the brain that are clinically silent. The McDonald 2001 criteria for the diagnosis of multiple sclerosis[4] allow an early diagnosis of multiple sclerosis to be made after one clinical attack if the patient also meets criteria for a positive result on a magnetic resonance imaging scan. The McDonald criteria have been adopted in England and Wales by the National Institute for Health and Clinical Excellence (NICE),[5] but they are not universally accepted.[1] Evidence shows that patients' wellbeing is affected by early diagnosis,[6–9] usually in a beneficial way but also occasionally in a negative way—for example, through increased insurance premiums and discrimination in the workplace.[10] Earlier diagnosis of multiple sclerosis could mean the availability of earlier treatment, such as the disease modifying therapies interferon beta and glatiramer acetate, provided under the "risk sharing scheme" in the United Kingdom (www.dh.gov.uk/assetRoot/04/01/22/14/04012214.pdf).

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, to investigate whether magnetic resonance imaging has the potential to alter diagnoses and patient management.

Methods

We identified studies, published and unpublished, by searching 12 databases from inception until September or November 2004. Search terms were "multiple sclerosis" combined with "magnetic resonance imaging" or "MRI". No language restrictions were applied. We undertook a citation search on the article reporting the McDonald 2001 criteria,[4] screened reference lists of included studies, and assessed studies included in the NICE multiple sclerosis guidelines.[5]

Studies were eligible that compared magnetic resonance imaging (or diagnostic criteria incorporating such imaging) to a reference standard for the diagnosis of multiple sclerosis and reported sufficient data to enable a 2×2 table of test



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

BMJ: first published as 10.1136/bmj.38771.583796.7C on 24 March 2006. Downloaded from http://www.bmj.com/ on 29 March 2024 by guest. Protected by copyright.

performance to be constructed. If studies were reported more than once, we included the publication that provided 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.

Two reviewers independently screened titles and abstracts for relevance. Screening for inclusion, data extraction, and quality assessment were carried out by one reviewer and checked by a second. Studies were assessed for methodological quality against the QUADAS (quality assessment of diagnostic accuracy studies) criteria.[11] (See bmj.com for a summary of how items were scored.) One item, the avoidance of disease progression bias, was omitted as it was not relevant to this topic. We grouped studies according to patient spectrum: prospective cohort studies that enrolled patients with suspected multiple sclerosis, and studies of other designs.

Data analysis

From each 2×2 table we computed sensitivity, specificity, and likelihood ratios, which combine data on sensitivity and specificity to give an indication of a test's ability to rule in or rule out a condition.[12]

We plotted all results from all included studies on a receiver operating characteristic plot of sensitivity against specificity, with the specificity axis reversed. To compare accuracy of cohort and other studies we selected the result with the median diagnostic odds ratio (defined as the odds of positivity among people with the disease, divided by the odds of positivity among people without the disease) for each study. We used random effects meta-analysis to obtain summary diagnostic odds ratios in each group, and we carried out a permutation test[13] to obtain a P value for their comparison. We restricted all further analyses to cohort studies that used a reference standard diagnosis of clinically definite multiple sclerosis, arrived at solely by clinical data.

As a final diagnosis of multiple sclerosis may be reached many years after a patient first presents with possible disease, we investigated the effect of duration of follow-up on estimates of diagnostic accuracy. We used the hierarchical summary receiver operating characteristic method proposed by Rutter and Gatsonis[14] to assess the effect of duration of follow-up on overall accuracy and threshold. An association with threshold would indicate that sensitivity increased as specificity decreased, or vice versa. We drew separate receiver operating characteristic plots for studies that evaluated commonly reported magnetic resonance imaging criteria, the Barkhof, Paty, and Fazekas criteria, and the McDonald 2001 criteria, which combine clinical information with findings on magnetic resonance imaging.

Further analysis was restricted to cohort studies with at least 10 years' clinical follow-up. We produced separate receiver operating characteristic plots for each of these studies and compared areas under the curves. The statistical software package Stata release 9 was used for all analyses, except the hierarchical summary receiver operating characteristic model, which was fitted in SAS.[15]

Results

Figure 1 shows the flow of studies through the review. Sixty one publications met the inclusion criteria, 21 of which were earlier reports of included studies and were not extracted.[w1-w43] Forty publications reporting the results of 29 studies (some reported results for different magnetic resonance imaging criteria, for imaging of the spine rather than the brain, or for patient subgroups) were included. Sample sizes were generally small (median 70), ranging from 15 to 1500 patients. The proportions

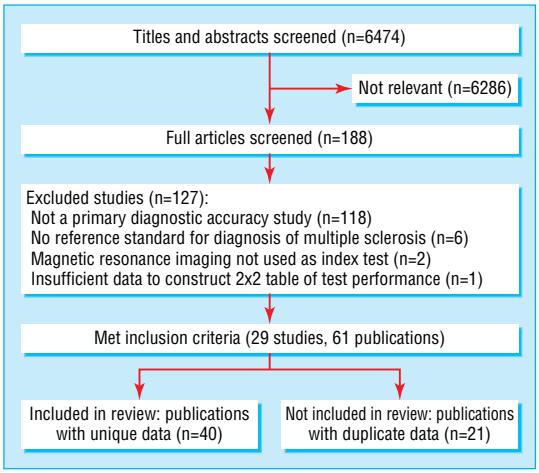


Fig 1 Flow of studies through review process

of dropouts ranged from 0 to 58% (median 4%), increasing with length of follow-up. Table 1 provides details of the 29 publications reporting the results of 18 cohort studies. Most of these studies used clinical follow-up as the reference standard. Most used the Poser criteria,[16] although some used the McDonald 1977 criteria.[17] The McDonald 1977 criteria, based on clinical information alone, are not the same as the McDonald 2001 criteria, which incorporate magnetic resonance imaging.[4] Table 2 provides details of the 11 studies of other designs. The studies differed according to population, quality, magnetic resonance imaging protocol, and criteria used to define a positive test result. Cohort studies varied in their inclusion criteria; some included only patients presenting with a particular clinically isolated syndrome (for example, optic neuritis or a spinal cord syndrome), whereas others included all patients being evaluated for possible multiple sclerosis. Publication dates ranged from 1986 to 2003. Over this time improvements occurred in magnetic resonance imaging technology; this is reflected in differences in scanning protocols (see table A on bmj.com).

Figure 2 summarises the results of the quality assessment (see table B on bmj.com for results of individual studies). Study quality was generally poor: only four QUADAS items were met by over 70% of studies (avoidance of partial and differential verification bias and reporting of uninterpretable results and withdrawals). Studies scored badly on three items: blinding, the use of an appropriate reference standard, and the availability of clinical information. Four publications, reporting results from three cohort studies, were susceptible to incorporation bias as magnetic resonance imaging contributed to the final diagnosis.[18–21] Three of these used a combination of clinical follow-up and paraclinical tests as the reference standard,[19–21] the other relied on paraclinical tests alone.[18] All other cohort studies used clinical follow-up alone as the reference standard.

Figure 3 shows that cohort studies produced lower estimated sensitivity and specificity than studies of other designs. The pooled diagnostic odds ratio was 9 (95% confidence interval 5 to 16) for cohort studies and 213 (85 to 535) for studies of other designs ($P < 0.001$, permutation test). Further analysis was restricted to the 15 cohort studies that used a diagnosis of clinically definite multiple sclerosis, arrived at by clinical information alone, as the reference standard.

The average duration of follow-up ranged from seven months to 14 years. The only criteria for which sufficient data were available to investigate the effects of duration of follow-up were presence of one or more lesions and presence of one or

BMJ: first published as 10.1136/bmj.38771.583796.7C on 24 March 2006. Downloaded from http://www.bmj.com/ on 29 March 2024 by guest. Protected by copyright.

Table 1 Study details and results of cohort studies recruiting patients with suspected multiple sclerosis (MS)

Author (year), country	No of patients* (% female)	Mean (range) age (years)	Presenting symptoms	Criteria for MS	Mean follow-up (range) in months	% with MS	MRI criteria	Sensitivity, specificity (%)	Positive, negative likelihood ratios
Lee (1991)[w1], Canada 1	184 (67)	43 (12-79)	NR	Schumacher[w2]	25 (NR)	30	≥1 lesions	95, 49	1.8, 0.1
							Paty	84, 63	2.2, 0.3
Ford (1992)[w3], Canada 2	15 (53)	33 (22-44)	Spinal cord lesion (n=15)	Poser[w4]	38.5 (NR)	80	≥1 non-clinical lesions	91, 50	1.8, 0.2
Reese (1986)[w5], Canada 3	47 (NR)	NR	NR	Diagnosed by referring neurologist (no details)		68	≥1 T2 lesions	88, 87	5.5, 0.2
Frederiksen (1989)[w6], Denmark	60 (70)	31 (12-53)	Optic lesion (n=60)	NR	Median 11 (1-28)	20	≥1 non-clinical lesions	100, 44	1.7, 0.1
Brex (2002)[w7], England 1	71 (62)	32 (13-49)	Optic lesion (n=36), spinal cord lesion (n=21), brainstem lesion (n=14)	Poser[w4]	168 (150-202)	68	≥1 non-clinical T2 lesions	92, 74	3.4, 0.1
							≥4 non-clinical T2 lesions	58, 83	3.1, 0.5
							>10 non-clinical T2 lesions	31, 91	3.0, 0.8
Filippi (1994)[w8], England 1	84 (70)	31 (13-50)	Optic lesion (n=40), spinal cord lesion (n=28), brainstem lesion (n=16)		63 (43-84)	40	Initial lesion load >1.23 cm^3	53, 94	7.7, 0.5
Miller (1988)[w9], England 1	53 (62)	33 (16-48)	Optic lesion (n=69)		12 (5-30)	28	≥1 non-clinical T2 lesions	80, 42	1.4, 0.5
Miller (1989)[w10], England 1	56 (63)	32 (13-49)	Spinal cord lesion (n=33)		14.7 (6-31)	30	≥1 non-clinical T2 lesions	90, 61	2.2, 0.2
			Brainstem lesion (n=23)			35	≥1 non-clinical T2 lesions	100, 40	1.6, 0.1
Sharief (1991)[w11], England 1	45 (62)	37.3 (<50)	Spinal cord lesion (n=25), brainstem lesion (n=20)	Poser[w4]†, paraclinical data contributed to diagnosis	18 (NR)	49	Paty	77, 70	2.4, 0.3
Brex (2001)[w12] (brain MRI), England 2	68 (57)	31 (16-50)	Optic lesion (n=46), spinal cord lesion (n=6), brainstem lesion (n=16)	Poser[w4]†	>12	26	≥1 non-clinical T2 lesions	89, 36	1.4, 0.4
							≥4 non-clinical T2 lesions	78, 54	1.7, 0.4
							≥9 non-clinical T2 lesions	61, 72	2.1, 0.6
							≥1 non-clinical enhancing lesions	61, 80	2.9, 0.5
							≥1 enhancing lesions at baseline and one new lesion at three months	39, 94	5.8, 0.6
							≥1 non-clinical T2 lesions at baseline and one new lesion at three months	83, 76	3.3, 0.2
Brex (1999)[w13] (spinal MRI), England 2	50 (57)	30 (16-49)	Optic lesion (n=46), spinal cord lesion (n=6), brainstem lesion (n=16)	Poser[w4]	12 (12-19)	26	≥1 non-clinical T2 lesions	77, 70	2.5, 0.4
							≥1 non-clinical enhancing lesions	15, 92	1.9, 0.9
Dalton (2003)[w14], England 2	56 (54)	32 (17-50)	Optic lesion (n=37), spinal cord lesion (n=5), brainstem lesion (n=14)	Poser[w4]	NR (>36)	34	New T2 lesion at three months	84, 89	7.0, 0.2
							McDonald 2001 criteria at three months	58, 95	8.7, 0.5
							McDonald 2001 criteria or new T2 lesions at three months	74, 92	7.9, 0.3

BMJ: first published as 10.1136/bmj.38771.583796.7C on 24 March 2006. Downloaded from http://www.bmj.com/ on 29 March 2024 by guest. Protected by copyright.

Author (year), country	No of patients* (% female)	Mean (range) age (years)	Presenting symptoms	Criteria for MS	Mean follow-up (range) in months	% with MS	MRI criteria	Sensitivity, specificity (%)	Positive, negative likelihood ratios
Dalton (2002)[w15], England 2	50 (59)	31 (16-50)	Optic lesion (n=90), spinal cord lesion (n=10), brainstem lesion (n=19)	Poser[w4]	37 (29-67)	38	≥1 non-clinical lesions (brain or spinal cord)	100, 35	1.5, 0.1
							Barkhof (brain or spinal cord MRI)	79, 77	3.3, 0.3
							Barkhof (brain MRI only)	63, 77	2.7, 0.5
							McDonald 2001 MRI at three months	59, 93	6.8, 0.5
							McDonald 2001 MRI at 12 months	83, 83	4.4, 0.2
							McDonald 2001 criteria at three months	65, 93	7.4, 0.4
							McDonald 2001 criteria at 12 months	94, 83	5.0, 0.1
Barkhof (1997)[w16], Holland 1	74 (NR)	NR	Optic lesion (n=40), spinal cord lesion (n=22), brainstem lesion (n=12)	Poser[w4]	39 (23-69)	45	≥1 T2 lesions	97, 29	1.4, 0.1
							Barkhof	82, 78	3.6, 0.2
							Fazekas	88, 54	1.9, 0.2
							Paty	88, 54	1.9, 0.2
Ghezzi (1999)[w17], Italy 1	102 (71)	29 (NR)	Optic lesion (n=143)	Poser[w4]	76 (>48)	36	≥1 white matter lesions	100, 48	1.9, 0
							Paty	86, 55	1.9, 0.3
Di Legge (2002)[w18], Italy 2	53 (66)	30.5 (NR)	NR	NR	>18	36	McDonald 2001 criteria at three months	89, 68	2.7, 0.2
Paolino (1996)[w19], Italy 3	44 (70)	30 (<50)	Spinal cord lesion (n=22), brainstem lesion (n=22)	McDonald 1977[w20]	Non-MS (80 months), MS (26 month), ranges NR	68	≥3 multifocal lesions	60, 71	2.0, 0.6
Filippini (1994)[w21], Italy 4	82 (65)	28 (14-51)	Optic lesion (n=21)	McDonald 1977[w20]	35 (NR)	34	≥1 MS-like abnormalities	96, 44	1.7, 0.1
							Paty	68, 69	2.1, 0.5
Sastre-Garriga (2004)[w22], Spain 1	153 (NR)	32 (14-50)	Optic lesion (n=56), spinal cord lesion (n=46), brainstem lesion (n=51);	Poser[w4]	34-40	26, 22, 35	Barkhof	63, 70	2.1, 0.5
			Optic lesion (n=56), spinal cord lesion (n=46)				Barkhof	52, 73	1.9, 0.7
			Brainstem lesion (n=51)				Barkhof	78, 61	1.9, 0.4
Sastre-Garriga (2003)[w23], Spain 1	51 (63)	29 (14-49)	Brainstem lesion (n=51)		37 (22-49)‡	35	≥1 lesions	94, 12	1.1, 0.6
							≥1 non-clinical lesions	94, 42	1.6, 0.2
							Fazekas	89, 48	1.7, 0.3
							Paty	89, 52	1.8, 0.3
Tintore (2001)[w24], Spain 1	112 (66)	28 (13-49)	Optic lesion (n=36), spinal cord lesion (n=41), brainstem lesion (n=23), other (n=12)		31 (12-60)	23	Fazekas	77, 51	1.6, 0.5
							Paty	77, 51	1.6, 0.5
Tintore (2003)[w25], Spain 1	86 (73)	30 (13-49)	Optic lesion (n=58), spinal cord lesion (n=39), brainstem lesion (n=34), other (n=8)		39 (12-77)	44	McDonald 2001 criteria at 12 months	74, 85	4.8, 0.3
Rio (1997)[w26], Spain 2	35 (80)	31 (17-47)	Optic lesion (n=35)	NR	29.4 (12-66)	20	≥1 T2 lesions	100, 67	2.7, 0.1
Soderstrom (1998)[w27], Sweden	147 (80)	34 (12-57)	Optic lesion (n=147)	Poser[w4]	25 (0-71)	41	≥1 lesions	85, 65	2.4, 0.2

BMJ: first published as 10.1136/bmj.38771.583796.7C on 24 March 2006. Downloaded from http://www.bmj.com/ on 29 March 2024 by guest. Protected by copyright.

Author (year), country	No of patients* (% female)	Mean (range) age (years)	Presenting symptoms	Criteria for MS	Mean follow-up (range) in months	% with MS	MRI criteria	Sensitivity, specificity (%)	Positive, negative likelihood ratios
Beer (1995)[w28], Switzerland	189 (57)	38 (16-67)	NR	Poser[w4]§, also incorporated paraclinical tests	No follow-up: diagnosis after testing	75	≥1 non-clinical lesions	84, 62	2.2, 0.3
							Fazekas	60, 87	4.4, 0.5
Beck (2003)[w29], USA 1	388 (77)	32 (18-46)	Optic lesion (n=388)	Poser[w4]	Estimates for MS at 10 years	37	≥1 non-clinical T2 lesions	68, 68	2.1, 0.5
							≥2 T2 lesions	51, 77	2.2, 0.6
							≥5 T2 lesions	32, 89	2.8, 0.8
							≥9 T2 lesions	12, 94	2.0, 0.9
Tumani (1998)[w30], USA 1	28 (NR)	32 (18-46)	Optic lesion (n=36)	Poser[w4]†	>48	61	≥2 lesions, at least one periventricular or ovoid	59, 73	2.0, 0.6
Jacobs (1997)[w31] USA 2	74 (69)	34 (12-61)	Optic lesion (n=74)	McDonald 1977[w20]	67 (4-228)	28	≥1 lesions	76, 51	1.5, 0.5
Mushlin (1993)[w32], USA and Canada	303 (73)	37 (14-75)	NR	Committee decision based on clinical and paraclinical data	7-8 (>6)	54	Paty	58, 91	6.1, 0.5
							Three lesions or two with one periventricular	75, 77	3.3, 0.3
							Multiple white matter and periventricular lesions	36, 99	20.1, 0.7

Several publications refer to same study, as indicated by numbers after country.
MRI=magnetic resonance imaging; NR=not reported.
*Total number who entered trial (includes withdrawals); this may differ from sum of number of patients with each of presenting symptoms as for most studies these were reported only for patients who completed the study.
†Patients with clinically probable MS were classified as having the disease.
‡Interquartile range.
§Patients with clinically probable or possible MS were classified as having the disease.

more non-clinical lesions. Figure 4 is a receiver operating characteristic plot for these criteria, with numbers showing the duration of follow-up in years. Evidence shows (P = 0.074 from hierarchical summary receiver operating characteristic analysis) that studies with longer follow-up produced higher estimated specificity and lower estimated sensitivity.

The longest average duration of follow-up was three years in studies that assessed the Barkhof, Fazekas, and McDonald 2001 criteria, and six years for studies that assessed the Paty criteria. It is therefore possible to draw conclusions regarding the ability of these criteria to predict the development of multiple sclerosis only over these relatively short periods. Figure 5 shows the receiver operating characteristic plots for these criteria. The study that developed the Barkhof criteria[22] showed higher estimated sensitivity and specificity than did the other studies of this criterion. The negative likelihood ratios for the Barkhof, Fazekas, and Paty criteria ranged from 0.2 to 0.5, suggesting that a negative result on magnetic resonance imaging on the basis of these criteria is of limited utility for ruling out the development of multiple sclerosis within three to six years. Positive likelihood ratios were <5: thus these criteria are also of limited utility in predicting the development of multiple sclerosis within three to six years. Positive likelihood ratios for the McDonald 2001 criteria ranged from 2.7 to 8.7, suggesting that they have more potential for predicting the development of multiple sclerosis within three years than any of the criteria based on magnetic resonance imaging alone.[23–26] Negative likelihood ratios were 0.1 in one study and 0.2 to 0.5 in three studies, suggesting that the McDonald 2001 criteria are of limited utility for ruling out the development of multiple sclerosis within three years.

Only two studies, one from the United States[2] and one from England,[3] followed patients for more than 10 years, long enough to be reasonably confident that almost all patients had been diagnosed as having multiple sclerosis who ever would be. Both studies fulfilled all but one QUADAS criterion (the availability of clinical information), and in the US study it was unclear whether review bias had been avoided (see bmj.com). The US study included 351 patients with optic neuritis; follow-up of more than 10 years was available for 302 (86%) of these. The study used survival analysis to estimate the cumulative proportions of patients diagnosed, with patients who did not receive a diagnosis of multiple sclerosis censored at the time of their last clinical follow-up. The English study included 135 patients with a range of presenting symptoms, of whom 71 (53%) were included in the final evaluation. Both studies evaluated thresholds based on the number of non-clinical T2 lesions present on magnetic resonance imaging of the brain.

Figure 6 shows the estimates of sensitivity and specificity, with confidence intervals, for each of the thresholds evaluated in these two studies. Sensitivity and specificity varied according to the number of lesions used to define a positive result on magnetic resonance imaging: sensitivity was higher with fewer lesions but specificity was lower. Estimates of specificity were similar for the two studies, but the English study tended to produce higher estimates of sensitivity. Comparison of areas under the curves suggested better accuracy in the English study than in the US study (P = 0.045). Estimates of the 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. Estimates of the negative likelihood ratio 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.

BMJ: first published as 10.1136/bmj.38771.583796.7C on 24 March 2006. Downloaded from http://www.bmj.com/ on 29 March 2024 by guest. Protected by copyright.

Table 2 Study details and results of case-control studies and studies of other designs

Author (year), country	Study design	No of patients* (% female)	Mean age (range) in years	No of patients with confirmed MS	No with other conditions	No of healthy volunteers	Criteria for MS	MRI criteria†	Sensitivity, specificity (%)	Positive, negative likelihood ratios
Offenbacher (1993)[w33], Austria 1	Consecutive scans done for multiple purposes (9% query MS)	1500 (52)	46 (12-93)	0	1251	115	Poser[w4]‡	Fazekas	81, 96	20, 0.2
								≥3 areas of high signal intensity	90, 71	3.3, 0.1
								Paty A	87, 74	3.3, 0.2
								Paty B	87, 92	10.2, 0.1
Fazekas (1988)[w34], Austria 2	Case-control	91 (NR)	46 (14-77)	50	0	49	NR†	Fazekas	88, 100	73.2, 0.1
								≥1 lesion >6 mm diameter	92, 95	15.3, 0.1
								≥1 infratentorial lesion	66, 98	18.4, 0.4
								≥1 periventricular	80, 98	22.2, 0.2
								≥3 lesions	96, 59	2.3, 0.1
								≥3 mm	98, 56	2.2, 0.1
								Three lesions >3 mm	96, 59	2.3, 0.1
Ravnborg (1992)[w35], Denmark 2	Cross-sectional study of patients with ≥1 neurological attacks	68 (56)	40 (18-63)	0	68	0	Poser[w4]	≥1 periventricular lesion	95, 65	2.7, 0.1
Van der Eerden (1990)[w36], Holland 2	Nested case-control	49 (50)	45 (21-77)	80	0	0	Schumacher[w2]	Presence of periventricular white matter lesions	93, 90	7.7, 0.1
Bot (2002)[w37], Holland 3	Case-control	91 (73)	43 (NR)	25	66	0	Poser[w4]	Barkhof	76, 92	9.1, 0.3
								Barkhof plus abnormal spinal cord	76, 100	101, 0.3
								Fazekas	92, 79	4.2, 0.1
								Fazekas plus abnormal spinal cord	84, 97	22.2, 0.2
								Paty	100, 50	2.0, 0.04
								Paty plus abnormal spinal MRI	92, 95	17.3, 0.1
								≥1 lesions	100, 36	1.5, 0.1
								Any abnormality in spinal cord	92, 94	13.5, 0.1
Rovaris (2002)[w38], Italy 5	Case-control	123 (67)	40 (NR)	64	59	0	Poser[w4]	Barkhof	80, 97	19, 0.2
								Barkhof or >1 spinal cord lesion	95, 100	114, 0.1
								≥1 lesions	100, 53	2.1, 0.01
Rovaris (2000)[w39], Italy 6	Case-control	54 (69)	40 (20-66)	10	44	0	NR	Any abnormality in spinal cord	90, 100	77.7, 0.1
Kuroda (1995)[w40], Japan 1	Case-control	72 (71)	44 (13-72)	36	36	0	Schumacher[w2]‡	≥1 lesions >6 mm diameter	56, 94	8.2, 0.5
								≥1 infratentorial lesion >3 mm diameter	39, 100	29, 0.6

BMJ: first published as 10.1136/bmj.38771.583796.7C on 24 March 2006. Downloaded from http://www.bmj.com/ on 29 March 2024 by guest. Protected by copyright.

Author (year), country	Study design	No of patients* (% female)	Mean age (range) in years	No of patients with confirmed MS	No with other conditions	No of healthy volunteers	Criteria for MS	MRI criteria†	Sensitivity, specificity (%)	Positive, negative likelihood ratios
								≥1 periventricular lesion >3 mm	72, 94	10.6, 0.3
								Three lesions >3 mm	86, 94	12.6, 0.2
Palmer (1999)[w41], USA 3	Case-control	50 (76)	40 (18-70)	6	25	0	NR	Subcallosal striations	100, 84	5.8, 0.05
Gean-Marton (1991)[w42], USA 4	Case-control	169 (46)	41 (22-87)	47	122	0	Poser[w4]	Focal abnormalities on callosal-septal interface	93, 98	34, 0.1
Yetkin (1991)[w43], USA 5	Case-control	260 (NR)	NR (31-83)	92	68	100	Schumacher[w2]	Abnormal MRI scan: equivocal findings classed as normal	76, 96	17.1, 0.3
								Abnormal MRI scan: equivocal findings classed as normal	80, 96	18.1, 0.2

MS=multiple sclerosis; MRI=magnetic resonance imaging; NR=not reported.
*Total number of patients who entered trial (includes withdrawals); in some cases this may differ from sum of number of patients with confirmed MS, other conditions, or healthy volunteers as for most studies these were only reported for patients who completed the study.
†Patients with clinically probable MS were classified as having the disease.
‡Patients with clinically probable or possible MS were classified as having the disease.

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 on the accuracy of magnetic resonance imaging for the diagnosis of multiple sclerosis included at least 10 years' follow-up. These suggested that the role of magnetic resonance imaging either in ruling in or ruling out a diagnosis of multiple sclerosis is limited. Studies that did not include an appropriate patient spectrum tended to overestimate both sensitivity and specificity. Studies that included shorter clinical follow-up tended to overestimate sensitivity and underestimate specificity. Specific criteria developed for the interpretation of magnetic resonance imaging scans as indicating multiple sclerosis, the Fazekas, Barkhof, and Paty criteria, have 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 McDonald 2001 were evaluated in studies with long term follow-up. It is therefore not possible to determine their accuracy for the diagnosis of multiple sclerosis.

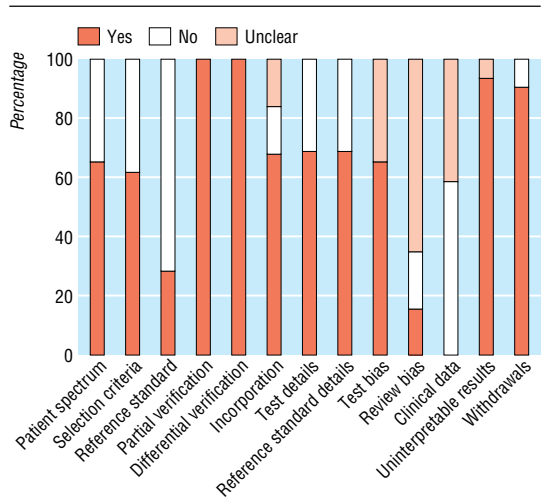


Fig 2 Results of quality assessment for appropriate patient spectrum studies

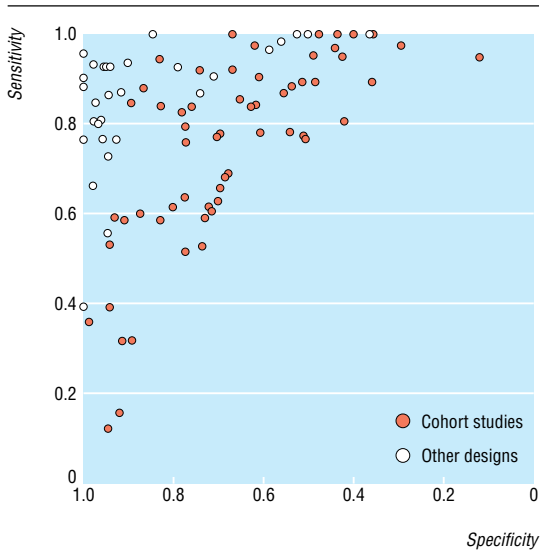


Fig 3 Receiver operating characteristic plots for cohort studies and for studies of other designs

BMJ: first published as 10.1136/bmj.38771.583796.7C on 24 March 2006. Downloaded from http://www.bmj.com/ on 29 March 2024 by guest. Protected by copyright.

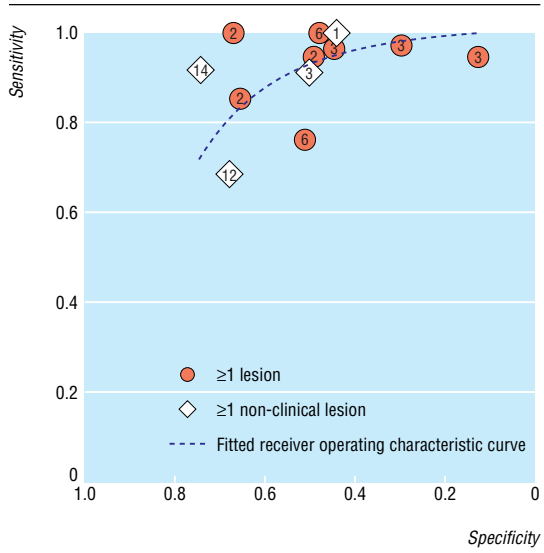


Fig 4 Receiver operating characteristic plots for studies included in hierarchical summary receiver operating characteristic analysis. Numbers are duration of follow-up in years

Strengths and weaknesses of the study

We carried out extensive literature searches, assessed study quality, and used recently developed statistical methods. 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 relatively short periods and so will have classified some patients as not having multiple sclerosis who had a second clinical attack after follow-up ended. Most studies included an inappropriate patient spectrum, which we found to be associated with considerably higher estimated diagnostic accuracy. Most of such studies used a case-control design—they selected people with clinically definite multiple sclerosis and a control group of people known not to have the disease, either healthy controls or patients with conditions that may present with similar symptoms to multiple sclerosis. That such studies tend to exaggerate the accuracy[27] of magnetic resonance imaging in the diagnosis of multiple sclerosis is to be expected; people with more advanced multiple sclerosis are more likely to have lesions on their magnetic resonance imaging scans than those presenting in the early stages of multiple sclerosis.

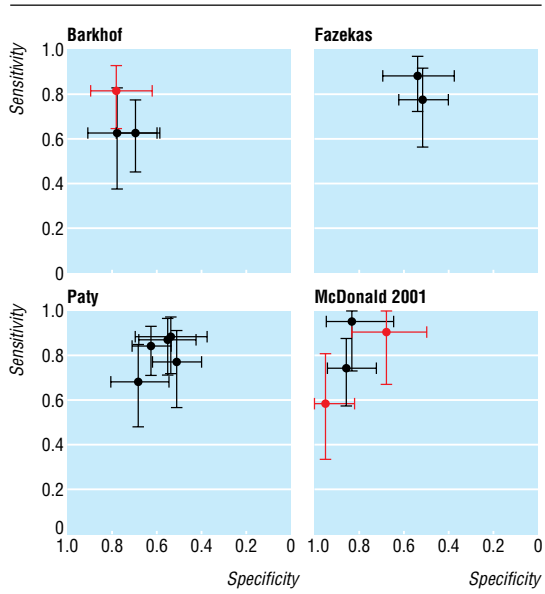


Fig 5 Receiver operating characteristic plots (95% confidence intervals) for Barkhof, Fazekas, Paty, and McDonald 2001 criteria. In Barkhof plot, red is study that proposed Barkhof criteria.[22] In McDonald 2001 plot, red indicates studies where McDonald 2001 criterion was applied after three months rather than 12 months

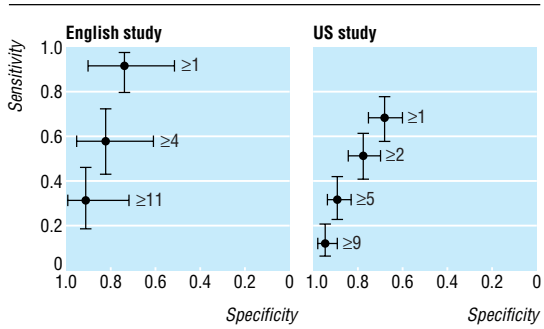


Fig 6 Sensitivity plotted against specificity (95% confidence intervals) for different thresholds (number of lesions shown next to plots) reported in English and US studies[2 3]

Strengths and weaknesses in relation to other studies

Although several reviews have assessed the accuracy of magnetic resonance imaging in the diagnosis of multiple sclerosis,[4 28 29] we are unaware of any systematic reviews. The McDonald 2001 criteria incorporate the Barkhof criteria to define a positive MRI scan.[4] The article reporting the McDonald 2001 criteria[4] refers to a small number of studies to justify its selection of the Barkhof criteria for this purpose. All these had methodological weaknesses: they either used a case-control design or had an average of less than three years clinical follow-up. This paper was published before the two long term cohort studies from England and the United States.[2 3]

A recently published detailed (but not systematic) review is the report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology.[28] This was limited to cohort studies and discussed in detail the problems associated with the lack of a true reference standard, an accurate method of determining whether or not a patient has multiple sclerosis that can be applied at the same time as the index test, for the diagnosis of the disease. It did not carry out any statistical synthesis and instead presents a narrative overview of the results of the English study and several other studies, also included in our review, which had relatively short clinical follow-up, and was published before the US study. It concluded, in contrast with our findings, that the presence of at least three lesions on a magnetic resonance imaging scan is a sensitive predictor of the development of multiple sclerosis in the next 7-10 years, and that normal results suggest that future development of multiple sclerosis is less likely. A more recent review article focused on the McDonald 2001 criteria but also draws on the results of the report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology.[29] It highlights the limitations of the evidence base for the McDonald 2001 criteria and draws on the results of the US[2] and English studies[3] to conclude, consistent with the results presented here, that presence of brain lesions does not guarantee development of multiple sclerosis over 10-14 years.

Unanswered questions and future research

The main clinical question is whether magnetic resonance imaging should be included in the work-up of patients with multiple

BMJ: first published as 10.1136/bmj.38771.583796.7C on 24 March 2006. Downloaded from http://www.bmj.com/ on 29 March 2024 by guest. Protected by copyright.

sclerosis. Several factors need to be considered, in particular the reasons why magnetic resonance imaging is ordered. This is not simply 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. If magnetic resonance imaging scans are ordered to inform the diagnosis of multiple sclerosis, and if the McDonald 2001 criteria that incorporate such imaging are to be used in practice, then further research, based on long term cohort studies, is required to evaluate these criteria. A limitation consequent on the need for long term clinical follow-up in studies that evaluate the accuracy of magnetic resonance imaging is that such studies inevitably use older technology. Studies with more advanced, and hence recent, technology inevitably had much shorter periods of follow-up. Differences in estimates of sensitivity and specificity according to magnetic resonance imaging technology were therefore confounded by differences in duration of follow-up.

The two studies that included follow-up of longer than 10 years produced differing results, with the US study reporting lower estimates of sensitivity than the English study for similar thresholds for magnetic resonance imaging. It is possible that these differences reflect the smaller sample size of the English study or that the large proportion of dropouts from this study biased results. An alternative explanation is that magnetic resonance imaging may be more accurate in patients presenting with brainstem or spinal cord symptoms than in patients with optic neuritis. Future studies should assess whether the accuracy of magnetic resonance imaging varies according to presenting symptoms.

Rather than the accuracy of magnetic resonance imaging alone in diagnosing multiple sclerosis, the issue of clinical relevance is, arguably, the added value of such imaging in diagnosing the disease compared with the patient's history and clinical examination alone.[30] None of the identified studies addressed 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, ranging from no lesions (in which case the probability of disease is low), to specific lesions (which may imply a greatly increased probability of disease), should be considered as an alternative to dichotomisation. This is probably consistent with how the results of magnetic resonance imaging are interpreted in practice.

Implications

In patients with clinically suspected multiple sclerosis, magnetic resonance imaging currently allows a diagnosis of the disease according to the McDonald 2001 criteria. Our results suggest that magnetic resonance imaging is a relatively poor test for both ruling in and ruling out multiple sclerosis. In clinical practice a false positive 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. Wrongly ruling out a diagnosis of multiple sclerosis after a first attack seems less dangerous: not all patients who experience a first attack will develop the disease and currently no treatment has been shown to delay conversion to clinically definite multiple sclerosis or impacts on long term disability. Neurologists should discuss with their patients the potential diagnosis, treatment, and ultimate effect of potential errors of false positive and false negative magnetic resonance imaging results. High quality clinical research based on improved magnetic resonance imaging techniques and measures in combination with a complete description of participants and long term clinical follow-up are needed for quantitative assessment of the clinical efficacy of magnetic resonance imaging in the diagnosis of multiple sclerosis. The disease remains a predominantly clinical diagnosis.

Contributors: PW, JACS, JJD, CM, and RH designed the study. PW carried out the literature searches. PW and CM screened the results of the searches for relevance, assessed inclusion, and carried out the data extraction and quality assessment. GF provided expert advice on magnetic resonance imaging and multiple sclerosis. RH, JJD, JACS, and PW developed the plan of analysis and RH carried out the analysis. PW, JACS, RH, JJD, and GF drafted the paper. All authors commented on drafts of the paper and approved the final manuscript. PW is guarantor for the paper.

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 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.
3 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.
4 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.
5 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.
6 Mushlin AI, Mooney C, Grow V, Phelps CE. The value of diagnostic information to patients with suspected multiple sclerosis. Rochester-Toronto MRI Study Group. *Arch Neurol* 1994;51:67-72.
7 Koopman W, Schweitzer A. The journey to multiple sclerosis: a qualitative study. *J Neurosci Nurs* 1999;31:17-26.
8 Heesen C, Kolbeck J, Gold SM, Schulz H, Schulz KH. Delivering the diagnosis of MS—results of a survey among patients and neurologists. *Acta Neurol Scand* 2005;107:363-8.
9 O'Connor P, Detsky AS, Tansey C, Kucharczyk W. Effect of diagnostic testing for multiple sclerosis on patient health perceptions. Rochester-Toronto MRI Study Group. *Arch Neurol* 1994;51:46-51.
10 O'Connor P, Canadian Multiple Sclerosis Working Group. Key issues in the diagnosis and treatment of multiple sclerosis. An overview. *Neurology* 2002;59:S1-33.
11 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.
12 Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ* 2004;329:168-9.
13 Higgins JP, Thompson SG. Controlling the risk of spurious findings from meta-regression. *Stat Med* 2004;23:1663-82.
14 Rutter CA, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001;20:2865-84.
15 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.
16 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.

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

BMJ: first published as 10.1136/bmj.38771.583796.7C on 24 March 2006. Downloaded from http://www.bmj.com/ on 29 March 2024 by guest. Protected by copyright.

17 McDonald WI, Halliday AM. Diagnosis and classification of multiple sclerosis. *Br Med Bull* 1977;33:4-9.
18 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.
19 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.
20 Sharief MK, Thompson EJ. The predictive value of intrathecal immunoglobulin synthesis and magnetic resonance imaging in acute isolated syndromes for subsequent development of multiple sclerosis. *Ann Neurol* 1991;29:147-51.
21 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.
22 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.
23 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.
24 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.
25 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.
26 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.
27 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.
28 Frohman EM, Goodin DS, Calabresi PA, Corboy JR, Coyle PK, Filippi M, et al. The utility of MRI in suspected MS: report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. *Neurology* 2003;61:602-11.
29 Miller DH, Filippi M, Fazekas F, Frederiksen JL, Matthews PM, Montalban X, et al. Role of magnetic resonance imaging within diagnostic criteria for multiple sclerosis. *Ann Neurol* 2004;56:273-8.
30 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

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: first published as 10.1136/bmj.38771.583796.7C on 24 March 2006. Downloaded from http://www.bmj.com/ on 29 March 2024 by guest. Protected by copyright.