

Review Article

Accuracy of diagnostic tests in multiple sclerosis – a systematic review

Schäffler N, Köpke S, Winkler L, Schippling S, Inglese M, Fischer K, Heesen C. Accuracy of diagnostic tests in multiple sclerosis – a systematic review.

Acta Neurol Scand: DOI: 10.1111/j.1600-0404.2010.01454.x.

© 2010 John Wiley & Sons A/S.

New diagnostic criteria for multiple sclerosis (MS) have been recently proposed and further updates are upcoming. This systematic literature review summarizes diagnostic studies in suspected MS to clarify the value of diagnostic tests. We included studies of at least 40 patients followed up for 2 years. All studies are limited by the fact that no gold standard to validate diagnostic tests is available. A second relapse is used as a surrogate in relapsing–remitting MS, but long follow-up of at least 5 years is necessary to detect all cases. Many studies showed selection bias, partly because of the vague definition of a clinically isolated syndrome. Based on these limitations, sensitivity of magnetic resonance imaging (MRI) criteria was between 35% and 100%, and specificity was between 36% and 92%. Cerebrospinal fluid (CSF) oligoclonal banding showed sensitivities between 69% and 91% with specificities between 59% and 94%. Combination studies of MRI and CSF indicate enhanced sensitivity (56–100%) and specificity (53–96%). Studies on evoked potentials did not justify conclusions about their value. A combination of simplified MRI criteria with CSF might be the best approach for an early MS diagnosis. However, the value of a very early diagnosis stays questionable as patients' benefit of new diagnostic criteria has never been addressed.

**N. Schäffler^{1*}, S. Köpke^{2*},
L. Winkler¹, S. Schippling¹,
M. Inglese³, K. Fischer¹,
C. Heesen¹**

¹Institute of Neuroimmunology and Clinical MS Research (INiMS), University Medical Center Hamburg-Eppendorf, Hamburg, Germany; ²Department of Health Sciences, University of Hamburg, Hamburg, Germany; ³Department of Radiology, New York University, New York, NY, USA

Key words: multiple sclerosis diagnosis; test accuracy; cerebrospinal fluid; evoked potentials; magnetic resonance imaging; systematic review

C. Heesen, Institute of Neuroimmunology and Clinical MS-Research, University Hospital Eppendorf, Martini-strasse 52, D-20246 Hamburg, Germany
Tel.: +49 40 7410 54076
Fax: +49 40 7410 56973
e-mail: heesen@uke.uni-hamburg.de
*Equally contributing authors.

Accepted for publication September 21, 2010

Introduction

In parallel to the development of the disease-modifying agents in multiple sclerosis (MS), new diagnostic criteria have been inaugurated in 2001 and revised in 2005, making a diagnosis possible after a first clinical episode of the disease the so-called clinically isolated syndrome (CIS) (1, 2). Meanwhile, all interferon- β preparations and glatirameracetate have shown that they can delay a second attack of the disease for almost a year in about 14 of 100 patients (3–5). This has emphasized the importance of an early diagnosis of MS for initiating treatment as early as possible (6, 7).

MS is still an ill defined entity with heterogeneous presentation especially in the early phase of the disease and might therefore be called a syndrome rather than a distinct disease, showing also great

variety of discussed disease mechanisms and patterns in neuropathological examination (8). The first diagnostic criteria (9) have implemented dissemination in time (DIT) and dissemination in space (DIS), i.e. the temporarily disseminated occurrence of symptoms and signs affecting at least two anatomically different locations in the brain. These criteria were modified by McAlpine in 1972 (10) and by McDonald & Halliday in 1977 (11) and, eventually, substituted by the Poser's criteria in 1983 (12). Aiming to better define the category of probable MS, Poser criteria included cerebrospinal fluid (CSF) findings and evoked potentials (EVOP) as paraclinical tests. This classification did not include diagnostic criteria for primary-progressive MS (PPMS). McDonald criteria (1) allowed to establish a MS diagnosis after a single clinical event by including magnetic resonance imaging (MRI)

findings based on studies by Barkhof et al. (13) and Tintoré et al. (14). Criteria for PPMS diagnosis were included. Nevertheless, a major limitation of McDonald criteria is the requirement of ‘no better explanation than MS’ in absence of guidelines for alternative diagnoses. Poser et al. (15) warned against a possible higher rate of false positive findings using McDonald criteria compared to Poser criteria from 1983. Charil et al. and Miller et al. (16, 17) have proposed a list of MRI ‘red flags’ which might be included in diagnostic criteria after more systematic evaluation.

In an earlier review, Whiting et al. (18) have reviewed the role of MRI as a diagnostic tool in MS, concluding that currently there is no sufficient data to rule in or rule out MS with MRI. A comprehensive review of all diagnostic studies clarifying as well the value of CSF and EVOP analysis in combination with MRI has never been carried out. EVOP have been used since the early 1970s (19), showing slowing of central conduction velocities. In CSF analysis, the detection of oligoclonal IgG-production has been considered a relatively specific finding since the 1960s (20, 21). We aimed at analyzing all diagnostic studies in MS to further clarify and compare the value of the major methods MRI, CSF analysis and EVOP.

Methods

Search strategy

We performed a systematic literature search in July 2010 via PubMed including

‘multiple sclerosis’ OR ‘transverse myelitis’ OR ‘optic neuritis’ OR ‘adem’ OR ‘neuromyelitis optica’
AND (specificity[Title/Abstract])

Also, the MeSH-Database was searched using the following search terms:

‘Multiple Sclerosis’ [Mesh] OR (‘Myelitis, Transverse’ [Mesh:noexp]) OR (‘Demyelinating Diseases’ [Mesh:noexp]) OR (‘Encephalomyelitis, Acute Disseminated’ [Mesh])
AND ‘Sensitivity and Specificity’ [Mesh]

Reference lists of retrieved articles were checked for further studies.

Eligibility criteria

We only included publications in English, French or German reporting on the diagnostic accuracy of

MRI, CSF and/or EVOP. Only studies with at least 2 years of follow-up and at least 40 patients were included referring to a systematic review on prognosis of MS (22). Study types eligible for inclusion were prospective and retrospective cohort studies or groups in randomized controlled treatment trials of patients with first symptoms suggestive of demyelination, which are called CIS in general or chosen typical CIS syndromes i.e. optic neuritis (ON), myelitis or brain stem syndrome. We excluded studies that solely analysed definitive MS cases or compared different laboratory techniques to analyse protein in the CSF.

Two independent reviewers (NS and CH) scanned titles and abstracts. In case of dissent, this was discussed and consent was achieved. Afterward, full-text articles were checked for eligibility by the same two authors using a self-developed assessment sheet. Again, if necessary, authors discussed studies in case of disagreement. In repeatedly published cohorts, we used the most recent update. NS extracted the data, which were checked again by CH and SK. CH and SK independently applied QUADAS criteria as proposed by Whiting et al. (23) in the context of MS. Agreement procedures were performed as outlined earlier.

Accuracy parameters

Accuracy of tests is determined by reporting sensitivity and specificity. To allow comparison of accuracy between studies, diagnostic odds ratios (DORs) and corresponding confidence intervals were calculated (24, 25). DOR describes the odds of positivity among diseased persons, divided by the odds of positivity among non-diseased (20) and therefore summarizes sensitivity and specificity in one measure. The higher the DOR, the higher the accuracy of the instrument (24). A DOR of 1 is similar to a prediction based on chance (26). There are no cut-off values for acceptable or sufficient DORs, although DORs below 5 certainly indicate low test accuracy (24).

Results

The literature search resulted in 2125 hits. Of these, 2022 were considered irrelevant after scanning abstracts. Of the remaining 103 papers, further 61 were excluded after reading full texts. Reasons for exclusion were insufficient follow-up (26), analysis of only definite MS cases (11), only comparison of laboratory methods (15), < 40 patients (5), other diagnostic tests studied (4). Finally, 42 studies were included, and accuracy parameters were extracted

or calculated. Most studies ($n = 39$) were cohort studies of patients with CIS in general or specific CIS subgroups as i.e. ON. Three studies reported on randomized controlled trials. Twenty-three MRI studies used a replacement approach comparing MRI criteria; in twelve studies, different MRI criteria were compared to the gold standard clinical definite multiple sclerosis (CDMS) defined by Poser criteria. All six CSF-focussed studies applied a replacement approach. Among twelve studies applying MRI and CSF analysis, eleven studied the add-on value to given MRI criteria. Four of five EVOP studies used a replacement approach, only one study assessed the additional value in comparison with MRI. All included studies compared diagnostic tests directly.

QUADAS ratings

As shown in Fig. 1, there are major quality deficits in MS diagnostic studies. Less than half of the studies clearly reported on consecutive patients with MS suspicious symptoms (patient spectrum). In only five studies, follow-up has been 10 year or more which seems necessary for not missing too many cases developing MS (reference standard). Only three studies report blinding of clinicians for test results when making a diagnosis using the reference standard (review bias). Finally, only two studies report handling of uninterpretable test results for e.g. low number of oligoclonal bands (OCB) or questionable lesions on MRI (Table S1).

MRI studies

Most MRI studies were performed in London, Barcelona and Amsterdam. Different MRI criteria

have been proposed since 1988 and are referred to in Table 1. Fazekas et al. (27) created criteria to differentiate established MS from other diseases based on a retrospective analysis of definitive MS patients. He demanded at least three lesions and the fulfilment of at least two of the three following criteria for a positive diagnosis of MS: size greater than or equal to 6 mm, periventricular or infratentorial location. Paty criteria demand at least four lesions, or three lesions, of which one is periventricular (28), Paty criteria, as well as all following criteria developments, are based on data of patients presenting with CIS suggestive of MS. The modified Barkhof criteria (BC), also called the Barkhof/Tintoré criteria, have been incorporated into the McDonald criteria for MS in 2001 for demonstrating DIS and require three of the following four criteria: one Gd-enhancing or ≥ 9 T2-hyperintense lesions, one infratentorial lesion, one juxtacortical lesion, three periventricular lesions. For demonstrating DIT, McDonald criteria 2001 demand a new gadolinium (Gd)-enhancing lesion on a scan performed three or more months after the initial clinical episode (1, 13, 14). Revision of these criteria in 2005 recommended that a spinal cord lesion is equivalent to a brain infratentorial lesion. Also, spinal cord lesions were considered equivalent to reach the required number of T2 lesions for DIS. Spinal Gd-enhancement was set equivalent to any enhancing cranial lesion (2). For DIT, the revised criteria propose the appearance of a new T2 lesion compared with a reference scan, carried out at least 30 days after the onset of the initial clinical event. Newer developments proposed by Swanton et al. (29) aimed at simplifying BC by demonstrating that DIS requires at least one lesion in at least two characteristic

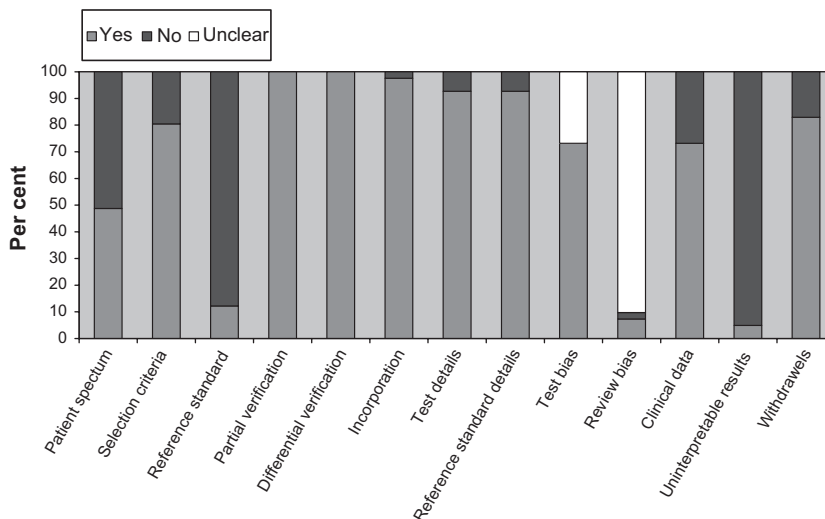


Figure 1. Results of quality assessment for appropriate patient spectrum studies (QUADAS).

Table 1 Diagnostic MRI studies in MS

Study	Centre	Patients	Follow-up (month)	MS criteria	MRI criteria	Sensitivity/specificity (%)	DOR (95% CI)	Consecutive	Prospective
Swanton (36)	London (UK)	142 ON	62	P	≥1 T2 lesions	86/24	8.8 (2.5–30.8)	+	+
Diaz-Sanchez et al. (53)	Madrid (Spain)	79 CIS	79	P	BC/Tintoré Swanton	72/77 91/68	8.7 (2.8–27.6) 22.3 (6.2–80.4)	+	+
Moraal et al. (50)	BENEFIT	468 CIS	36	P	BC/Tintoré	64/36	1.0 (0.7–1.5)	–	–
Nielsen et al. (49)	BENEFIT	176 CIS (PG)	24	McDo	≥9 T2 lesions Gd+	*	*	–	+
Chausson et al. (52)	Martinique (France)	66 CIS (Afro-caribbean)	34	P	McDo 2001 McDo 2005	35/68 38/63	1.2 (0.3–4.1) 1.1 (0.3–3.6)	–	–
Jafari et al. (48)	Rotterdam (NL)	158 CIS	39	P	Swanton BC	38/68 51/73	1.4 (0.4–4.7) 2.8 (1.4–5.4)	–	?
Korteweg et al. (31)	MAGNIMS	119 CIS	59	P	BC + CC lesion CC lesion	35/80 55/67	2.2 (1.1–4.6) 2.4 (1.2–4.8)	–	?
Rovira et al. (30)	MAGNIMS	250 CIS	24	P	BC/Tintoré ≥1 T2 lesions & 1 PV lesion	49/79 64/70	3.6 [†] 4.1 [†]	–	–
Lo et al. (51)	Taiwan	64 CIS	48	P	Swanton BC	77/60 48/49	3.7 (2.0–6.6) 4.9 (2.9–9.5)	?	–
Fisniku et al. (35)	London (UK)	107 CIS	242	P	Swanton ≥3 T2 lesions	100/53 100/60	5.3 (2.5–11.0) ‡	?	+
Rocca et al. (47)	Milan (Italy)	208 CIS	37	P	Swanton ≥1 T2 lesion	90/68	17.8 (6.4–49.6)	–	+
Swanton et al. (41)	MAGNIMS	208 CIS	40	P	≥4 T2 lesions ≥10 T2 lesions	63/78 37/85	5.8 (2.4–14.1) 3.4 (1.2–9.2)	–	+
Masjuan et al. (46)	Madrid (Spain)	52 CIS	72	P	Bark./Gd+	*	*	–	+
Korteweg et al. (40)	MAGNIMS	349 CIS	43	P	McDo 2001 McDo 2005	47/91 60/88	9.1 (4.3–17.2) 10.8 (5.4–21.6)	–	–
Tintoré et al. (39)	Barcelona (Spain)	157 CIS	84	P	Swanton BC	72/87 74/88	17.0 (8.4–34.4) 20.9 [†]	–	+
Minneboo et al. (32)	Amsterdam (NL)	42 CIS	104	P	BC/Tintoré BC (3 of 4)	49/79	3.7 (2.3–6.0)	–	+
Sastre-Garriga et al. (37)	Barcelona (Spain)	153 CIS	34	P	IT lesion	68/82	9.6 (4.6–20.4)	–	+
Beck et al. (45)	ONSG	388 ON	120 (at least)	P	Bark. ≥1 lesion	63/70 68/68	4.0 (1.9–8.4) 4.6 (2.9–7.3)	?	+
Sastre-Garriga et al. (38)	Barcelona (Spain)	51 CISB	37	P	BC	78/61	5.4 (1.4–20.0)	?	+
Dalton et al. (34)	London (UK)	41 ON	36	McDo	≥1 lesion & new spinal lesion	77/96	90 (8.4–969.2)	?	+
Dalton et al. (33)	London (UK)	56 CIS	36	McDo	New T2 lesion New Gd+ lesion	74/84 58/92	14.5 (3.8–55.5) 15.6 (3.5–69.2)	?	+
Barkhof et al. (13)	Amsterdam (NL)	74 CIS	39	P	BC	82/78	16.2 [†]	?	–
Jacobs et al. (44)	Buffalo (USA)	74 ON	67	McDo 77	BC (3 of 4)	61/88	11.1 (3.4–35.6)	?	–
Paolino et al. (42)	Ferrara (Italy)	44 CIS	84	McA	≥1 lesion ≥3 lesions	76/51 60/71	3.3 (1.1–10.4) 3.8 (1.0–14.8)	?	?

+ = Yes, – = no, ? = unclear.

BC, Barkhof criteria; CC, corpus callosum; CI, confidence interval; CIS, clinically isolated syndrome; DOR, diagnostic odds ratio; Gd+, gadolinium-enhancement; IF, infratentorial; McA, McAlpine criteria; McDo, McDonald criteria; McDo 77, McDonald and Halliday 1977 criteria; OCB, oligoclonal bands; ONSG, optic neuritis study group; ON, optic neuritis; P, Poser criteria; PG, placebo group; PV, periventricular.

*No data available.

[†]Confidence interval not computable because of lack of raw data.[‡]Not computable because of sensitivity of 100%.

locations: periventricular, juxtacortical, posterior fossa or spinal cord. To also simplify criteria for DIT, Swanton et al. (25) propose for DIT any new T2-lesion on a follow-up MRI scan, irrespective of the timing distance to the baseline scan. Rovira et al. (30) have recently proposed that a single brain MRI study demonstrating DIS and showing both gadolinium-enhancing and non-enhancing lesions suggesting DIT is highly specific for predicting development of CDMS, even when the MRI is performed within the first 3 months after the onset of a CIS. Korteweg et al. (31) investigated a model of simplified criteria based on the presence of ≥ 1 deep white matter lesion and ≥ 1 periventricular lesion.

Amsterdam cohort data – In 1997, Barkhof et al. published a cohort of 74 patients with CIS, whose MRI scans were retrospectively analysed to find the best possible predictive value (13). Thirty-three patients (41%) developed CDMS during a mean follow-up of 39 months (range 23–96). The study comprised three cohorts, investigated with different scan protocols (42 CIS from Amsterdam, 19 patients with spinal symptoms from Milan and 13 patients with ON from London). The criteria developed from the study showed the highest sensitivity (82%) and specificity (78%) reported until now. Major limitations of the study were the retrospective analysis, unclear patient selection, small patient number and short follow-up time.

A subgroup of 42 patients of the initial cohort (13), in which no CDMS had been diagnosed in 1997, was further followed up for a mean period of 8.7 years (32). Twenty-six patients developed a second relapse during follow-up and were thus diagnosed with CDMS. Adding this follow-up time, 59 (79%) of the initially 74 patients could be diagnosed with CDMS. As the study did not include details on MRI data, it is not possible to derive information about the sensitivity or specificity of BC over the longer follow-up.

London cohort data – Dalton et al. (2003) studied 56 patients with CIS for 3 years to compare the diagnostic criterion ‘occurrence of a new Gd+ lesion’ (1) to the criterion of a ‘new T2 lesion’ (33). Twenty of 50 patients showed new T2 lesions after 3 months while only 14 of them showed Gd-enhancing lesions. Eight of the patients not fulfilling McDonald criteria because of the lack of at least one Gd-enhancing lesion developed CDMS after 3 years. By allowing a new T2-lesions as an alternative for DIT, only five patients who developed CDMS did not fulfil these modified criteria. The inclusion of the new T2-lesion criterion did

increase MRI sensitivity 3 months after symptom onset from 58% to 74% for the development of CDMS after 3 years, specificity decreased from 95% to 84%.

Also in 2003, Dalton et al. (34) studied the diagnostic value of an additional spinal MRI in 115 patients with ON. MRI cord imaging allowed a diagnosis of MS in only one additional asymptomatic patient out of 64 at 1-year follow-up and in two additional asymptomatic patients out of 44 after 3 years. By including new spinal lesions 1 year after the first MRI as indicator of DIT, specificity decreased from 79% to 75%, sensitivity remained unchanged at 85%. Therefore, additional spinal MRI is only of limited value in diagnosing definite MS in patients with clinically isolated ON.

Fisniku et al. (2008) were able to follow 109 patients of a cohort of 144 recruited patients with CIS for 20 years. MRI was obtained at baseline, at 5, 10 and 14 years (35). Sixty of 73 (82%) patients showing at least one asymptomatic lesion on baseline MRI developed CDMS during follow-up compared to only 7 of 34 (21%) without lesions. Sixty-seven patients of 107 (63%) developed CDMS. Increasing required lesion numbers from 1 to 10 led to decreased sensitivity but increased specificity.

Swanton et al. (2010) studied a cohort of 142 patients with ON to investigate the predictive value of different MRI parameters at baseline and 3-month follow-up scan (36). One hundred and fourteen of 142 patients showed at least one lesion on baseline scan, 57 (40%) of all patients developed CDMS during follow-up (median 62 months for non-converters). At least one lesion on a baseline scan showed a high sensitivity but low specificity for the development of CDMS. Strongest predictors for conversion to CDMS were new T2-lesions on a follow-up scan or at least three periventricular lesions on the baseline scan. All other MRI parameters did not show any significantly higher risk of conversion.

Barcelona cohort data – Sastre-Garriga et al. (37) compared different MRI criteria in 51 patients with clinically isolated brain stem syndromes. MRI at symptom onset was evaluated based on the development of CDMS after 37 months of follow-up. Sensitivity/specificity of BC was 78/61% (Fazekas criteria 89/48%, Paty criteria 89/52%) while 18 (35%) of the patients developed CDMS during follow-up.

In 2004, the same authors compared the validity of BC in CIS patients (38). In myelitis and ON-presentations ($n = 102$), BC had a sensitivity of 52% and a specificity of 73% for the development

of CDMS within 34 months. In patients with brainstem syndromes ($n = 51$), sensitivity decreased to 63% and specificity to 70%. Therefore, BC seem less specific for patients with brainstem syndromes than for other CIS patients.

Tintoré et al. (39) studied 157 patients with CIS followed up for a median of 7 years. Forty-two per cent of the patients developed CDMS, while 57% fulfilled McDonald criteria. Of the 102 patients with baseline MRI fulfilling BC, 66% could be diagnosed according to Poser criteria, and 72% according to McDonald criteria. BC showed a high sensitivity (94%) but a low specificity (54%) supporting only a questionable diagnostic value of BC.

MAGNIMS cohort data – Korteweg et al. (2006) validated BC on 532 CIS patients from seven MS centres. One hundred and seventy-three (33%) of the patients were diagnosed with CDMS after a mean follow-up of 3.6 years (40). MRI sensitivity (49%) and specificity (79%) for three or four BC were much lower than in the study conducted by Barkhof (13). Nevertheless, the higher the number of fulfilled criteria in the initial MRI, the higher was the likelihood of developing CDMS during follow-up.

Swanton et al. (41) compared the revised McDonald 2001 criteria with their new proposed criteria, which required one or more lesions in each of at least two characteristic locations: periventricular, juxtacortical, fossa posterior, spinal. For DIT, one new T2 lesion in any MRI following the initial clinical episode was required. Two hundred and eight CIS patients from four different clinical centres were selected based on the availability of a baseline and a 1-year follow-up MRI scan. During a clinical follow-up of 3 years, 41% of the patients developed CDMS. Specificities for both McDonald and Swanton criteria were between 87% and 91%. Sensitivity of the Swanton criteria was higher than of McDonald criteria. This study with the largest cohorts of patients of all studies performed until now showed that the new criteria are simpler than the McDonald criteria without substantially compromising accuracy.

Rovira et al. (30) conducted a study to investigate the diagnostic value of a single early MRI for the diagnosis of MS. Data of 250 CIS patients from the MAGNIMS multicenter database were analysed with applying BC and Swanton criteria. To show DIT in a single scan, at least one asymptomatic gadolinium-enhancing lesion in conjunction with one or more non-enhancing lesions was postulated as a new criterion. This additional criterion for DIT decreased sensitivity (Barkhof

47%, Swanton 53%) and increased specificity (Barkhof 88%, Swanton 87%).

Korteweg et al. (31) analysed data of 349 CIS patients. During a mean follow-up time of 4.9 years, 132 (37.8%) converted to CDMS. Patients were divided into a training set ($n = 230$) and a test set ($n = 119$). Data analysis led to simplified criteria requiring ≥ 1 deep white matter lesion and ≥ 1 periventricular lesions on baseline MRI scan to increase the accuracy and sensitivity of the BC without losing specificity and to enhance their applicability in clinical practice. These simplified criteria applied to the test set showed increased sensitivity but reduced specificity compared to BC.

Other studies – Paolino et al. (42) studied 44 CIS patients over 7 years. After a mean follow-up of 26 months, 30 (68%) developed CDMS according to the McAlpine criteria (10). The appearance of at least one lesion in MRI showed a lower sensitivity (60%) and higher specificity (71%) than the comparable study of Filippini et al. (43). The study was highly selective as only patients with spinal cord or brain stem symptoms were included.

Jacobs et al. (44) followed a cohort of 74 patients with ON for a mean of 5.6 years. During follow-up, 21 patients (28%) developed CDMS, of those 16 (76%) showed at least one asymptomatic lesion on MRI. Five CDMS patients showed normal MRI at baseline, which stayed normal in four of these five patients during follow-up. Twenty-six of the 53 patients, who did not develop CDMS also showed abnormal MRI, leading to low specificity and moderate sensitivity. Therefore, accuracy for patients with ON seemed rather low.

In 2003, data of the ON treatment trial were published by Beck et al. (45). They studied 388 patients with ON followed up for at least 10 years, 145 (38%) patients developed CDMS during follow-up. Of the 160 patients showing one or more brain lesions on initial MRI, 56% developed CDMS while only 22% of the 191 patients with no lesions. Sensitivity and specificity of at least one lesion were moderate.

Masjuan et al. (46) studied 52 CIS patients for 6 years in Madrid. Initial MRI fulfilled at least three BC in 28 patients (54%). Twenty-six of these patients converted to CDMS during follow-up. By contrast, CDMS could be diagnosed only in 9 of 24 patients with initial negative MRI. Sensitivity and specificity of BC were about 80%.

Rocca et al. (47) conducted a study involving 208 patients with CIS suggestive of MS to define the overall brain damage (lesion volumes, normalized brain volume and magnetization transfer ratio

of normal-appearing white and grey matter) and its predictive value for the development of CDMS. During the mean follow-up period of 3.1 years, 43% of the patients converted to CDMS. The only significant independent predictors of evolution of definite MS were the positivity of International Panel criteria for DIS and occurrence of Gd-enhancing lesions. Data on sensitivity or specificity were not reported in the paper.

Jafari et al. (48) assessed the additional impact of callosal lesions to BC on prediction of CDMS. A cohort of 158 patients with CIS was followed up for a mean of 39 months. During follow-up, 64 (41%) patients developed CDMS. For prediction of CDMS BC vs. only one callosal lesion showed similar sensitivities with 51% vs. 55%, as well as similar specificities of 73% vs. 67%. Combining the two criteria, sensitivity dropped to 35% but specificity improved to 80%. Callosal lesions seem to be a risk factor in developing CDMS independent of BC, combination of the two might help to avoid false positive diagnosis.

Nielsen et al. (49) studied the placebo group of the BENEFIT multicenter interferon-beta-1b treatment study in CIS. In this highly selected study cohort (at least two asymptomatic lesions on MRI) followed up for 2 years, clinically defined DIS (monofocal vs. multifocal onset) was compared to MRI criteria for the prediction of CDMS. Of 176 patients, 45% developed CDMS during follow-up, with similar conversion rates of 47% for patients with monofocal onset (*n* = 93) and 44% for patients with multifocal onset (*n* = 83). Interestingly, the risk of CDMS was significantly higher in patients with monofocal presentation, who showed more than nine T2-lesions at baseline or Gd-enhancement (at baseline or at moth 3 or 6). In contrast, for patients with multifocal onset lesion, number was not predictive,

Moraal et al. (50) studied the 3-year follow-up data of 468 CIS patients of the BENEFIT-cohort to investigate, whether the prognostic value of

MRI parameters was altered by timing of treatment interventions. After 3 years, conversion rate to CDMS was 42% (37% for patients with early treatment and 51% for patients receiving delayed treatment either after 2-year placebo group or after conversion to CDMS). Altogether, conversion to CDMS for patients with three or more BC could be predicted with a sensitivity of 64% (early treatment 67% vs. delayed treatment 61%) and a specificity of only 36% (early treatment 36% vs. delayed treatment 34%).

In Taiwan, Lo et al. (51) conducted a study to determine which of two sets of MRI criteria for DIS (revised McDonald versus Swanton criteria) better predicted the conversion from CIS to CDMS. Data of 64 Asian patients with CIS were analysed. During a mean follow-up time of 48.3 month, 30 (46.9%) patients converted to CDMS. McDonald and Swanton criteria showed a high specificity and a positive predictive value of 100% for developing CDMS. However, sensitivity was low with 53% for McDonald and 60% for Swanton criteria.

In another non-Western setting, the Caribbean island Martinique, Chausson et al. (52) retrospectively applied MRI diagnostic criteria for DIS to 45 patients with CIS with a follow-up time of at least 2 years (median 55 months). Sensitivity was low for each criterion. These low accuracy estimates may indicate that MS MRI appearance might be different in Afro-Caribbean patients.

In Spain, Diaz-Sanchez et al. (53) compared the accuracy of DIS criteria of BC and Swanton criteria in predicting a conversion to CDMS in a cohort of 79 CIS patients followed up for a mean of 57 month. They found a lower specificity of Swanton's criteria but a considerably higher sensitivity compared to BC criteria.

CSF studies

In London, Moulin et al. (54) conducted a study with 183 patients with monosymptomatic demye-

Table 2 Diagnostic CSF studies in MS

Study	Centre	Patients	Follow-up (month)	MS criteria	CSF markers	Sensitivity/specificity (%)	DOR (95% CI)	Consecutive	Prospective
Masjuan et al. (46)	Madrid (Spain)	52 CIS	72	P	OCB	91/94	170.7 (16.4–1774.2)	+	+
Nilsson et al. (57)	Lund (Sweden)	86 ON	312	P	OCB pleocytosis	*	*	–	+
Cole et al. (56)	ONSG (USA)	76 ON	60	P	OCB	73/60	3.9 (1.3–11.5)	–	–
Paolino et al. (42)	Ferrara (Italy)	44 CIS	84	McA	OCB	80/86	24.6 (4.2–137.3)	+	+
Corridori et al. (55)	Milan (Italy)	44 ON & suspected MS	44	McDo 77	OCB	72/79	9.6 (2.4–39.4)	–	+
Moulin et al. (54)	London (UK)	183 possible MS	34	P	OCB	69/59	3.2 (1.4–7.5)	+	+

+ = yes; – = no.

CI, confidence interval; CIS, clinically isolated syndrome; DOR, diagnostic odds ratio; McA, McAlpine criteria; McDo 77, McDonald and Halliday 1977 criteria; OCB, oligoclonal bands; ONSG, optic neuritis study group; ON, optic neuritis; P, Poser criteria.

*No data available.

lination suggestive of MS. Of the 83 patients with positive OCB, 20 (24%) developed CDMS, whereas out of the 100 OCB negative patients only nine (9%) developed CDMS, leading to a cumulative sensitivity of 69% and specificity of 59% (Table 2).

Corridori et al. (55) reported the clinical evolution of 10 ON patients and 34 patients with suspected (19 patients), possible or probable MS (15 patients) according to the criteria of McDonald and Halliday (1977) for 2–10 years. In the ON cohort, three of four patients with positive OCB developed MS, but also three of six OCB negative patients. In the whole cohort, 25 (57%) of 44 patients developed CDMS during follow-up, of those 18 (72%) showed positive OCB.

Paolino et al. (42) investigated the CSF of 44 CIS patients during a follow-up of 7 years. Thirty patients (68%) developed CDMS. As only in-patients under 50 years with spinal or brain stem symptoms were included these results are selective.

An US-study by Cole et al. (56) analysed the usefulness of CSF diagnostic tests in 76 patients with ON, who were followed up for 5 years. Twenty-two (29%) developed CDMS. Sixteen (73%) of those had positive OCB. Forty-two per cent of 38 positively tested developed CDMS. Additional diagnostic value of CSF was only detected in patients with negative MRI. Normal MRI and normal CSF had a negative predictive value of 96%.

In the study by Nilsson et al. (57), 34 (40%) of 86 patients with ON developed CDMS during a follow-up of 30 years. Twenty of these patients were diagnosed in the first 3 years, all cases were diagnosed within 14 years after ON CSF findings from 50 patients at first presentation could be analysed. The risk of CDMS was 49% in patients with OCB, without OCB risk was 23%. Sensitivity and specificity cannot be obtained from the published data.

Masjuan et al. (46) studied 52 patients with CIS for 6 years. At study onset, 33 patients (63%) showed OCB and 32 of them developed CDMS. Only 3 of 19 patients with negative OCB developed CDMS. Sensitivity of OCB was 91% and specificity 94%. As the method used for OCB analysis was new, limited conclusions are possible without further studies with the same method.

Combination studies of MRI and CSF

Lee et al. (58) applied MRI and CSF analysis in a cohort of 184 patients after a first episode suggestive of MS, followed up for 2 years. Fifty-five patients (30%) developed CDMS during follow-

up. Paty-MRI criteria showed better test accuracy than OCB. Data for combination of both methods could not be obtained from the publication (Table 3).

Jin et al. (59) observed 115 patients with ON over 3 years in Sweden. After 3 years, 37 were diagnosed CDMS. CSF was obtained in only 82 patients, 95% of these showed positive OCB, in 76% IgG-index was elevated, while 58% of 45 patients without a definite MS diagnosis during the follow-up had positive findings as well (in 44% also IgG-index elevated). Sensitivity of OCB was high with 95% but specificity was only 42%. Combining results for OCB and MRI, sensitivity decreased while specificity was unaffected.

Filippini et al. (43) analysed MRI and CSF of 82 patients with suspected MS consecutively admitted to four clinical centres. As the appearance of one or more lesions on MRI was considered as positive MRI, they reached a high sensitivity (96%) with low specificity (44%). Nineteen of 43 OCB positive patients developed CDMS during a mean follow-up of 2.9 years. OCB showed only low accuracy for evolution to CDMS. Combining positive OCB and positive Paty criteria did not lead to better predictive values.

Rolak et al. (60) analysed data of 83 patients with ON included in an American multicenter study for the treatment of ON. Thirteen (16%) developed CDMS during a 2-year follow-up, of which 11 showed at least two T2 lesions on MRI. Nine of 11 patients with CDSM and positive MRI also showed OCB. Specificities of MRI and CSF were low.

In Sweden, Söderström et al. (61) followed 147 patients with ON for a mean of 25 months. Fifty-three patients (36%) developed CDMS during follow-up. Of the 103 patients showing OCB in CSF, 50 (49%) developed CDMS. Of the 64 patients with positive MRI, it was 40 (63%). OCB showed a higher sensitivity than the occurrence of three or more lesions on MRI while specificity was higher on MRI. Combining both, progression to CDMS could be predicted with an increased sensitivity of 100% and rather low specificity of 53%.

Ghezzi et al. (62) reported data of 102 patients with ON, of whom 37 developed CDMS during a mean follow-up of 2.3 years. All patients developing CDMS showed at least one lesion on MRI. With respect to a positive IgG-index in CSF, both sensitivity and specificity were low at 65% and 52%, respectively. Patients who had both a positive MRI and a positive IgG-index had a insignificant higher risk of developing CDMS than the patients with normal CSF.

Table 3 Diagnostic combination studies of MRI and CSF in MS

Study	Centre	Patients	Follow-up (month)	MS criteria	MRI criteria	CSF markers	Sensitivity/specificity (%)	DOR (95% CI)	Consecutive	Prospective
Rojas et al. (68)	Buenos Aires (Argentina)	40 CIS	60	P	BC (2 of 4)		93/44	11.0 (1.2–97.0)	?	+
						OCB	93/64	24.9 (2.8–221.7)	?	+
Tintoré et al. (65)	Barcelona (Spain)	392 CIS	50	P	BC (2 of 4)	OCB	87/80	26.0 (0.4–154.5)	?	+
						OCB	83/52	5.2 (3.1–8.5)	?	+
Zipoli (67)	Florence (Italy)	118 suspected DE	46	McDo	BC		63/76	5.3 (3.4–8.3)	?	+
						OCB	56/82	5.8 (3.7–9.2)	?	+
						OCB	70/49	2.2 (1.0–4.7)	+	+
						OCB	78/65	6.5 (2.8–15.0)	+	+
Permural (66)	Detroit (USA)	58 ATM	62	P	2 T2 lesions BC	OCB, IgG-index	80/57	5.4 (2.3–12.5)	+	+
							60/76	4.9 (2.2–10.8)	+	+
Villar et al. (64)	Madrid (Spain)	58 CIS	72	P	Normal MRI		100/63	‡	–	+
Jin et al. (59)	Huddinge (Sweden)	82 ON	46	P	2 T2 lesions BC		94/96	363.0 (31.0–4250.2)	–	+
						OCB	74/88	19.4 (4.7–81.2)	–	+
Tintoré et al. (63)	Barcelona (Spain)	139 CIS	39	P	≥3 lesions BC		95/42	12.8 (2.7–59.8)	–	–
						OCB	81/67	8.6 (3.1–24.0)	–	–
						OCB	74/68	6.0 (2.3–15.7)	–	–
Ghezzi et al. (62)	Milan (Italy)	112 ON	6.3	P	≥1 lesion		71/69	5.4 [†]	?	+
						OCB	78/63	6 [†]	?	+
Söderström et al. (61)	Stockholm (Sweden)	147 ON	25	P	≥3 lesions		100/48	‡	–	+
						IgG-synthesis	65/52	2.0 (0.8–4.7)	–	+
						IgG-synthesis	*	*	–	+
Rolak et al. (60)	USA	83 ON	24	P	≥3 lesions 2 T2 lesions		85/65	10.7 (4.2–27.5)	?	+
						OCB	69/42	17.9 (4.1–78.2)	?	+
Filippini et al. (43)	Milan (Italy)	82 suspected MS	35	McD77	P		100/53	‡	?	+
						OCB	85/57	7.3 (1.5–35.6)	–	+
Lee et al. (58)	Vancouver (Canada)	184 suspected MS	25	P	P		85/57	7.33 (1.5–35.9)	–	+
						OCB	68/69	4.7 (1.7–12.2)	+	+
						OCB	68/56	2.7 (1.0–6.9)	+	+
						OCB	100/62	‡	+	+
						OCB	84/63	8.6 (3.9–19.2)	?	+
						OCB	69/59	3.2 (1.6–6.3)	?	+

+ = yes, – = no, ? = unclear.

BC, Barkhof criteria; CI, confidence interval; CIS, clinically isolated syndrome; DE, demyelinating event; DOR, diagnostic odds ratio; McDo 77, McDonald and Halliday 1977 criteria; OCB, oligoclonal bands; ON, optic neuritis; P, Poser criteria.

*No data available.

[†]Confidence interval not computable because of lack of raw data.

[‡]Not computable because of sensitivity of 100%.

Tintoré et al. (63) studied in Barcelona the usefulness of McDonald criteria in 139 CIS patients followed for 3 years. Eighty per cent of the patients fulfilling McDonald criteria after 1 year developed a second relapse within 3 years of follow-up. Sensitivity of McDonald criteria was therefore 74% and specificity 86% to predict CDMS in 3 years. BC showed a sensitivity of 71% and specificity of 69%. Allowing both criteria for DIS (positive BC or two lesions plus OCB) increased sensitivity to 87% but specificity decreased to 56%.

Villar et al. (64) compared the accuracy of BC with the combination of at least two T2-lesions on MRI plus the presence of OCB. Fifty-eight consecutive CIS patients were followed up for 6 years. Twenty-eight patients fulfilled BC at disease onset. Twenty-five of these 28 patients developed CDMS. At least two lesions on MRI plus OCB were shown

by 34 patients, 33 of whom converted to CDMS. Combining MRI and CSF results lead to a sensitivity of 94% and a specificity of 96%.

Tintoré et al. (65) conducted in 2008 a large prospective study of 572 patients with CIS, to determine whether the presence of OCB adds information to a MRI scan in predicting the development of CDMS. After a mean follow-up of 50 month, 16% of OCB negative patients and 50% of OCB positive patients developed CDMS. Combining OCB and positive BC, sensitivity remained at 84% and specificity increased to 74%. OCB were positive in 31% patients who did not fulfil BC and in 85% of those with three or four BC. Thirty-one per cent of 146 (37%) patients with normal MRI had positive OCB and 7% (16) of these developed MS. The presence of OB was significantly associated with the development of

CDMS independently of MRI. The authors concluded that the assessment of OCB is of additional diagnostic value to MRI.

Perumal et al. (66) analysed the validity of intrathecal IgG-production in the diagnosis of MS on 58 patients with acute transverse myelitis and normal brain MRI. In the mean follow-up of 5 years, 17 patients developed MS (seven had a second attack, 10 fulfilled the McDonald criteria). All of these 17 patients showed an abnormal CSF (sensitivity 100%) at onset. In addition, 15 of 41 (37%) patients, who did not develop MS, did show intrathecal IgG-synthesis. As this study only shows data of a highly selected subgroup of cases, applicability for clinical use in CIS patients apart from myelitic syndromes is questionable.

Zipoli et al. (67) assessed a cohort of 118 consecutive patients with a first event suggestive of MS over a mean follow-up period of 3.8 years. Ninety cases were classified as CIS while 28 patients were given alternative diagnoses. During follow-up, 78% of CIS patients developed MS according to McDonald criteria, whereas only 56% converted to CDMS showing a second clinical event. OCB were not found in any subject who received an alternative diagnosis, whereas they were found in 12 of the 28 patients fulfilling BC. Patients with OCB and at least two lesions on MRI showed a sensitivity of 80% and specificity of 57% for the development of CDMS during follow-up. Adding OCB to BC increased sensitivity for 8% and a specificity for 17%.

In 2010, Rojas et al. (68) published a cohort of 40 CIS patients from Buenos Aires, followed up for 60 months. Fifteen patients (37%) converted to

CDMS during follow-up, sensitivity was high for MRI and OCB (93%) while specificity was low (MRI: 44%, OCB: 64%). Combining these two methods, sensitivity decreased to 87%, but specificity was clearly increased to 80%.

EVOP studies

Lee et al. (58) applied visual evoked potentials (VEP) and somatosensory evoked potentials (SEP) in a cohort of 184 patients after a first episode suggestive of MS, which they followed up for 2 years. Fifty-five patients (30%) developed CDMS during follow-up. Of the 87 patients showing abnormal VEP 38 (42%) developed CDMS, for SEP it was 35 (40%) of 87 patients with abnormal SEP results. Conversion to CDMS during follow-up could be predicted with only moderate accuracy (Table 4).

Filippini et al. (43) studied 82 patients with suspected MS with visual, somatosensory and brainstem auditory evoked potentials (VEP, SEP, BAEP). In addition to MRI, EVOP did hardly improve accuracy of MS diagnostic tests.

Another study from Milan analysed the prognostic value of EVOP in 100 patients with ON (62). Motoric evoked potentials were obtained in addition to VEP, SEP and BAEP, but with no substantial gain in making a MS diagnosis.

Sastre-Garriga et al. (37) in Barcelona studied 45 patients with clinically isolated syndromes of the brainstem (CISB) with VEP, SEP und BAEP for 3 years. EVOP had no predictive value.

Pelayo et al. (69) studied 245 patients with CIS for a mean follow-up of 76 month with MRI and

Table 4 Diagnostic EVOP studies in MS

Study	Centre	Patients	Follow-up (month)	MS Criteria	EVOP	Sensitivity / specificity (%)	DOR (95% CI)	Consecutive	Prospective
Pelayo (69)	Barcelona (Spain)	245 CIS	76.4	P	BC	65/90	16.2 (9.0–28.9)	+	+
					BC + 1 EVOP	51/97	29.2 (13.1–65.1)	+	+
					BC + 2 EVOP	25/98	15.1 (8.6–40.6)	+	+
					BC + 3 EVOP	10/99	15.4 (1.9–121.4)	+	+
Sastre-Garriga et al. (37)	Barcelona (Spain)	51 CIS	37	P	VEP	47/77	2.9 (0.8–10.8)	?	+
					SEP	33/63	0.9 (0.2–3.2)	?	+
					BAEP	33/57	0.7 (0.2–2.4)	?	+
Ghezzi et al. (62)	Milan (Italy)	100 ON	62	P	VEP	72/25	0.9 (0.3–2.2)	?	+
					BAEP, SEP or MEP	89/18	1.9 (0.4–10.1)	?	+
Filippini et al. (43)	Milan (Italy)	82 suspected MS	35	McDo77	VEP	26/62	0.6 (0.2–1.6)	+	+
					SEP	14/93	2.1 (0.5–9.1)	+	+
					BAEP	18/91	2.1 (0.6–8.1)	+	+
Lee et al. (58)	Vancouver (Canada)	184 suspected MS	25	P	VEP	69/62	3.7 (1.9–7.2)	?	+
					SEP	64/60	2.6 (1.4–5.0)	?	+

+ = yes, ? = unclear.

BAEP, brainstem auditory evoked potentials; BC, Barkhof criteria; CI, confidence interval; CIS, clinically isolated syndrome; DOR, diagnostic odds ratio; EVOP, evoked potentials; Gd McDo, McDonald criteria; McDo 77, McDonald and Halliday 1977 criteria; OCB, oligoclonal bands; ON, optic neuritis; MEP, motoric evoked potentials; P, Poser criteria; SEP, somatosensory evoked potentials; VEP, visually evoked potentials.

multimodal EVOP (VEP, SEP, BAEP) to investigate the additional information of EVOP to MRI. CDMS was diagnosed in 108 patients (44%) and in 70 patients (75%), of those 94 with three or more BC. For BC alone, sensitivity was 65% and specificity was 90%, in combination with abnormal EVOP sensitivity decreased (one abnormal 51%, two abnormal 25%, three abnormal 10%) but specificity increased up to 99% for three abnormal EVOP (two abnormal 98%, one abnormal 97%).

Discussion

This review aimed at analysing accuracy of diagnostic tests for MS compared to the gold standard of Poser criteria (12). Taken together, all the considered MRI studies demonstrate that McDonald criteria (1) lead to a strong increase in early MS diagnoses. With further developments of MRI criteria used in combination with CSF findings, even more patients will be diagnosed. However, studies showed a wide range of sensitivities and specificities. Sensitivity of BC in the analysed MRI studies was reported between 35% and 100%, specificity between 36% and 92%. Apart from differences in cohort selections, development of MRI methodology may account for the observed differences. Recently, simplified MRI criteria have been proposed (29–31, 76). Applied in multicenter cohorts, their approaches seem to show enhanced sensitivity with only a minor loss in specificity. Nevertheless, DOR of Swanton criteria differ between 1.4 and 22.3 in different populations.

Since very few studies have systematically evaluated the value of CSF, results are hardly comparable. This is partly explained by differences in analysis methodologies (70). Among the analysed studies, sensitivities of OCB differed between 69% and 91%, specificities between 59% and 94%.

EVOP are included in MS diagnostic criteria for PPMS but studies on their value in diagnosing MS are scarce, and current data do not justify applying them to diagnose MS (71).

Only recent studies addressed the value of the combination of diagnostic tests. While Sastre-Garriga et al. demonstrated a decrease in sensitivity and an increase in specificity by combining MRI and CSF parameters (37), Villar et al. and Zipoli et al. (64, 67) showed an increase in both when using new MRI criteria. The study of Villar with a follow-up of 6 years gives the strongest indication that combination of simple MRI criteria with an indicator of intrathecal IgG-production might be the best strategy to rule in or out MS. But sample size was low and confirmation in another centre necessary to advocate this approach.

The validation of studies for a new diagnostic test requires three major quality factors: accuracy of the reference standard, representativeness of the study cohort and information on the time period between test and reference standard (18, 72). As indicated in the QUADAS ratings, MS diagnostic studies are hampered by substantial quality deficits, some of them based on the many scientific uncertainties in this disease. In MS, as in many other neurological diseases, a diagnostic gold standard does not exist (73). Therefore, Poser criteria have been used as the best available reference standard for the diagnosis of relapsing–remitting MS with the corollary that evolution over time is a necessary prerequisite. As a consequence, the assessment of sensitivity, specificity and predictive values of Poser criteria requires longitudinal observation of patients. The same holds true for the diagnosis of primary-progressive MS for which a diagnostic gold standard is even less established. The question arises, which time frame is needed to diagnose MS according to Poser criteria. In a longitudinal study of 107 patients with CIS followed up for 20 years, 43% of the patients converted to CDMS within 5 years, 50% within 10 years and an additional 20% converted within 20 years (35). In contrast, in a cohort of patients with ON followed up for a mean of 30 year, no patient converted to CDMS later than after 14 years (57). Therefore, a minimum follow-up of 10 years has been postulated to be necessary to validate a diagnostic test in MS (18), an approach that would reduce the number of relevant MRI studies drastically. Short-term diagnostic studies have constitutively a low specificity as only longitudinal observation enables a clear diagnosis. Also, a short follow-up may lead to an underestimation of sensitivity as a considerable number of patients may already have MS but not yet a second relapse.

Also, representativeness of a given sample is not straightforward in MS. Spectrum and selection bias may strongly influence pretest probability and with this the validity of test results. Studies on diagnostic tests have often investigated already diagnosed patients compared to normal individuals or subsequently excluded patients with other diagnoses from the studied cohort. These selection processes lead to overestimation of the discriminatory power of a given test (15). O'Connor et al. (74) nicely showed in 303 patients how different pretest probabilities, which are hardly considered in MS diagnostic studies, influence sensitivities and specificities of Paty-MRI criteria.

Despite the number of clinical studies in patients with CIS and the approval of several disease-

modifying treatments, CIS lacks a clear clinical definition (14). While the onset of ON, myelitis and brainstem syndromes is considered ‘typical CIS’ when occurring in young adults, concrete symptoms at disease manifestation may make the diagnosis of ‘typical CIS’ difficult. For instance, minor visual field defects without slowing abnormalities in VEP may be a manifestation of ON but the diagnosis of CIS is not straightforward. Likewise, a first episode of ON at age of 50 might raise hesitations on making the diagnosis of CIS. These issues become even more relevant in CIS with atypical clinical presentations. Even though the criterion ‘typical CIS’ is not tight enough to have distinct comparable cohorts, many studies described the diagnostic cohorts as ‘patients with suspected MS’. Taken together, all these factors demonstrate that a clear definition of the clinical presentation is needed to address differences in test accuracy.

Interestingly, only one study addressed the relevance of clinical details for making a MS diagnosis. Nielsen (49) showed in the BENEFIT placebo cohort that multifocal vs. monofocal presentation showed no difference in their predictive value for the development of CDMS. Since Neuhaus et al. (75) have indicated that Nadir EDSS, i.e. the lowest disability level after a relapse, is the strongest predictor for further disease course, clinical evolution might nevertheless add to clarify diagnosis and prognosis and need further studying. Most studies do not address the effect of treatments on test performance. Moraal (50) demonstrated no major difference in test accuracy between early vs delayed treated patients in BENEFIT.

Follow-up data of early treatment studies comparing early and 2-year delayed treatment up to 5 years of follow-up do not allow definite conclusions about the gain of very early treatment in patients with possible MS (77, 78). Therefore, obligatory early diagnostic testing in patients with symptoms suggestive of MS is questionable.

To our knowledge, no study reported how intermediate test results were handled, i.e. questionable lesions on MRI or 1–2 OCB in the CSF leading to lowest QUADAS ratings in item 12.

Limitations of this review are that we did not search databases other than PubMed and we focussed on three languages only. However, we do not think that we would have missed larger studies with longer observation periods, which might have altered the result of our review.

The efficacy of diagnostic tests in MS in relation to patients’ well-being is still unknown as the tests’ impact on clinically relevant endpoints has never

been investigated. This could only be determined by randomized controlled trials (79, 80). Ideally, a new diagnostic test should positively influence relevant health outcomes and result in considerable improvement in mortality, morbidity, symptom control or quality of life (79, 81). This view on the value of diagnostic tests might contradict patients’ and clinicians’ intuition that results of diagnostic tests are always valuable, particularly if the test is not painful and without directly tangible, unwanted side effects. Against this background of the many uncertainties related to diagnostic testing in MS, patients should be involved in diagnostic and decision-making processes as recently underlined by the British General Medical Council (82). This is only possible through balanced, evidence-based, validated information on diagnostic tests, which have recently been proposed (83).

Acknowledgement

S. Köpke is supported by the National MS Society of the US (NMSS).

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Detailed QUADAS ratings

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

References

1. McDONALD WI, COMPSTON A, EDAN G 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.
2. POLMAN CH, REINGOLD SC, EDAN G et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the “McDonald Criteria”. *Ann Neurol* 2005;**58**:840–6.
3. COMI G, FILIPPI M, BARKHOF F et al. Effect of early interferon treatment on conversion to definite multiple sclerosis: a randomised study. *Lancet* 2001;**357**:1576–82.
4. KAPPOS L, POLMAN CH, FREEDMAN MS et al. Treatment with interferon beta-1b delays conversion to clinically definite and McDonald MS in patients with clinically isolated syndromes. *Neurology* 2006;**67**:1242–9.
5. JACOBS LD, BECK RW, SIMON JH et al. Intramuscular interferon beta-1a therapy initiated during a first demyelinating event in multiple sclerosis. CHAMPS Study Group. *N Engl J Med* 2000;**343**:898–904.
6. GOODIN DS. The causal cascade to multiple sclerosis: a model for MS pathogenesis. *PLoS ONE* 2009;**4**:e4565.
7. FREEDMAN MS. ‘Time is brain’ also in multiple sclerosis. *Mult Scler* 2009;**15**:1133–4.
8. COMPSTON A, COLES A. Multiple sclerosis. *Lancet* 2008;**372**:1502–17.

9. SCHUMACKER GA, BEEBE G, KIBLER RF et al. Problems of experimental trials of therapy in multiple sclerosis: report by the panel on the evaluation of experimental trials of therapy in multiple sclerosis. *Ann N Y Acad Sci* 1965;**122**:552–68.
10. MCALPINE D, LUMSDEN CE, ACHESON ED. Multiple sclerosis. A reappraisal. London: Churchill Livingstone, 1972; 83–307.
11. MCDONALD WI, HALLIDAY AM. Diagnosis and classification of multiple sclerosis. *Br Med Bull* 1977;**33**:4–9.
12. POSER CM, PATY DW, SCHEINBERG L et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann Neurol* 1983;**13**:227–31.
13. BARKHOF F, FILIPPI M, MILLER DH et al. Comparison of MRI criteria at first presentation to predict conversion to clinically definite multiple sclerosis. *Brain* 1997;**120**(Pt 11):2059–69.
14. TINTORÉ M, ROVIRA A, MARTINEZ MJ et al. Isolated demyelinating syndromes: comparison of different MR imaging criteria to predict conversion to clinically definite multiple sclerosis. *AJNR Am J Neuroradiol* 2000;**21**:702–6.
15. POSER CM, BRINAR VV. Diagnostic criteria for multiple sclerosis: an historical review. *Clin Neurol Neurosurg* 2004;**106**:147–58.
16. CHARIL A, YOUSRY TA, ROVARIS M et al. MRI and the diagnosis of multiple sclerosis: expanding the concept of “no better explanation”. *Lancet Neurol* 2006;**5**:841–52.
17. MILLER DH, WEINSHENKER BG, FILIPPI M et al. Differential diagnosis of suspected multiple sclerosis: a consensus approach. *Mult Scler* 2008;**14**:1157–74.
18. WHITING P, HARBORD R, MAIN C et al. Accuracy of magnetic resonance imaging for the diagnosis of multiple sclerosis: systematic review. *BMJ* 2006;**332**:875–84.
19. HALLIDAY AM, MCDONALD WI, MUSHIN J. Delayed visual evoked response in optic neuritis. *Lancet* 1972;**1**:982–5.
20. TOURTELLOTTE WW, PARKER JA. Multiple sclerosis: correlation between immunoglobulin-G in cerebrospinal fluid and brain. *Science* 1966;**154**:1044–5.
21. FELGENHAUER K, SCHADLICH HJ, NEKIC M, ACKERMANN R. Cerebrospinal fluid virus antibodies. A diagnostic indicator for multiple sclerosis? *J Neurol Sci* 1985;**71**:291–9.
22. LANGER-GOULD A, POPAT RA, HUANG SM et al. Clinical and demographic predictors of long-term disability in patients with relapsing-remitting multiple sclerosis: a systematic review. *Arch Neurol* 2006;**63**:1686–91.
23. WHITING PF, WESWOOD ME, RUTJES AW, REITSMA JB, BOSSUYT PN, KLEIJNEN J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med Res Methodol* 2006;**6**:9.
24. DEEKS JJ. Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001;**323**:157–62.
25. HAINES TP, HILL K, WALSH W, OSBORNE R. Design-related bias in hospital fall risk screening tool predictive accuracy evaluations: systematic review and meta-analysis. *J Gerontol A Biol Sci Med Sci* 2007;**62**:664–72.
26. 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**:iii, 1–234.
27. FAZEKAS F, OFFENBACHER H, FUCHS S et al. Criteria for an increased specificity of MRI interpretation in elderly subjects with suspected multiple sclerosis. *Neurology* 1988;**38**:1822–5.
28. PATY DW, OGER JJ, KASTRUKOFF LF et al. MRI in the diagnosis of MS: a prospective study with comparison of clinical evaluation, evoked potentials, oligoclonal banding, and CT. *Neurology* 1988;**38**:180–5.
29. SWANTON JK, FERNANDO K, DALTON CM et al. Modification of MRI criteria for multiple sclerosis in patients with clinically isolated syndromes. *J Neurol Neurosurg Psychiatry* 2006;**77**:830–3.
30. ROVIRA A, SWANTON J, TINTORE M et al. A single, early magnetic resonance imaging study in the diagnosis of multiple sclerosis. *Arch Neurol* 2009;**66**:587–92.
31. KORTEWEG T, TINTORE M, UITDEHAAG BM et al. A search for new MRI criteria for dissemination in space in subjects with a clinically isolated syndrome. *Eur Radiol* 2009;**19**:2244–8.
32. MINNEBOO A, BARKHOF F, POLMAN CH, UITDEHAAG BM, KNOL DL, CASTELJNS JA. Infratentorial lesions predict long-term disability in patients with initial findings suggestive of multiple sclerosis. *Arch Neurol* 2004;**61**:217–21.
33. DALTON CM, BREX PA, MISZKIEL KA et al. New T2 lesions enable an earlier diagnosis of multiple sclerosis in clinically isolated syndromes. *Ann Neurol* 2003;**53**:673–6.
34. DALTON CM, BREX PA, MISZKIEL KA et al. Spinal cord MRI in clinically isolated optic neuritis. *J Neurol Neurosurg Psychiatry* 2003;**74**:1577–80.
35. FISNIKU LK, BREX PA, ALTMANN DR et al. Disability and T2MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. *Brain* 2008;**131**:808–17.
36. SWANTON JK, FERNANDO KT, DALTON CM et al. Early MRI in optic neuritis: the risk for clinically definite multiple sclerosis. *Mult Scler*. 2010;**16**:156–65.
37. SASTRE-GARRIGA J, TINTORE M, ROVIRA A et al. Conversion to multiple sclerosis after a clinically isolated syndrome of the brainstem: cranial magnetic resonance imaging, cerebrospinal fluid and neurophysiological findings. *Mult Scler* 2003;**9**:39–43.
38. SASTRE-GARRIGA J, TINTORE M, ROVIRA A et al. Specificity of Barkhof criteria in predicting conversion to multiple sclerosis when applied to clinically isolated brainstem syndromes. *Arch Neurol* 2004;**61**:222–4.
39. TINTORÉ M, ROVIRA A, RIO J et al. Baseline MRI predicts future attacks and disability in clinically isolated syndromes. *Neurology* 2006;**67**:968–72.
40. KORTEWEG T, TINTORE M, UITDEHAAG B et al. MRI criteria for dissemination in space in patients with clinically isolated syndromes: a multicentre follow-up study. *Lancet Neurol* 2006;**5**:221–7.
41. SWANTON JK, ROVIRA A, TINTORE M et al. MRI criteria for multiple sclerosis in patients presenting with clinically isolated syndromes: a multicentre retrospective study. *Lancet Neurol* 2007;**6**:677–86.
42. PAOLINO E, FAINARDI E, RUPPI P et al. A prospective study on the predictive value of CSF oligoclonal bands and MRI in acute isolated neurological syndromes for subsequent progression to multiple sclerosis. *J Neurol Neurosurg Psychiatry* 1996;**60**:572–5.
43. FILIPPINI G, COMI GC, COSI V et al. Sensitivities and predictive values of paraclinical tests for diagnosing multiple sclerosis. *J Neurol* 1994;**241**:132–7.
44. JACOBS LD, KABA SE, MILLER CM, PRIORE RL, BROWNSCHIEDLE CM. Correlation of clinical, magnetic resonance imaging, and cerebrospinal fluid findings in optic neuritis. *Ann Neurol* 1997;**41**:392–8.
45. BECK RW, TROBE JD, MOKE PS 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.
46. MASJUAN J, ALVAREZ-CERMENO JC, GARCIA-BARRAGAN N et al. Clinically isolated syndromes: a new oligoclonal band test

- accurately predicts conversion to MS. *Neurology* 2006;**66**:576–8.
47. ROCCA MA, AGOSTA F, SORMANI MP et al. A three-year, multi-parametric MRI study in patients at presentation with CIS. *J Neurol* 2008;**255**:683–91.
 48. JAFARI N, KREFT KL, FLACH HZ, JANSSENS AC, HINTZEN RQ. Callosal lesion predicts future attacks after clinically isolated syndrome. *Neurology* 2009;**73**:1837–41.
 49. NIELSEN JM, POHL C, POLMAN CH et al. MRI characteristics are predictive for CDMS in monofocal, but not in multifocal patients with a clinically isolated syndrome. *BMC Neurol* 2009;**9**:19.
 50. MORAAL B, POHL C, UITDEHAAG BM et al. Magnetic resonance imaging predictors of conversion to multiple sclerosis in the BENEFIT study. *Arch Neurol* 2009;**66**:1345–52.
 51. LO CP, KAO HW, CHEN SY et al. Prediction of conversion from clinically isolated syndrome to clinically definite multiple sclerosis according to baseline MRI findings: comparison of revised McDonald criteria and Swanton modified criteria. *J Neurol Neurosurg Psychiatry* 2009;**80**:1107–9.
 52. CHAUSSON N, OLINDO S, SIGNATE A, SMADJA D, CABRE P. Evaluation of the 2005 McDonald MRI criteria for dissemination in space in Afro-Caribbean patients with clinically isolated syndromes. *Eur J Neurol* 2009;**16**:1191–6.
 53. DIAZ-SANCHEZ M, MAYRA GOMEZ-MORENO S, ASUNCION MORALES-OTAL M, RAMOS-GONZALEZ A, BENITO-LEON J. Accuracy of MRI criteria for dissemination in space for the diagnosis of multiple sclerosis in patients with clinically isolated syndromes. *Mult Scler* 2010;**16**:576–80.
 54. MOULIN D, PATY DW, EBERS GC. The predictive value of cerebrospinal fluid electrophoresis in 'possible' multiple sclerosis. *Brain* 1983;**106**(Pt 4):809–16.
 55. CORRIDORI F, SALMAGGI A, BORTOLAMI C et al. Prognostic value of cerebrospinal fluid electrophoresis in optic neuritis and suspected multiple sclerosis. *Ital J Neurol Sci* 1987;**6**:77–80.
 56. COLE SR, BECK RW, MOKE PS, KAUFMAN DI, TOURTELLOTTE WW. The predictive value of CSF oligoclonal banding for MS 5 years after optic neuritis. *Optic Neuritis Study Group. Neurology* 1998;**51**:885–7.
 57. NILSSON P, LARSSON EM, MALY-SUNDGREN P, PERFEKT R, SANDBERG-WOLLHEIM M. Predicting the outcome of optic neuritis: evaluation of risk factors after 30 years of follow-up. *J Neurol* 2005;**252**:396–402.
 58. LEE KH, HASHIMOTO SA, HOOGE JP et al. Magnetic resonance imaging of the head in the diagnosis of multiple sclerosis: a prospective 2-year follow-up with comparison of clinical evaluation, evoked potentials, oligoclonal banding, and CT. *Neurology* 1991;**41**:657–60.
 59. JIN YP, PEDRO-CUESTA J, HUANG YH, SODERSTROM M. Predicting multiple sclerosis at optic neuritis onset. *Mult Scler* 2003;**9**:135–41.
 60. ROLAK LA, BECK RW, PATY DW, TOURTELLOTTE WW, WHITAKER JN, RUDICK RA. Cerebrospinal fluid in acute optic neuritis: experience of the optic neuritis treatment trial. *Neurology* 1996;**46**:368–72.
 61. SODERSTROM M, YA-PING J, HILLERT J, LINK H. Optic neuritis: prognosis for multiple sclerosis from MRI, CSF, and HLA findings. *Neurology* 1998;**50**:708–14.
 62. GHEZZI A, MARTINELLI V, TORRI V et al. Long-term follow-up of isolated optic neuritis: the risk of developing multiple sclerosis, its outcome, and the prognostic role of para-clinical tests. *J Neurol* 1999;**246**:770–5.
 63. TINTORÉ M, ROVIRA A, RIO J et al. New diagnostic criteria for multiple sclerosis: application in first demyelinating episode. *Neurology* 2003;**60**:27–30.
 64. VILLAR LM, GARCIA-BARRAGAN N, SADABA MC et al. Accuracy of CSF and MRI criteria for dissemination in space in the diagnosis of multiple sclerosis. *J Neurol Sci* 2008;**266**:34–7.
 65. TINTORÉ M, ROVIRA A, RIO J et al. Do oligoclonal bands add information to MRI in first attacks of multiple sclerosis? *Neurology* 2008;**70**:1079–83.
 66. PERUMAL J, ZABAD R, CAON C et al. Acute transverse myelitis with normal brain MRI : long-term risk of MS. *J Neurol* 2008;**255**:89–93.
 67. ZIPOLI V, HAKIKI B, PORTACCIO E et al. The contribution of cerebrospinal fluid oligoclonal bands to the early diagnosis of multiple sclerosis. *Mult Scler* 2009;**15**:472–8.
 68. ROJAS JI, PATRUCCO L, CRISTIANO E. Oligoclonal bands and MRI in Clinically Isolated Syndromes: predicting conversion time to multiple sclerosis. *J Neurol* 2010;**257**:1188–91.
 69. PELAYO R, MONTALBAN X, MINOVES T et al. Do multimodal evoked potentials add information to MRI in clinically isolated syndromes? *Mult Scler* 2010;**16**:55–61.
 70. FREEDMAN MS, THOMPSON EJ, DEISENHAMMER F et al. Recommended standard of cerebrospinal fluid analysis in the diagnosis of multiple sclerosis: a consensus statement. *Arch Neurol* 2005;**62**:865–70.
 71. GRONSETH GS, ASHMAN EJ. Practice parameter: the usefulness of evoked potentials in identifying clinically silent lesions in patients with suspected multiple sclerosis (an evidence-based review): Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 2000;**54**:1720–5.
 72. LEEFLANG MM, DEEKS JJ, GATSONIS C, BOSSUYT PM. Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 2008;**149**:889–97.
 73. KNOTTNERUS JA, VAN WEEL C, MURIS JW. Evaluation of diagnostic procedures. *BMJ* 2002;**324**:477–80.
 74. O'CONNOR PW, TANSAY CM, DETSKY AS, MUSHLIN AI, KUCHARCZYK W. The effect of spectrum bias on the utility of magnetic resonance imaging and evoked potentials in the diagnosis of suspected multiple sclerosis. *Neurology* 1996;**47**:140–4.
 75. NEUHAUS A, GÜNTNER M, MORRISSEY SP. Risk factors for conversion to multiple sclerosis in patients with a clinically isolated syndrome, ECTIMS October 11, 2007; Abstr. 16.
 76. MONTALBAN X, TINTORE M, SWANTON J et al. MRI criteria for MS in patients with clinically isolated syndromes. *Neurology* 2010;**74**:427–34.
 77. KINKEL RP, KOLLMAN C, O'CONNOR P et al. IM interferon beta-1a delays definite multiple sclerosis 5 years after a first demyelinating event. *Neurology* 2006;**66**:678–84.
 78. KAPPOS L, FREEDMAN MS, POLMAN CH et al. Effect of early versus delayed interferon beta-1b treatment on disability after a first clinical event suggestive of multiple sclerosis: a 3-year follow-up analysis of the BENEFIT study. *Lancet* 2007;**370**:389–97.
 79. GLUUD C, GLUUD LL. Evidence based diagnostics. *BMJ* 2005;**330**:724–6.
 80. WINDELER J. Prognosis – what does the clinician associate with this notion? *Stat Med* 2000;**19**:425–30.
 81. SCHUNEMANN HJ, OXMAN AD, BROZEK J et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 2008;**336**:1106–10.
 82. GENERAL MEDICAL COUNCIL. Consent: patients and doctors making decisions together. London: British General Medical Council, 2008; http://www.gmc-uk.org/guidance/ethical_guidance/consent_guidance/index.asp
 83. HEESSEN C, SCHÄFFLER N, KASPER J, MÜHLHAUSER I, KÖPKE S. Suspected multiple sclerosis – what to do? Evaluation of a patient information leaflet. *Mult Scler* 2009;**15**:1103–12.