NEURO

T. Korteweg
M. Tintore
B. M. J. Uitdehaag
D. L. Knol
H. Vrenken
A. Rovira
J. Frederiksen
D. H. Miller
K. Fernando
M. Filippi
F. Agosta
M. A. Rocca
F. Fazekas
C. Enzinger
A. Parry
C. H. Polman
X. Montalban
F. Barkhof

A search for new MRI criteria for dissemination in space in subjects with a clinically isolated syndrome

T. Korteweg (✉) · F. Barkhof
Department of Radiology,
VU University Medical Centre,
De Boelelaan 1118,
1081 HV,
Amsterdam, The Netherlands
e-mail: T.Korteweg@vumc.nl
Tel.: +31-204-442852
Fax: +31-204-442831

M. Tintore · X. Montalban
Department of Neuroimmunology,
Hospital Vall d'Hebron,
Barcelona, Spain

B. M. J. Uitdehaag · D. L. Knol
Department of Clinical Epidemiology
and Biostatistics,
VU University Medical Centre,
Amsterdam, The Netherlands

B. M. J. Uitdehaag · C. H. Polman
Department of Neurology,
VU University Medical Centre,
Amsterdam, The Netherlands

H. Vrenken
Department of Physics and Medical
Technology,
VU University Medical Centre,
Amsterdam, The Netherlands

A. Rovira
Department of Radiology,
Magnetic Resonance Unit,
Hospital Vall d'Hebron,
Barcelona, Spain

J. Frederiksen
The MS Clinic, Department of
Neurology, University of Copenhagen,
Glostrup Hospital,
Copenhagen, Denmark

D. H. Miller · K. Fernando
MS NMR Research Unit, Department
of Neuroinflammation, Institute of
Neurology, University College London,
London, United Kingdom

M. Filippi · F. Agosta · M. A. Rocca
Neuroimaging Research Unit,
Department of Neurology,
Scientific Institute and University
Ospedale San Raffaele,
Milan, Italy

F. Fazekas · C. Enzinger
Department of Neurology,
Medical University Graz,
Graz, Austria

A. Parry
Department of Clinical Neurology,
University of Oxford,
John Radcliffe Hospital,
Oxford, United Kingdom

Abstract The International Panel on the Diagnosis of Multiple Sclerosis (MS) incorporated the Barkhof/Tintoré (B/T) magnetic resonance criteria into their diagnostic scheme to provide evidence of dissemination in space of central nervous system lesions, a prerequisite for diagnosing MS in patients who present with clinically isolated syndromes (CIS). Although specific for MS, the B/T criteria were criticised for their low sensitivity and relative complexity in clinical use. We used lesion characteristics at onset from 349 CIS patients in logistic regression and recursive partitioning modelling in a search for simpler and more sensitive criteria, while maintaining current specificity. The resulting models, all based on the presence of periventricular and deep white matter lesions, performed roughly in agreement with the B/T criteria, but were unable to provide higher diagnostic accuracy based on information from a single scan. Apparently, findings from contrast-enhanced and follow-up magnetic resonance scans are needed to improve the diagnostic algorithm.

Keywords Magnetic resonance imaging · Multiple sclerosis · Diagnosis

Introduction

Magnetic resonance imaging (MRI) is often used to support the diagnosis of multiple sclerosis (MS). The McDonald International Panel (IP) on the diagnosis of multiple sclerosis incorporated the Barkhof/Tintoré MRI criteria into their diagnostic scheme to gather specific and objective evidence of dissemination in space of central nervous system lesions [1]. Additionally, the IP formulated criteria for the use of MRI to demonstrate dissemination in time, allowing for an earlier diagnosis of MS to be made in patients with a clinically isolated syndrome (CIS).

The Barkhof/Tintoré (B/T) criteria for dissemination in space consist of four individual MRI criteria including one or more juxtacortical lesions, one or more enhancing lesions, one or more infratentorial lesions and three or more periventricular lesions. A threshold is set as at least three positive criteria with the possibility to substitute the need for an enhancing lesion with nine or more T2 lesions [2, 3]. The four individual criteria were derived using logistic regression in a sample of 74 CIS patients, predicting the development of clinically definite multiple sclerosis (CDMS) with increasing risk as more criteria were fulfilled [2]. The threshold, or dichotomisation, at three criteria and the possibility to substitute an enhancing lesion were determined in another sample of 70 CIS patients [3].

The choice of the cutoff criterion in the B/T criteria was fuelled by a desire to maximise specificity, and this was the reason for incorporation into the IP criteria. Further validation studies have confirmed the high specificity, but also revealed the low sensitivity of the B/T criteria for dissemination in space [4]. Related to this low sensitivity, the IP criteria have been criticised for being overly restrictive, preventing an appropriate diagnosis from being made [5]. Another drawback of the B/T criteria is their relative complexity, requiring the user to have experience before using them reliably in a clinical setting [6].

In this study we sought to increase the accuracy and usability of the current B/T criteria by redefining prediction models based on the predictive properties of a single MRI examination at onset of CIS. We made use of more sophisticated statistical methods and took advantage of the large MAGNIMS database.

Materials and methods

The dataset from the MAGNIMS multicentre follow-up study was used containing regional brain lesion counts of initial MRI examinations at onset of CIS and the clinical follow-up of 532 patients (for details see Korteweg et al. 2006 [4]). From this dataset the 349 cases with a minimal clinical follow-up of 2 years were selected to allow conversion to CDMS. The MRI lesions scored included those needed for the B/T criteria, i.e. number of juxtacortical, periventricular, deep white matter, enhancing and infratentorial lesions. In addition, the following items were scored: the number of corpus callosum, basal ganglia, temporal, brainstem and cerebellar lesions, the number of hypointense lesions and the number of lesions bigger than 5 mm. During clinical follow-up, the occurrence of a second clinical event was assessed, indicating the development of CDMS [7]. Because of the retrospective nature of this study, contrast-enhanced scans were not systematically available, and to prevent selection bias these lesions were left out of the analysis.

Statistical analysis

To ensure independent exploration and validation of prediction models, the whole dataset was randomly divided (2:1) into a training (230 observations) and a test set (119 observations). Two approaches to create a prediction model for conversion to CDMS were evaluated. In the first approach logistic regression was used, modelling all MRI items as independent variables with conversion to CDMS as a dependent variable. Using forward and backward stepwise analysis in the training set (Wald statistic with P value for entry of 0.05 and a P value for removal of 0.10), the relative contribution of individual MRI items was assessed for predicting conversion to CDMS. The items with the highest contribution were selected to build a regression model, using ROC (receiver operating characteristic) curves to determine the optimal cutoff point for dichotomisation of the continuous covariates. The resulting model was applied to the test set to calculate the sensitivity and specificity.

In the second approach a multivariate statistical methodology, recursive partitioning, was used to classify cases as either CIS or CDMS. Recursive partitioning splits the data into segments that are as homogeneous as possible with respect to the dependent variable. The method is non-parametric and non-linear in nature, imposing no implicit assumption regarding the relation between the predictor variables and the dependent variable. At each step, a recursive partitioning procedure determines for each variable a cutoff that optimally splits all of the cases into CIS and CDMS and selects the variable that performs best. It then takes the resulting subpopulations and repeats the process, until no additional partitioning is necessary: either a subpopulation contains one class of cases or the subpopulation is too small to be divided. The final results can be summarised in a series of logical if-then conditions or a decision tree. Additionally, to overcome the decision tree instability, an analysis was performed using Random Forest, proposed by Breiman [8]. This method uses bootstrapping to construct multiple independent decision trees, each constructed with a random subset of the predictor variables. In the end, a simple majority vote from all trees is taken for prediction.

The program RPART, freely available on the Internet, implemented in the R language, was used to generate a decision tree depicting the classification rules generated through recursive partitioning. When growing the tree, we

used priors estimated by the class proportions of the sample and equal misclassification costs for CIS and CDMS. The method used to measure impurity (homogeneity) was based on Gini. Pruning the tree (to correct for overtraining) was undertaken using the 1-standard error (SE) rule described by Breiman et al. [9]. Similar to the first approach, the decision tree was developed in the training set and evaluated using the test set. All analyses were performed using SPSS 11.04 for Apple Macintosh, R version 2.5.1 for Apple Macintosh including RPART 3.1–38 and Random Forest 4.5–19.

Results

The clinical characteristics of the cases included (n=349) are shown in Table 1. The MRI was performed within 3 months after the onset of CIS, and the mean clinical follow-up time was 4.9 years (SD 1.9), ranging from 1 to 10 years. During follow-up, CDMS was diagnosed in 132 cases (37.8%) after a median time to conversion of 14 months (IQR 7.6–33.6) within this group. Those patients converting showed a median of 11 T2 lesions (IQR 3–22) versus a median of 1 lesions (IQR 0–9) in the non-converted group. Conversion occurred in 17 of 113 cases (15.0%) without any abnormalities at MRI examination.

Forward stepwise regression analysis with the continuous MRI covariates in the training set found a significant contribution of deep white matter (B=0.082, p=0.04) and periventricular (B=0.055, p<0.01) lesions for predicting conversion to CDMS. The backward stepwise regression confirmed the importance of these covariates without any significant contribution of other covariates. The number of lesions showing the highest accuracy for predicting CDMS for periventricular and deep white matter lesions was determined using ROC curves, resulting in three lesions for deep white matter and two for periventricular lesions. A dichotomised model was built using these cutoffs and applied in the test dataset. This model showed sensitivity of 0.43 (95% CI: 0.28–0.59), specificity of 0.82 (95% CI: 0.71–0.90) and accuracy of 0.68 (95% CI: 0.59–0.76). An additional analysis including
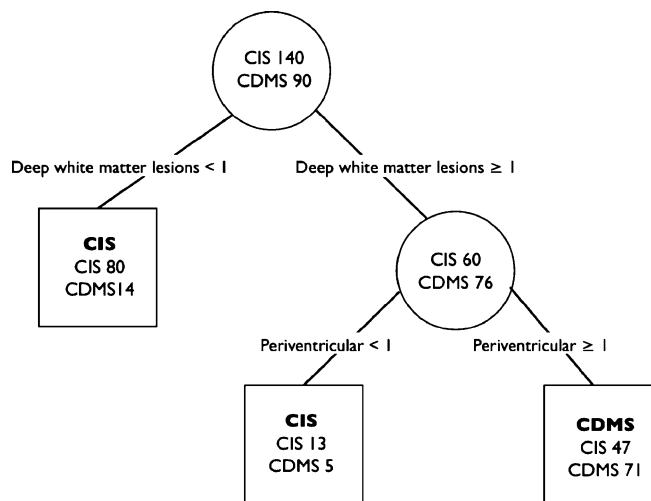
Table 1 Clinical characteristics at onset of CIS

N=349	
Mean age in years (range)	31 (10–52)
Sex (f/m ratio)	235/114
CIS presenting symptoms (%)	
Optic neuritis	164 (47.0)
Spinal cord	80 (22.9)
Brainstem	64 (18.3)
Multiregional	21 (6.0)
Hemispheric	14 (4.0)
Undetermined	6 (1.7)
Median days from onset to MRI (IQR)	23 (10–49)

Fig. 1 Classification tree derived from the training set data. Values represent number of CIS or CDMS cases. The predicted class is displayed in each terminal node of the tree in bold

the age and symptoms at onset in the stepwise regression analysis revealed no significant contribution of these covariates to the model. Furthermore, when added to the analysis, the total number of brain lesions was found to be a single significant covariate predicting conversion, replacing the periventricular and deep white matter lesions.

The classification tree analysis included the same predictors as those found using the regression analysis. The first split was made using the presence of one or more deep white matter lesions, dividing the development dataset into a group of 94 and 136 cases (Fig. 1). The risk of conversion was 15% (14/94) in the left node without deep white matter lesions versus 56% (76/136) in the right node with deep white matter lesions. A final split was made within the group with deep white matter lesions, based on the presence of a periventricular lesion. This split increased the risk of conversion to 60% in the right lower node with a periventricular lesion. When applied to the test set, this model showed sensitivity of 0.64 (95% CI: 0.48–0.78), specificity of 0.70 (95% CI: 0.59–0.80) and accuracy of 0.68 (95% CI: 0.59–0.76). No improvement in accuracy was found using Random Forest analysis. Table 2 summarises the performance of the models in the test set, including the original B/T criteria. Similar to the regression analysis, including the total number of lesions as an independent covariate resulted in a tree with a single split based on the presence of four lesions regardless of their topographical location.

Discussion

This study was fuelled by the desire to find MRI criteria for dissemination in space in the initial MRI examination at onset of CIS that are more sensitive than the existing B/T criteria (without sacrificing specificity) and more simple to

Table 2 Performance of the models in the test set including 95% confidence intervals

*The B/T criteria did not include data on enhancement status

Test dataset (n=119)	Sensitivity	Specificity	Accuracy
B/T cutoff 3*	0.43 (0.28–0.59)	0.87 (0.77–0.94)	0.71 (0.62–0.79)
Regression analysis with cutoffs	0.43 (0.28–0.59)	0.82 (0.71–0.90)	0.68 (0.59–0.76)
Classification tree	0.64 (0.48–0.78)	0.70 (0.59–0.80)	0.68 (0.59–0.76)

use. We took advantage of the large database available from the MAGNIMS collaborative network and applied advanced statistical models. The predictors found were roughly in agreement with those incriminated previously [2, 10–12]. While there was a slight increase in sensitivity (64% versus 49%) and the models were relatively simple, this came at the cost of the slightly reduced specificity (70% versus 79%) and overall similar accuracy (68%) reported earlier for the B/T criteria in this dataset [4]. More complicated models were dismissed, since they tended to "over fit" the data, were instable in the test set and would violate our desire to reduce the complexity of the current B/T criteria.

The failure to achieve higher diagnostic accuracy using information about dissemination in space criteria may reflect several problems inherent in a single brain MRI scan. First, there are many diseases that may produce brain MRI lesions similar to the type seen in MS that cannot be discriminated using simple topological and morphological criteria. Second, incidental brain lesions increase with age (with or without known risk factors), and without the additional use of spinal cord imaging (where age-related changes do not occur) they may interfere with the detection of MS lesions. Third, there are a substantial number of patients without (substantial numbers of) cerebral MRI lesions. Lowering the threshold to include such patients will interfere with specificity. Again, spinal cord imaging may provide significant improvement in this group of patients.

Assuring a diagnosis of MS obviously does not simply rely on demonstration of dissemination in space (using MRI). Demonstration of dissemination in time is equally important, and the IP criteria allow the use of MRI to this end [1]. Using MRI information only, the relative weight of dissemination in time is stronger than that obtained from dissemination in space criteria [13]. In fact, when dissemination in time is fulfilled, less stringent MRI criteria for dissemination in space suffice, whilst guaranteeing a high accuracy. In fact, two clinically silent lesions in locations characteristic of MS (i.e. periventricular, juxtacor-

tical, infratentorial or spinal cord) performed very well in this setting [13].

The strengths of our study included the large sample size from a multi-centre study and the fact that all images were evaluated by the same observer. Patients were recruited from specialised centres and diagnosed by experienced neurologists, resulting in a dataset with predominantly young adults with clinically typical CIS. It may be viewed as a weakness that other diagnoses had already been ruled out before entry into this study. Other weaknesses are the fact that gadolinium was not administered systematically and that the spinal cord was not depicted.

In conclusion, with this approach we have not been able to improve the diagnostic performance of the current B/T criteria for dissemination in space alone using a single unenhanced brain MRI examination. The combined use of gadolinium-enhanced scans with simpler criteria for dissemination in space and follow-up scans to ascertain dissemination in time appears to be a more fruitful avenue.

References

1. McDonald WI, Compston A, Edan G et al (2001) Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. Ann Neurol 50(1):121–127

2. Barkhof F, Filippi M, Miller DH et al (1997) Comparison of MRI criteria at first presentation to predict conversion to clinically definite multiple sclerosis. Brain 120(Pt 11): 2059–2069

3. Tintore M, Rovira A, Martinez MJ et al (2000) Isolated demyelinating syndromes: comparison of different MR imaging criteria to predict conversion to clinically definite multiple sclerosis. AJNR Am J Neuroradiol 21(4):702–706

4. Korteweg T, Tintore M, Uitdehaag B et al (2006) MRI criteria for dissemination in space in patients with clinically isolated syndromes: a multicentre follow-up study. Lancet Neurol 5(3):221–227

5. Frohman EM, Goodin DS, Calabresi PA et al (2003) The utility of MRI in suspected MS: report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. Neurology 61(5):602–611

6. Korteweg T, Uitdehaag BM, Knol DL et al (2007) Interobserver agreement on the radiological criteria of the International Panel on the diagnosis of multiple sclerosis. Eur Radiol 17(1):67–71

7. Poser CM, Paty DW, Scheinberg L et al (1983) New diagnostic criteria for multiple sclerosis: guidelines for research protocols. Ann Neurol 13(3):227–231

8. Breiman L (2001) Random forests. Machine Learning. Vol. 45: Kluwer Academic Publishers

9. Breiman L, Friedman JH, Olshen RA et al (1984) Classification and regression trees. Wadsworth International Group, Belmont, CA

10. Fazekas F, Offenbacher H, Fuchs S et al (1988) Criteria for an increased specificity of MRI interpretation in elderly subjects with suspected multiple sclerosis. Neurology 38(12):1822–1825

11. Paty DW, Oger JJ, Kastrukoff LF et al (1988) MRI in the diagnosis of MS: a prospective study with comparison of clinical evaluation, evoked potentials, oligoclonal banding, and CT. Neurology 38(2):180–185

12. Frederiksen JL, Larsson HBW, Christiansen P et al (1997) Assessment of MRI criteria for abnormal brain MRI in acute monosymptomatic optic neuritis. Comparison to findings in healthy persons. Eur J Neurol 4:561–566

13. Swanton JK, Rovira A, Tintore M et al (2007) MRI criteria for multiple sclerosis in patients presenting with clinically isolated syndromes: a multicentre retrospective study. Lancet Neurol 6(8):677–686