MRI Predictors for disability at 5 years after the diagnosis of MS

Arjan Minneboo, Bas Jasperse, Frederik Barkhof, Bernard M.J. Uitdehaag, Dirk L. Knol, Vincent de Groot, Chris H. Polman, and Jonas A. Castelijns

Submitted
Abstract

Background and purpose: Most studies in Multiple Sclerosis (MS)-patients where longitudinal MRI parameters are used to predict future clinical disability focus on brain MRI parameters. Our goal is to assess predictive value of both brain and spinal cord MRI parameters.

Methods: 83 recently diagnosed MS patients with relapse-onset disease were followed prospectively for a median of 5.5 years. MRI was performed at baseline and after 2 years. MRI parameters included the values of baseline and change over 2 years for brain (T2 and T1 lesion loads, rate of atrophy) and spinal cord (focal lesions and diffuse abnormalities). EDSS scores were available at baseline and after 2- and 5.5-year follow-up (FU). MSFC scores were available in a subgroup of 64 patients. Outcome measures were EDSS and MSFC at 5.5 years. For the EDSS score logistic regression was used (EDSS score <4 or ≥4). Multiple linear regression analysis was used to assess which MRI parameters were associated with MSFC at 5.5 years. All models were corrected for age, sex, disease duration and EDSS score/MSFC at baseline.

Results: At baseline median age was 36.5 years (IQR 29.4–45.0), disease duration was 1.6 years (IQR 0.7–4.1) and disability was low with a median EDSS score of 2.0 (IQR 2.0–3.0), increasing to 3.0 (IQR 2.0–4.5) at 5.5 years FU. At baseline 8 (9.6%) patients had an EDSS score ≥4 increasing to 33 patients (39.8%) at FU. FU EDSS and MSFC were correlated with normalized brain volume at baseline (-0.32, P=0.003; 0.49, P<0.001) and percentage brain volume change (PBVC)/year (-0.21, P=0.057; 0.45, P<0.001). In the multiple regression, EDSS score at 5.5 years was predicted by the presence or absence of diffuse spinal cord abnormalities (OR 8.1, 95% CI 1.4–46.3) and the change in number of segments of focal lesions (OR 1.5, 95% CI 1.1–2.2). The MSFC at 5.5 years was predicted by baseline T2 lesion load and PBVC/year, adjusted R² 0.45.

Conclusions: Both spinal cord- and brain parameters are predictive of development of disability at 5.5 years after the diagnosis MS.
Introduction

In Multiple Sclerosis (MS), progression rate is highly variable between patients. This results in a wide spectrum of outcome ranging from no or minimal disability at long-term follow-up (FU) through being wheelchair bound within years after the diagnosis to death due to MS.\textsuperscript{1-5} Now that effective treatment is available, although at considerable costs and not without side-effects, it becomes crucial to select the patients that benefit most from early treatment.\textsuperscript{6} Consequently it is obvious that much effort is put into the process of identifying predictors for future disability. Magnetic Resonance Imaging (MRI) is one of the most studied paraclinical tests. MRI is widely used not only in diagnosing MS but also as outcome measure in clinical trials and in monitoring disease progression in individual patients.\textsuperscript{7-10} However, most studies do not include spinal cord parameters and only sparse data are available on the predictive value of spinal cord lesion on the progression of disability. Most documentation is on conventional lesion loads as measured on T2- and T1-weighted MRI that are, at best, mildly predictive for (future) disability as measured by the expanded disability status scale (EDSS) or MS Functional Composite (MSFC).\textsuperscript{11-18} Histopathologic studies have pointed out that besides inflammatory changes, axonal loss occurs both within and outside focal MS lesions.\textsuperscript{19,20} Neurodegeneration going undetected on conventional MRI is likely irreversible and associated with clinical disability. The MRI parameter that probably reflects neurodegeneration most closely is brain atrophy. Accumulating evidence suggests that brain atrophy measures are clinically relevant.\textsuperscript{14-18,21-25} Our previous study showed brain atrophy rate measured over the first two years after diagnosis to be the MRI predictor that is most strongly associated with progression of disability during the same interval.\textsuperscript{26} Now that longer FU is available, we evaluated the predictive value of combined brain and spinal cord MRI parameters (both baseline and change over 2 years) for clinical disability, as measured by EDSS and MSFC-scores at 5.5-year FU.
Methods

Patients

Patients originate from a cohort with recently diagnosed MS patients that are prospectively followed-up in a long-term natural history study. The original cohort consisted of 133 patients. Previously, 89 patients with complete MRI- and clinical data at baseline (at the time of the diagnosis) and at FU after a median of 2.2 years (interquartile range (IQR): 2.0–2.4) were studied. Six of these patients declined clinical assessment at 5.5 years and could not be included in the present study. The remaining 83 patients (32 men, 51 women) had further clinical assessment including EDSS after a median of 5.5 years (IQR 5.5–5.6). For a subgroup of 64 patients MSFC scores were available as well. This composite score was calculated for internal population and external population from the time to perform the timed 10-meter walk test (TWT, calculated 25-foot value for comparison with external population), score on paced auditory serial addition test (PASAT) and time to complete the nine-hole PEG test (9HPT). No differences in disease duration, age, use of disease modifying therapy (DMT), EDSS or MSFC at baseline and first FU were observed between studied and excluded patients.

MRI

MRI data were available at baseline and first FU after 2 years. All MRI scans were performed on the same 1.0 Tesla scanner (Magnetom Impact, Siemens, Erlangen, Germany) according to the same scanning protocol. Axial T2 (2700/45, 90) and T1-weighted (700/15) MR images were acquired covering the whole brain: 25 slices with a slice thickness of 5mm, gap 0.5mm. At baseline we acquired T1-weighted MR images before and after the administration of i.v. Gadolinium. At FU no Gadolinium was used. Baseline and FU T2 hyperintense lesion loads (T2LL), T1 hypointense or black holes lesion loads (T1LL) and baseline gadolinium lesion load (GADLL) were quantified using home-developed semi-automated software based on a thresholding technique after identification of lesions by an experienced reader. Baseline Normalized Brain Volume (NBV) and percentage brain volume change (PBVC) were measured on the pre-contrast T1-weighted images using an automated method called SIENAX and SIENA respectively. Spinal cord scanning included a cardiac-triggered sagittal T2-weighted dual-echo spin echo (2400 to
2900/20, 80) and a sagittal T1-weighted spin echo sequence (500/15) with a slice thickness of 3 mm, gap 0.3mm. The number and size (expressed as their extension over a number of corresponding vertebral segments) of spinal cord abnormalities were scored by two readers in consensus. Focal lesions (i.e., sharply delineated areas of increased signal intensity) were considered to be present if seen on intermediate and T2-weighted MRI. Diffuse abnormalities were defined as areas with a subtle, poorly delineated areas of increased signal intensity compared to signal intensity of spinal CSF on intermediate-weighted images.29

Statistics
Because most data were not normally distributed, medians and interquartile range (IQR) were used to describe the data. Spearman rank correlations with two-tailed p-values were used for correlations between clinical and MRI measures. The Mann-Whitney U test was used to test differences in MRI parameters between patients with an EDSS score <4 and EDSS score ≥4 at 5.5 years FU. Pearson chi-square test was used for categorical parameters. The Wilcoxon signed rank test was used to test for differences between baseline and FU values. Two regression models were constructed. Main outcome was disability at last FU (at 5.5 years) as measured by the EDSS or MSFC. The following brain MRI parameters were used in the regression analysis as independent variables: gadolinium-enhancing lesion load (GdLL) at baseline, T2LL at baseline, change in T2LL during the first FU interval (cT2LL), T1LL at baseline, change in T1LL during the first FU interval (cT1LL), NBV at baseline and annualized PBVC. The next spinal cord parameters were used as independent variables: presence or absence of diffuse abnormalities, number of focal cord lesions and number of segments with focal cord abnormalities, change in number of focal cord lesions, change in number of segments. For the EDSS model, logistic regression was used with a forward stepwise selection procedure. The dependent variable was obtained by dichotomizing patients according to EDSS score at 5.5 years: EDSS score <4 or ≥4, as EDSS 4 marks the onset of limitations in mobility. Performance of this model is reported by area under the receiver operator characteristics (ROC) curve. For the MSFC model multiple linear regression was used with a stepwise selection procedure. MSFC score at 5.5 years was the dependent variable. Performance of this model is reported by adjusted R². Both regression models were corrected for relevant clinical parameters e.g.
Chapter 3.3

sex, age, disease duration and EDSS/MSFC score at baseline were entered into the model. Adding data on the use of DMT did not change the models and were therefore not included.

Results

Clinical data
Included patients had a short disease duration (1.6 years, IQR 0.7–4.1) and a median age of 36.5 years (IQR 29.4–45.0) at baseline (Table 3.8). In the subgroup with available MSFC scores at 5.5 years, the age was about 2 years lower (34.1 years, \( P=0.036 \)), no other significant differences were observed. Disability was low with a median EDSS score of 2.0 (IQR 2.0–3.0). Only 8 (9.6%) patients had an EDSS score \( \geq 4 \). Median EDSS increased to 3.0 (IQR 2.0–4.5) at 5.5 years FU; at that time 33 (39.8%) patients had an EDSS score \( \geq 4 \). Baseline and 5.5 years FU scores on the MSFC and its components are reported in Table 3.8 that also shows the MSFC scores compared to an external population. Performance on the PASAT was significantly better at 5.5 years compared to baseline (57.0 compared to 50.0; \( p<0.001 \)) whereas the walking ability (time needed to complete the TWT) and scores on the 9HPT did not change significantly during FU.

MRI data
MRI data for the first 2 years of FU were available. MRI data were not different for the whole cohort and MSFC subgroup. T2LL and T1LL were low at both baseline and at first FU (Table 3.9). Median annual PBVC was -0.9% (IQR -1.4–-0.3). At baseline, 13 (15.7%) patients showed diffuse spinal cord abnormalities whereas 70 (84.3%) patients had focal spinal cord lesions. Median number of focal spinal cord lesions did not increase during FU although 39 (47%) had evidence of at least 1 new focal spinal cord lesion (Figure 3.4).

Correlations between MRI and clinical data
EDSS score at 5.5 years correlated with NBV at baseline \(-0.32, P=0.003\) and a trend was observed for PBVC/year \(-0.21, P=0.057\) but not with lesion loads. Patients with a EDSS score of \( \geq 4 \) at 5.5 years had significantly more diffuse spinal cord
abnormalities than patients with a EDSS score<4 at 5.5 years: 10 out of 33 (30.3%) compared to 3 out of 50 (6.0%), \( P<0.001 \). NBV at baseline was significantly lower in the patients with an EDSS score of 4 or more than in patients with an EDSS score<4 at 5.5 years: 1460 ml compared to 1479 ml (\( P=0.017 \)). Other spinal cord measures or brain MRI parameters did not differ between subgroups. MSFC score at 5.5 years correlated with atrophy measures: strongest correlations were found for NBV at baseline (0.49, \( p<0.001 \)) and PBVC/year (0.45, \( p<0.001 \)) (Figure 3.5). MSFC at 5.5 years was also correlated with brain lesion loads: T1LL at baseline (-0.25, \( P=0.047 \)) and FU (-0.37, \( P=0.003 \)), T2LL at baseline (-0.29, \( p=0.020 \)) and FU (-0.32, \( P=0.011 \)) and cT1LL during FU (-0.34, \( P=0.006 \)) but not with spinal cord measures.

Table 3.8. Clinical characteristics

<table>
<thead>
<tr>
<th>Measurement</th>
<th>EDSS only (n=83)</th>
<th>EDSS and MSFC (n=64)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)</td>
<td>36.5 (29.4–45.0)</td>
<td>34.1 (28.6–40.0)</td>
</tr>
<tr>
<td>Sex (Male–Female)</td>
<td>32–51</td>
<td>18–46</td>
</tr>
<tr>
<td>DMT (Yes–No)</td>
<td>20–63</td>
<td>19–45</td>
</tr>
<tr>
<td>Disease duration at baseline (y)</td>
<td>1.6 (0.7–4.1)</td>
<td>1.5 (0.6–4.1)</td>
</tr>
<tr>
<td>EDSS at baseline</td>
<td>2.0 (2.0–3.0)</td>
<td>2.0 (2.0–2.5)</td>
</tr>
<tr>
<td>EDSS at follow-up 1</td>
<td>2.5 (2.0–3.5)</td>
<td>2.5 (2.0–3.0)</td>
</tr>
<tr>
<td>EDSS at 5.5 y</td>
<td>3.0 (2.0–4.5)</td>
<td>2.5 (2.0–4.0)</td>
</tr>
<tr>
<td>TWT at baseline (sec)</td>
<td>5.4 (4.8–6.3)</td>
<td>50.0 (43.3–56.0)</td>
</tr>
<tr>
<td>PASAT at baseline (correct answers)</td>
<td>19.0 (17.0–22.7)</td>
<td>0.15 (0.55)</td>
</tr>
<tr>
<td>9HPT at baseline (sec)</td>
<td>0.00 (0.76)</td>
<td>5.3 (4.6–6.1)</td>
</tr>
<tr>
<td>MSFC external pop. at baseline*</td>
<td>57.0 (53.3–59.0)</td>
<td>19.3 (17.8–21.7)</td>
</tr>
<tr>
<td>MSFC internal pop. at baseline*</td>
<td>0.64 (0.37)</td>
<td>0.00 (0.74)</td>
</tr>
</tbody>
</table>

Values are presented as means with interquartile range shown between parentheses for both subgroups, * indicates mean with standard deviation between the parentheses. Except for age (\( p=0.036 \), Mann-Whitney \( U \)), no statistical significant differences exist between the subgroups. EDSS = expanded disability status scale; MSFC = MS functional composite; DMT = disease modifying therapy; TWT = timed 10-meter walk test; PASAT = paced auditory serial addition test; 9HPT = nine-hole PEG test.
Figure 3.4. Example of progressive spinal cord abnormalities with new focal lesions. Intermediate- and T2-weighted spin-echo MR images at baseline (A,B) and at follow-up (C,D). At baseline there are no abnormalities but the follow-up images show several focal lesions (indicated by arrows). EDSS score was 1.5 at baseline and 4.0 at follow-up after 5 years.

Regression models
The EDSS model was constructed to find the strongest associations between MRI parameters and EDSS score at 5.5 years. The change in number of segments with spinal cord focal abnormalities (odds ratio (OR) 1.5, 95% confidence interval (CI) 1.1–2.2) as well as the presence or absence of diffuse spinal cord abnormalities (OR 8.1, CI 1.4–46.3) were selected in this model. Performance of this model was good with area under the ROC curve 0.84 (CI 0.75–0.93). When only entering the clinical parameters in the model, area under the ROC curve was 0.80 (CI 0.69–0.90).
Table 3.9. MRI predictors

<table>
<thead>
<tr>
<th>Measurement</th>
<th>EDSS only (n=83)</th>
<th>EDSS and MSFC (n=64)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2LL at baseline (ml)</td>
<td>3.9 (1.7–11.3)</td>
<td>3.8 (1.7–8.5)</td>
</tr>
<tr>
<td>T1LL at baseline (ml)</td>
<td>0.3 (0.0–0.9)</td>
<td>0.3 (0.0–0.7)</td>
</tr>
<tr>
<td>GdLL at baseline (ml)</td>
<td>0.0 (0.0–0.2)</td>
<td>0.0 (0.0–0.2)</td>
</tr>
<tr>
<td>NBV at baseline (ml)</td>
<td>1468 (1420–1513)</td>
<td>1471 (1439–1523)</td>
</tr>
<tr>
<td>T2LL at follow-up (ml)</td>
<td>4.7 (2.0–12.5)</td>
<td>4.4 (2.0–10.3)</td>
</tr>
<tr>
<td>T1LL at follow-up (ml)</td>
<td>0.3 (0.0–0.9)</td>
<td>0.2 (0.0–0.9)</td>
</tr>
<tr>
<td>Change in T2LL/y (ml/y)</td>
<td>0.2 (-0.2–0.7)</td>
<td>0.2 (-0.1–0.7)</td>
</tr>
<tr>
<td>Change in T1LL/y (ml/y)</td>
<td>0.0 (-0.1–0.1)</td>
<td>0.0 (-0.1 0.1)</td>
</tr>
<tr>
<td>PBVC/y (%)</td>
<td>-0.9 (-1.4–0.3)</td>
<td>-0.9 (-1.5–0.3)</td>
</tr>
<tr>
<td>No. focal cord lesion at baseline</td>
<td>3.0 (1.0–4.0)</td>
<td>2.5 (1.0–4.8)</td>
</tr>
<tr>
<td>No. segments with focal abnormalities at baseline</td>
<td>2.0 (1.0–3.3)</td>
<td>2.0 (1.0–3.0)</td>
</tr>
<tr>
<td>Diffuse abnormalities at baseline (Yes-No)</td>
<td>13–70 7–57</td>
<td></td>
</tr>
<tr>
<td>No. focal cord lesion at follow-up</td>
<td>3.0 (2.0–5.0)</td>
<td>3.5 (2.0–5.0)</td>
</tr>
<tr>
<td>No. segments with focal abnormalities at follow-up</td>
<td>2.5 (1.5–5.0)</td>
<td>2.5 (1.6–5.0)</td>
</tr>
<tr>
<td>change in No. focal cord lesion during follow-up</td>
<td>0.5 (0.0–1.0)</td>
<td>0.5 (0.0–1.4)</td>
</tr>
<tr>
<td>change in No. segments with focal abnormalities during follow-up</td>
<td>0.5 (0.0–1.0)</td>
<td>1.0 (0.0–2.0)</td>
</tr>
</tbody>
</table>

Values are presented as means with interquartile range shown between parentheses for both subgroups. No statistical significant differences exist between the subgroups. EDSS = expanded disability status scale; MSFC = MS functional composite; T2LL = T2 lesion load; T1LL = T1 hypointense lesion load, GdLL = gadolinium-enhancing lesion load; NBV = normalized brain volume; PBVC = percentage brain volume change.

For the prediction of the MSFC score at 5.5 years, a stepwise procedure selected T2LL at baseline (beta=-0.35, P=0.003) and PBVC/year (beta=0.30, P=0.005). Thereby, percentage of variance in MSFC score at 5.5 years as explained by the model (adjusted R²) increased from 24% (clinical parameters only) to 45% (clinical and MRI parameters). To exclude the possibility that finding different MRI predictors was caused by differences in characteristics of the subgroups, we repeated the logistic regression analysis for the EDSS model in the MSFC subgroup. This resulted
in similar predictors (diffuse spinal cord abnormalities, change in number of segments with focal cord lesions) and performance of the model (area under the ROC curve 0.81).

**Figure 3.5.** Scatterplot for the correlation between MSFC at follow-up and rate of atrophy. pbvc = percentage brain volume change, MSFC = MS functional composite. Scatterplot for the correlation between MSFC at 5.5 year follow-up and annualized pbvc between baseline and follow-up after two years. Regression line ($r=0.45$, $P<0.001$) with 95% confidence-interval is also shown.

**Discussion**

The current study describes the MRI predictors for disability at 5.5 year FU of patients with relapse-onset MS that were included in the study shortly after the diagnosis. Importantly both brain and spinal cord MRI predictors are assessed. Main findings are that both brain and spinal cord MRI parameters can help to predict disability at 5.5 years. There is much debate about the use of clinical outcome measures in MS. To date, no single best outcome measure exists. All tests have specific weaknesses
and measure different aspects of impairment and disability. We chose to measure disability by the EDSS and MSFC: the two most frequently used clinical rating scales in MS that seem to provide complementary information. The EDSS depends on the measurement of function of several functional systems in the lower regions of the scale. The higher regions of the scale are strongly dependent on leg/walking function. Its main strength is its widespread use and ease of use for the clinician. There are several drawbacks: it is an ordinal and non-equal interval scale, partially subjective, relatively unresponsive to clinical change and it is known that staying times are different for different levels of disability. The MSFC is a composite of three different tests that, when taken together, give an insight in leg function, arm function and cognitive function. Visual and sensory function is not tested. It is frequently used in clinical trials but has also several disadvantages that include practice effects, poor interstudy comparability and poor performance in the more disabled patients. Furthermore, it is time consuming. Amongst its strengths are the good responsiveness to clinical change, and the inclusion of cognitive function. From the above it is clear that the two most used disability rating scales have important limitations and measure different aspects of disability: changes in MSFC are driven by changes in cognitive function and arm function and changes in EDSS by changes in leg function. Since different rating scales measure different aspects of disability and due to the absence of a gold-standard for overall disability, there is also no single best MRI predictor for disability. For example, spinal cord measures will better predict ambulation than cognitive dysfunction. We found the selected MRI predictors to differ according to the used clinical rating scale: EDSS score is predicted by spinal cord measures (diffuse abnormalities and change in focal lesions) whereas the outcome on the MSFC is predicted by the cerebral atrophy rate and T2LL. Similar findings have been reported in patients with primary progressive MS. Most studies on the predictive value of MRI for future clinical deterioration do not evaluate the spinal cord, thereby possibly underestimating the role of MRI. The spinal cord is a clinically eloquent structure and pathology in the cord is likely to become clinically manifest although asymptomatic lesions do occur. Even in early MS spinal cord pathology is common and well depicted by MRI. Several spinal cord measures like atrophy and diffusion tensor derived predictors that could not be embedded in our study are known to correlate with - and predict - disability but the role of conventional spinal cord measures is less clear: although
important for the diagnosis, (focal spinal cord lesions are incorporated in the diagnostic criteria for MS\textsuperscript{7,10}) the correlations between number of focal lesions and disability is poor in most studies and almost no data are available on the predictive value at medium or long-term FU\textsuperscript{34-36} in the previous report on this cohort, evaluating the MRI predictors for EDSS progression over two years, spinal cord parameters did not make it into the final model, probably due to relatively low disability and small changes on the EDSS. The results of our present study indicate that there still is a role for conventional imaging of the spinal cord since the presence of diffuse abnormalities and the increase in focal cord lesions predict the development of locomotor disability (EDSS≥4).

Brain atrophy is an attractive MRI parameter since it can be measured accurately and predicts disability (measured by EDSS or MSFC) better than T2LL\textsubscript{d}.\textsuperscript{22,17,22,37-39} Conceptually, brain atrophy is also interesting: atrophy seems to be the MRI parameter that depicts neurodegeneration most closely and since there are only modest to poor correlations with conventional lesion loads (in the early stages of the disease thought to represent inflammation), additional information is gained when both are studied. In the previous report on this cohort, we showed that rate of brain atrophy as measured by PBVC over two years was the MRI measure most strongly associated with progression of the EDSS score. The current study confirms the predictive value of the PBVC for progression of disability measured by the MSFC over an extended FU period. T2 hyperintense brain lesions are seen in almost all MS patients. Despite their importance in making the diagnosis of MS and predicting conversion to MS and disability in patients presenting with CIS, their role as predictor of disability is not undisputed.\textsuperscript{40-42} T2LL were reported to be poorly predictive for clinical disability in most FU studies.\textsuperscript{22,15,17,21-23,43} In our study T2LL at baseline are predictive for the MSFC score at 5.5 years. Recently T2LL regained attention as predictor for short-term\textsuperscript{43} and long-term FU disability in RR patients and long-term risk factor for conversion from RR to SP MS.\textsuperscript{44,45} Probably monitoring of T2LL in order to predict further disease progression is most effective in the first years after the diagnosis of MS.\textsuperscript{42}

Our study has several limitations. Although the data where collected prospectively, clinical evaluation was relatively infrequent (FU interval was 1 year) and no 3-month control examinations were performed to confirm disease progression. Therefore, for example, we could not use the clinical relevant outcome
measure of time to reach EDSS score of 4. Relatively few patients declined further participation and could not be included for this report (6 patients). Still, many of the included patients had no MSFC scores but only EDSS scores obtained at 5.5 years (19 patients), precluding a one to one comparison of predictors for EDSS and MSFC. Furthermore, the performance on the PASAT improved during FU, probably due to practise effect. These problems with longitudinal registration have been described previously. The most important limitation from imaging perspective is the lack of a volume scan of the spinal cord and the resulting inability to include spinal cord atrophy measures.

In conclusion our results indicate a role for focal and diffuse spinal cord abnormalities as predictor for disability at 5.5 years after the diagnosis of MS as measured by EDSS. We also show the value of brain atrophy rate and T2LL as MRI predictors for clinical disability as measured by MSFC.

Acknowledgements

A. Minneboo is supported by a grant (98-348 MS) of the Dutch MS Research Foundation. The MS Center Amsterdam is supported by the Dutch MS Research Foundation and the VU University Medical Center.

References

Chapter 3.3


