Predicting Clinical Progression in Multiple Sclerosis With the Magnetic Resonance Disease Severity Scale

Rohit Bakshi, MD; Mohit Neema, MD; Brian C. Healy, PhD; Zsuzsanna Liptak, MD; Rebecca A. Betensky, PhD; Guy J. Buckle, MD; Susan A. Gauthier, DO; James Stankiewicz, MD; Dominik Meier, PhD; Svetlana Egorova, MD; Ashish Arora, MD; Zachary D. Guss, BA; Bonnie Glanz, PhD; Samia J. Khoury, MD; Charles R. G. Guttmann, MD; Howard L. Weiner, MD

**Background:** Individual magnetic resonance imaging (MRI) disease severity measures, such as atrophy or lesions, show weak relationships to clinical status in patients with multiple sclerosis (MS).

**Objective:** To combine MS-MRI measures of disease severity into a composite score.

**Design:** Retrospective analysis of prospectively collected data.

**Setting:** Community-based and referral subspecialty clinic in an academic hospital.

**Patients:** A total of 103 patients with MS, with a mean (SD) Expanded Disability Status Scale (EDSS) score of 3.3 (2.2), of whom 62 (60.2%) had the relapsing-remitting, 33 (32.0%) the secondary progressive, and 8 (7.8%) the primary progressive form.

**Main Outcome Measures:** Brain MRI measures included baseline T2 hyperintense (T2LV) and T1 hypointense (T1LV) lesion volume and brain parenchymal fraction (BPF), a marker of global atrophy. The ratio of T1LV to T2LV (T1:T2) assessed lesion severity. A Magnetic Resonance Disease Severity Scale (MRDSS) score, on a continuous scale from 0 to 10, was derived for each patient using T2LV, BPF, and T1:T2.

**Results:** The MRDSS score averaged 5.1 (SD, 2.6). Baseline MRI and EDSS correlations were moderate for BPF, T1:T2, and MRDSS and weak for T2LV. The MRDSS showed a larger effect size than the individual MRI components in distinguishing patients with the relapsing-remitting form from those with the secondary progressive form. Models containing either T2LV or MRDSS were significantly associated with disability progression during the mean (SD) 3.2 (0.3)–year observation period, when adjusting for baseline EDSS score.

**Conclusion:** Combining brain MRI lesion and atrophy measures can predict MS clinical progression and provides the basis for developing an MRI-based continuous scale as a marker of MS disease severity.

Arch Neurol. 2008;65(11):1449-1453

©2008 American Medical Association. All rights reserved.
which measures current severity rather than relapses or acute activity.10 Fifth, we chose a priori to equally weight the components to ensure that the scale was descriptive of the cerebral state and could be used for many different potential comparisons. We refer to this composite as the Magnetic Resonance Disease Severity Scale (MRDSS). We tested the validity of the MRDSS vs clinical status in a 3-year longitudinal study of patients representing a wide range of disability, disease duration, and clinical phenotype.

### METHODS

We retrospectively identified 103 patients (Table 1) from the Comprehensive Longitudinal Investigation of MS at Brigham and Women’s Hospital. Inclusion criteria were as follows: (1) age 18 to 60 years; (2) brain MRI performed with our MS-designated protocol; (3) baseline EDSS scoring26 by a neurologist specializing in MS within 3 months of MRI; (4) follow-up EDSS scoring 2½ to 3½ years later; and (5) established MS12 at baseline (relapsing-remitting [RRMS], secondary progressive [SPMS], or primary progressive [PPMS]).13 Patients with clinically isolated syndromes were not included because many of them will not develop MS.14 All patients except 12 were treated with immunotherapy during the observation period (Table 1). Baseline characteristics (Table 1) were comparable to those of a large population-based MS cohort.15 This study was approved by our institutional review board.

Ninety-six patients (93.2%) had clinical follow-up 3.2 (0.3) years later (mean [SD]). Follow-up disability was categorized as stable or progressed (1-point progression on follow-up EDSS if the baseline score was <6 [or 0.5-point progression if the baseline score was ≥6], sustained for 3 months) (Table 1).

Patients underwent 1.5-T MRI with axial T1-weighted spin-echo (repetition time [TR]/echo time [TE], 725/20 milliseconds) and dual-echo T2-weighted (TR/TE/TE, 3000/80/30 milliseconds) imaging (voxel size, 0.9375 × 0.9375 × 3 mm). T1 imaging was repeated 5 minutes after intravenous administration of gadopentetate dimeglumine, 0.1 mmol/kg. Our analysis, based on a hypothesis of a more informative composite measure, was tested without prior knowledge of the component MRI measures in the sample.

We used template-driven segmentation with partial volume correction16 to determine whole-brain T2LV and brain parenchymal fraction (BPF) (an estimate of whole-brain atrophy)17 from the dual-echo T2-weighted images.16 Although fluid attenuation inversion recovery images would likely have yielded a higher lesion load than dual-echo T2 images, we have not yet developed an automated segmentation method for fluid attenuation inversion recovery and thus relied on our established method. The presence of T1 hypointense lesions (black holes) was determined by consensus of 2 trained observers (including M.N., J.S., and A.A.18) and showed at least partial hyperintensity on T2 images but no gadolinium enhancement (to reduce the likelihood of including transient lesions).19

Template-driven segmentation with partial volume correction achieves an intraclass correlation of 0.994, interscan coefficient of variation of 4.98%, and mean (SD) volume bias of 0.01 (0.68) mL.16 Our T1LV measurement shows intraobserver and interobserver coefficients of variation of 1.7% and 4.5%, respectively.16 In the current cohort, MRDSS achieved intraobserver and interobserver coefficients of variation of 2.3% and 4.4%, respectively.

All calculations necessary to derive the MRDSS score from the MRI data were scaled with the use of the cohort studied, not from external data. The MRI data were rounded to 2 decimal places. When nonrounded data were used instead, the results did not change (data not shown). We did not use absolute T1LV for the MRDSS because of high collinearity with T2LV (r = 0.79). Because the distributions of T2LV and T1:T2 were skewed, log (T2LV) or logistic (T1:T2) transformation was performed. The T2LV, BPF, and T1:T2 were then standardized (z*) by subtracting the mean and dividing by the standard deviation. The individual components of the MRDSS had only moderate intercorrelation with one another, both before and after transformations (r = 0.53-0.58). Patients with zero T1LV were not included in the normalization but were assigned a value more extreme (~2.5) than the most extreme nonzero standardized T1:T2 (~1.92), similar to the MS Functional Composite.20 We selected the more extreme value because patients with zero T1 hypointensities have meaningfully less severe disease than patients with a small amount. The value ~2.5 approximated the magnitude of the difference between the 2 smallest

---

**Table 1. Description of Patients**

<table>
<thead>
<tr>
<th>Characteristic Finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total No. of patients</td>
</tr>
<tr>
<td>Multiple sclerosis type, No.</td>
</tr>
<tr>
<td>Sex ratio, M:F</td>
</tr>
<tr>
<td>Age, y</td>
</tr>
<tr>
<td>Disease duration from first symptoms, y</td>
</tr>
<tr>
<td>Expanded Disability Status Scale score</td>
</tr>
<tr>
<td>Duration of observation period, y</td>
</tr>
<tr>
<td>Brain parenchymal fraction, mean (SD)</td>
</tr>
<tr>
<td>T2 hyperintense lesion volume, mean (SD), cm³</td>
</tr>
<tr>
<td>T1 hypointense lesion volume, mean (SD), cm³</td>
</tr>
<tr>
<td>T1:T2 lesion volume ratio, mean (SD)</td>
</tr>
<tr>
<td>Magnetic Resonance Disease Severity Scale score</td>
</tr>
<tr>
<td>Use of disease-modifying therapy during observation period, No.</td>
</tr>
</tbody>
</table>
values on the standardized scale. The individual standardized scores were equally weighted and summed for each patient: \( \text{MRDSS} = \frac{z_{T2LV}}{3} + \frac{z_{T1LV/T2LV}}{3} + \frac{z_{BPF}}{3} \) (where \( z = \frac{\text{raw score} - \text{mean}}{\text{standard deviation}} \)). Each subject’s composite value was then transformed to a continuous 0 to 10 MRDSS score (zero is lowest severity).

Spearman correlation or the Wilcoxon rank sum test assessed the association between baseline MRDSS score and other variables. Univariate logistic regression tested the association between baseline MRI (BPF, T2LV, T1:T2, and MRDSS) and 3-year clinical progression. Multivariate regression controlled for covariates. The area under the receiver operating characteristic curve investigated the predictive ability of the model. The 95% confidence interval for the area under the curve was the percentiles of a bootstrap distribution. Comparing the MRDSS with the raw MRI data demonstrated the value of the MRDSS over conventional approaches, whereas comparisons with standardized scores investigated the effect of combining measures into a composite scale. \( P < .05 \) was considered significant.

### RESULTS

The distribution of MRDSS scores is shown in Figure 1, Table 1, and Table 2, and representative MRIs are shown in Figure 2. During follow-up, 24 patients developed sustained progression of disability (Table 1); this included 3 of 12 patients (25%) who were not receiving disease-modifying therapy.

Baseline MRI-EDSS comparisons showed moderate correlations for BPF (\( r = -0.47, P < .001 \)), T1:T2 (\( r = 0.46, P < .001 \)), and MRDSS (\( r = 0.48, P < .001 \)) and weak correlation for T2LV (\( r = 0.25, P < .01 \)). The MRI correlations with disease duration were either weakly significant (BPF) or nonsignificant (T2LV, T1:T2, MRDSS) (data not shown). All MRI measures differed between the RRMS and SPMS groups, but more so for MRDSS (Table 2).

When adjusting for baseline EDSS score, only MRDSS and T2LV showed a significant association with clinical progression (Table 3). Standardized T2LV showed a closer association with clinical progression than did nonstandardized T2LV, indicating that standardization contributed to the improvement in MRDSS vs individual MRI measures. In the receiver operating characteristic curves for the ability of MRI to predict clinical progression, both T2LV and MRDSS showed higher areas under the curve than did BPF or T1:T2 (Table 3). The unstandardized T2LV showed the highest area under the curve (best predictive ability), but the 95% confidence interval overlapped with the 95% confidence intervals for standardized T2LV and MRDSS (Table 3). Adding disease duration as a covariate or using the standardized vs raw EDSS score did not affect the models (data not shown).

### COMMENT

The MRDSS encompasses 3 equally weighted whole-brain MRI measures of lesions and atrophy. T1 hypointense lesions are expressed as a ratio to T2LV because only a subset of T2 hypointense lesions are T1 hypointense (particularly destructive lesions). We have developed and tested this scale in a 3-year longitudinal MS cohort with a wide range of disease duration and disease severity, including minimal, mild, moderate, and severe physical disability. The MRDSS has concurrent validity when compared with clinical status. It showed the largest effect size in differentiating the RRMS and SPMS groups compared with the individual MRI components. The MRDSS predicts the risk of developing sustained progression of physical disability 3 years later. However, the MRDSS offers only a modest improvement compared with current metrics.

Although the individual MRI measures also had some association with disability, the MRDSS showed the unique

### Table 2. Comparison of MRI Data Among the MS Clinical Subgroups

<table>
<thead>
<tr>
<th>Clinical Subgroup</th>
<th>RRMS (n = 62)</th>
<th>SPMS (n = 33)</th>
<th>Primary Progressive (n = 8)</th>
<th>RRMS vs SPMS, ( P ) Value</th>
<th>RRMS vs SPMS, ( b ) Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain parenchymal fraction</td>
<td>0.83 (0.05)</td>
<td>0.81 (0.05)</td>
<td>0.83 (0.04)</td>
<td>.005</td>
<td>0.62</td>
</tr>
<tr>
<td>T2 hypointense lesion volume, cm(^3)</td>
<td>6.1 (5.2)</td>
<td>9.5 (7.6)</td>
<td>4.2 (2.8)</td>
<td>.006</td>
<td>0.52</td>
</tr>
<tr>
<td>T1:T2 lesion ratio</td>
<td>0.13 (0.17)</td>
<td>0.24 (0.17)</td>
<td>0.11 (0.11)</td>
<td>&lt;.001</td>
<td>0.67</td>
</tr>
<tr>
<td>Magnetic Resonance Disease Severity Scale score</td>
<td>4.5 (2.7)</td>
<td>6.6 (2.4)</td>
<td>4.6 (2.0)</td>
<td>&lt;.001</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Abbreviations: MRI, magnetic resonance imaging; MS, multiple sclerosis; RRMS, relapsing-remitting MS; SPMS, secondary progressive MS.

\( a \) Wilcoxon rank sum test.

\( b \) Comparisons did not change for standardized brain parenchymal fraction, T2 hypointense lesion volume, and T1:T2 lesion ratio (data not shown).
combination of both concurrent and predictive validity. For example, although baseline MRDSS, BPF, and T1:T2 were moderately correlated with baseline EDSS score, T2LV showed a weak correlation. Furthermore, BPF and T1:T2 were not significantly associated with clinical progression in the regression modeling. In the receiver operating characteristic analysis based on the regression models, all of the confidence intervals were wide, but the estimates were largest for T2LV, followed by MRDSS. Although these findings support the utility of the MRDSS, it is likely that our sample is too small to ensure generalizability, ie, the 0 to 10 scaling of the MRDSS derived from the current cohort may not apply to a larger population.

This was not a natural history study because most patients were receiving therapy during the observation period, which could affect some MRI measures differently and confound the results. Disability progression was defined by EDSS worsening, which is limited by nonlinearity, variability, and heavy weighting toward ambulation. Future studies should test the MRDSS in a larger sample size and assess whether it is related to other clinical manifestations such as cognition.

Previous MS studies combined brain MRI data into composite or multiparametric assessments. The z4 composite6,8,9 combines T2LV, T1LV, BPF, and gadolinium-enhancing lesions. In a follow-up study of patients with RRMS or SPMS, the z4 at 3 months predicted the change in physical disability (P < .02).6 Correlations with EDSS were not reported, nor was it reported whether the z4 had better association with clinical change than did the individual MRI measures. In a cross-sectional study of patients with RRMS, z4 differentiated patients who had been stable, worse, or improved during the previous 2 years; however, brain atrophy showed a stronger association with clinical status than did z4.9 There are notable differences between our MRDSS and the z4. First, we did not include gadolinium-enhancing lesions as part of the scale, based on the rationale presented in the introduction to this article. Second, gadolinium enhancement poorly predicts sus-

![Figure 2. Representative proton-density (PD) and T1-weighted brain magnetic resonance images showing mild (A), moderate (B), and severe (C) Magnetic Resonance Disease Severity Scale (MRDSS) scores. A, Images from a 39-year-old woman with relapsing-remitting multiple sclerosis (MS); mild physical disability (Expanded Disability Status Scale [EDSS] score, 2.5); brain parenchymal fraction (BPF), 0.88; T2 hyperintense lesion volume (T2LV), 4.1 cm³; T1 hypointense lesion volume (T1LV), 0.012 cm³; ratio of T1LV to T2LV (T1:T2), 0.0030; and MRDSS score, 1.93. B, Images from a 53-year-old woman with relapsing-remitting MS; mild to moderate disability (EDSS score, 3.5); BPF, 0.80; T2LV, 8.4 cm³; T1LV, 0.22 cm³; T1:T2, 0.026; and MRDSS score, 5.67. C, Images from a 38-year-old woman with secondary progressive MS; severe disability (EDSS score, 8.0); BPF, 0.77; T2LV, 17.6 cm³; T1LV, 10.83 cm³; T1:T2, 0.61; and MRDSS score, 9.94.](image)

![Table 3. Regression Modeling: Odds Ratios and Area Under the ROC Curve for the Relationship Between Baseline MRI Measure and Risk of Developing Sustained Disability Progression](table)
tained changes in physical disability, whereas the other MRI measures assessed in our study have shown such predictive value. The T1LV is an absolute measure in the z4 but is divided by T2LV in the MRDSS. We chose to use T1:T2 in part because of the collinearity problems related to the high correlation between T1LV and T2LV.

Another group² combined brain T2LV, T1LV, magnetization transfer, diffusion, and magnetic resonance spectroscopy data in a cross-sectional study of 23 patients. The correlations between EDSS and many of the individual MRI metrics were weak. In contrast, composite MRI scores showed better associations with EDSS (P < .004 to <.001). In contrast to the present study, brain atrophy was not considered, nor was the predictive relationship between MRI and longitudinal clinical change.

More validation work is necessary to evaluate the MRDSS. We are in the process of assessing longitudinal MRDSS change, which we shall report in subsequent publications. We are planning to add MRI measures of occult damage, such as magnetization transfer, diffusion imaging, or magnetic resonance spectroscopy, or MRI measures of gray matter²³ and spinal cord damage,²⁶ to the MRDSS. The inclusion of a spinal cord metric, for example, might increase the predictive value for disease progression, especially if the outcome is EDSS score.

Accepted for Publication: May 9, 2008.
Correspondence: Rohit Bakshi, MD, Departments of Neurology and Radiology, Brigham and Women’s Hospital, 77 Avenue Louis Pasteur, Ste HIM 730, Boston, MA 02115 (rbakshi@bwh.harvard.edu).

Additional Contributions: Dr Bakshi had full access to all data and takes responsibility for the integrity of the data and accuracy of the analysis. Study concept and design: Bakshi, Healy, Meier, Khoury, Guttmann, and Weiner. Acquisition of data: Bakshi, Liptak, Buckle, Gauthier, Meier, Egorova, Arora, Glanz, Khoury, Guttmann, and Weiner. Analysis and interpretation of data: Bakshi, Neema, Healy, Betensky, Buckle, Stankiewicz, Meier, Guss, Khoury, Guttmann, and Weiner. Drafting of the manuscript: Bakshi. Critical revision of the manuscript for important intellectual content: Bakshi, Neema, Healy, Betensky, Buckle, Stankiewicz, Meier, Egorova, Arora, Glanz, Khoury, Guttmann, and Weiner. Statistical analysis: Healy and Betensky. Obtained funding: Bakshi, Khoury, Guttmann, and Weiner. Administrative, technical, and material support: Bakshi, Liptak, Buckle, Gauthier, Stankiewicz, Meier, Egorova, Arora, Guss, Glanz, Khoury, Guttmann, and Weiner. Study supervision: Bakshi, Neema, Buckle, Khoury, Guttmann, and Weiner.

Financial Disclosure: None reported.

Funding/Support: This study was supported by research grants from the National Institutes of Health (1R01NS050833-01) and the National Multiple Sclerosis Society (RG3705A1 and RG3798A2).

Role of the Sponsors: The funders had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; and the preparation, review, or approval of the manuscript.

Previous Presentation: This study was presented at the 59th annual meeting of the American Academy of Neurology; May 1, 2007; Boston, Massachusetts.

Additional Contributions: Sophie Tamm, MA, provided assistance with manuscript preparation.

REFERENCES