



Published in final edited form as:

Mult Scler. 2010 September ; 16(9): 1117–1125. doi:10.1177/1352458510374892.

Composite MRI Scores Improve Correlation with EDSS in Multiple Sclerosis

Aziz H. Poonawalla, PhD^{1,2}, Sushmita Datta, PhD¹, Vaibhav Juneja, MS¹, Flavia Nelson, MD³, Jerry S. Wolinsky, MD³, Gary Cutter, PhD⁴, and Ponnada A. Narayana, PhD^{1,*}

¹Department of Diagnostic and Interventional Radiology, University of Texas Medical School at Houston, Houston TX (USA)

²Department of Radiology, University of Wisconsin – Madison, Madison WI (USA)

³Department of Neurology, University of Texas Medical School at Houston, Houston TX (USA)

⁴Department of Biostatistics, University of Alabama at Birmingham School of Public Health, Birmingham AL (USA)

Abstract

Background—Quantitative measures derived from MRI have been widely investigated as noninvasive biomarkers in multiple sclerosis (MS). However, the correlation of single measures with Expanded Disability Status Scale (EDSS) is poor, especially for studies with large population samples.

Objective—To explore the correlation of MRI-derived measures with EDSS through composite MRI scores.

Methods—Magnetic resonance images of 126 patients with relapsing-remitting MS were segmented into white and gray matter, CSF, T2-hyperintense lesions, gadolinium contrast-enhancing lesions, T1-hypointense lesions (“black holes (BH)”). The volumes and average T2 values for each of these tissues and lesions were calculated and converted to a z-score (in units of standard deviation from the mean). These z-scores were combined to construct composite z-scores, and evaluated against individual z-scores for correlation with EDSS.

Results—Composite scores including relaxation times of different tissues and/or volumetric measures generally correlated more strongly with EDSS than individual measures. The maximum observed correlation of a composite with EDSS was $r = 0.344$ ($p < 0.0001$), which is an improvement over the highest-performing single MRI measure (BH; $r = 0.298$, $p < 0.001$).

Conclusion—Z-transformation permits construction of composite scores including volumetric and T2-relaxation measures. Inclusion of multiple MRI measures in the composite can provide a broader characterization of the disease process, resulting in more robust correlations with EDSS.

Keywords

Multiple Sclerosis; EDSS; composite score; MRI; biomarkers

*Corresponding Author: Ponnada A. Narayana, Ph.D., University of Texas Health Science Center at Houston, 6431 Fannin Street, Houston TX 77030, Tel: (713) 500-7677 Fax: (713) 500-7684, ponnada.a.narayana@uth.tmc.edu.

Conflicts of interest: None declared.

Introduction

Multiple Sclerosis (MS) is a chronic disease of the central nervous system, for which there is currently no cure. Any evaluation of therapy must be based on meaningful clinical outcomes. Despite the use of clinical instruments such as the extended disability status score (EDSS) to assess neurological disability [1,2], the highly variable and complex course of MS impedes objective disease evaluation [3]. Therefore, identification of validated noninvasive biomarkers would have an important impact on disease management and assessing treatment efficacy.

Magnetic resonance imaging (MRI) is the most sensitive noninvasive imaging modality for visualization and characterization of MS lesions. There is a general enthusiasm for using MRI measures as biomarkers in clinical trials and treatment evaluation [4-9]. Numerous quantitative measures derived from conventional MRI [10,11] and advanced MRI [5,11-22] have been proposed as biomarkers in MS. However, the correlations between various individual MRI scores and EDSS have been modest and variable. The present consensus, therefore, is that no single MRI-derived measure is sufficient to serve broadly as an imaging biomarker in MS [5, 23].

Since one of the major strengths of MRI is its multi-modal nature, a promising approach is to construct informative composites from a combination of MRI measures, each capturing a different aspect of the disease's clinical course [10,12,23,24]. For example, T2 lesions could represent demyelination, gliosis, edema etc. Similarly, the T1-hypointense lesions could represent edema or tissue necrosis. Therefore, even though the pathologic specificity of each MRI measure is unclear, combinations of these various measures could better capture the pathologic spectrum of the disease.

Different MRI measures have different dimensions, such as lesion volume (measured in cc), tissue atrophy (measured as a volume fraction), or T2 value (measured in milliseconds). In addition, the numerical values of each measure may differ widely, with tissue volumes several orders of magnitude larger than lesion volumes. These limitations may be overcome by employing a z-transformation, which standardizes each measure into a dimensionless parameter (the "z-score") that represents the number of standard deviation units from the mean [25]. After z-transformation, different measures may simply be added together algebraically, providing a more complete representation of the disease process (and thus, correlation to disability) than any single MRI-derived measure alone.

An initial composite measure of this type, incorporating volume of gadolinium-enhancing lesions, total lesion volume ("burden of disease"), T1 hypointense lesion volume, and normalized CSF volume (nCSF) (a measure of brain atrophy), was introduced in the North American Linomide Trial and shown to distinguish treatment groups at 3 months [4]. The same score was also shown to discriminate the duration of therapy and clinical outcomes in the US open label glatiramer acetate (GA) extension trial in relapsing-remitting (RR) MS [24]. More recently, the predictive value of this composite has been demonstrated in a prospectively followed cohort of subjects on GA [26]. While these studies have demonstrated the potential of composite measures for treatment monitoring, the observed correlations remained modest. For example, in the linomide trial, the observed r-value for correlation with EDSS was 0.146, increasing to 0.25 when excluding all but relapsing-remitting (RRMS) patients [4]. In comparison, r-values have been observed as high as 0.60 for correlation of T2-hyperintense lesion volume with EDSS in RRMS patients with a relatively small sample of 39 patients [27].

It is possible that inclusion of other MRI measures in the composite may improve the correlation with EDSS. One promising measure is the transverse relaxation time (T2), which can be quantitatively estimated from the dual-echo scan routinely used in MS acquisition protocols

[28,29]. Quantitative T2 values have been used to probe normal-appearing MS tissue for evidence of subclinical disease activity [30-32]. The sensitivity of T2 to the microscopic disease process suggests that T2 values could complement the conventional volumetry-based measures, which are more macroscopic in scope.

In the present work, we construct new composite measures which incorporate T2 values and conventional measures derived from images commonly acquired during routine patient management or clinical trials, using a relatively large cohort (n = 126) of RRMS patients. These new composite measures were compared and evaluated with respect to correlation with EDSS.

Materials and Methods

Subjects

This study was approved by our institution's Committee for the Protection of Human Subjects and was HIPAA compliant. One hundred and twenty-six RRMS patients (101 female, 25 male) were recruited over a period of 29 months from September 2004 to January 2007. The patients had a median age of 40.7 years (range 20 – 64), and median EDSS score of 1.5 (range 0 – 6.5). Written informed consent was obtained from all participants.

MRI Acquisition

In this cross-sectional study, all patients were imaged once at a single MRI center on a 3T Philips Intera scanner with a Quasar gradient system (maximum gradient amplitude 80 mT/m, slew rate 200 T/m/s) and a 6-channel SENSE-compatible head coil (Philips Medical Systems, Best, Netherlands). The MRI protocol included dual-echo fast spin echo (dFSE, 8.2/90/6800 ms), fluid-attenuated inversion recovery (FLAIR, 80/10000 ms, TI 2600 ms), and pre- and post-gadolinium contrast T1-weighted spin echo (T1W, 9.2/600 ms) images. For each sequence, 44 contiguous and interleaved axial slices of 3 mm thickness were acquired for full-brain coverage with an image matrix of 256×256 and a 24 cm square field of view. All sequences employed a sensitivity encoded (SENSE) acceleration factor of 2.0.

All patient images were reviewed by an experienced neurologist with extensive clinical and imaging experience in MS (JW). This neurologist also performed the EDSS scoring of all subjects based on their examinations prior to imaging.

Processing and Analysis

All image pre-processing and analyses were overseen by two medical physicists specializing in MRI (AP and PN). First, image stripping to remove skull and extract brain was performed on the T2-weighted (long-echo FSE) image, and the resulting tissue mask was applied to the proton density-weighted (short-echo FSE) image. Then, the rigid body registration module in SPM2 [33] was used to align the FLAIR and T1-weighted (T1W) images to the dual-echo FSE space, and the stripping mask applied.

Image segmentation was performed using an automated hybrid parametric/nonparametric algorithm [34,35] implemented on a PC running Interactive Data Language (Research Systems, Inc., Boulder, CO). Images were segmented into white matter (WM), gray matter (GM), CSF, T2-hyperintense lesions [34], gadolinium contrast-enhancing lesions [35], and T1-hypointense lesions (T1 lesions, aka “black holes”) [35]. Note that the T1 lesions are also hyperintense on the T2-weighted images, so for consistency with published literature we define “burden of disease” (BOD) as the total volume of *all* T2-hyperintense lesions and define “T2 lesions” as a subset of BOD, namely those lesions which are hyperintense on T2 but not hypointense on T1. Thus, BOD is the sum of the T1 lesion and T2 lesion volumes. In addition, we calculated

the total volume of gadolinium-enhancing lesions when available, and the normalized CSF/whole brain volume fraction for each patient using the segmented images.

Quantitative T2 values were estimated on a voxel-by-voxel basis using the two intrinsically co-registered images from the dFSE scan using the expression $T2 = (TE_2 - TE_1) / \ln(S_1/S_2)$, where TE_1 and TE_2 are the two echo times and S_1 and S_2 are their respective signal intensities. The mean T2 values for each tissue, including lesions, were then obtained by applying the relevant tissue masks generated by the segmentation algorithm and averaging the T2 values over all voxels in that compartment. If the estimated T2 value in any GM or WM voxel exceeded +3 standard deviations from the mean compartment value, that voxel was excluded from the analysis as an outlier due to noise and/or partial volume contamination with CSF. This threshold was chosen so that outliers would fall outside the range of 99% of the T2 values for each compartment.

The z-scores were calculated for each measure according to the formula $z = (x - u) / \sigma$, where x is the MRI-derived measure, u the mean and σ is the standard deviation of the measure over the patient population. Henceforth, we employ the notation Z_X to denote the z-score for any given measure, X, defined in Table 1. For example, Z_{BOD} is the z-score associated with BOD, Z_{T2-WM} the z-score associated with T2 of WM, and so on.

Statistical Analysis

All statistical analyses were performed using SAS/STAT® Version 9.3 (Cary, NC). The correlations of each z-score with EDSS, and cross-correlations between each pair of z-scores were computed using Pearson's linear correlation coefficient.

In principle an infinite number of composite measures could be constructed. In order to keep the search for useful composites manageable and informative, the number of components was limited to 2-4 and each component was equally weighted. The correlation results were then used to guide the construction of the composite so that each individual component had high correlation with EDSS (to maximize the sensitivity), and the cross-correlation between components was low (to capture the maximum amount of information with the composite). Once final composites were chosen, Pearson's correlation coefficient as well as regression analysis was used to assess the performance of the composite measures with EDSS.

Results

Segmentation

A typical T2-weighted image and the corresponding segmented images are shown in Figure 1. The map of the quantitative T2 values of WM and GM tissues at the same slice location is also shown for comparison.

Volumetric measures

The correlation coefficients of the volumetric measures with EDSS are listed in Table 1a. The volume of T1 lesions (BH) and burden of disease (BOD) correlated most strongly with EDSS, followed by normalized CSF (nCSF). The volume of gadolinium-enhancing lesions (GD) had a very weak and non-significant correlation ($p = 0.101$).

Cross-correlations between volumetric measures are listed in Table 2. While volumetric measures in general had moderate cross-correlations, the cross-correlation of nCSF with the other measures was substantially lower, demonstrating that it was an independent measure. The correlation between nCSF and BOD just failed to meet significance ($p = 0.054$).

Relaxation measures

The correlation coefficients of the mean T2 values with EDSS are listed in Table 1b. The mean T2 value of T1-lesions (T2-BH) had the strongest correlation with EDSS, followed by the mean T2 of GM (T2-GM). The T2 values for WM (T2-WM) and T2-lesions (T2-LES) were poor and not significant ($p = 0.223$ and 0.161 , respectively).

Histograms of the mean T2 values for each tissue are shown in Figure 2; these showed a relatively narrow distribution for most tissues. In contrast, the histogram for T2-BH was more broadly distributed, suggesting a more heterogeneous microarchitecture of the tissue within this lesion component across patients.

Cross-correlations between the relaxation measures are also listed in Table 2. T2-WM and T2-GM had a moderately high correlation ($r = 0.76$), as did T2-BH and T2-LES ($r = 0.61$). Cross correlations generally were lower between relaxation measures than for their volumetric counterparts.

Composite measures

A number of composite scores were constructed, using the individual measures' correlations with EDSS (Table 1) and the cross-correlations between them (Table 2) as a guide. The final composite scores are defined in Table 3, along with their coefficients of correlation to EDSS and p-values.

The strongest composite scores (as assessed by the correlation with EDSS) were those with a mix of metrics from either volumetry alone or a combination of volumetry and T2 measures. Relaxation-only composite measures had the lowest correlation with EDSS, and underperformed the single measure, T2-BH. The volumetric-only composites consistently outperformed the single measures.

The Z4f and Z3d composites were the strongest performers, with correlation of $r = 0.344$ ($p < 0.0001$) with EDSS. The sole difference between these composites was the inclusion of BOD in Z4f, which had negligible effect on the correlation. A variant of Z3d in which BOD was substituted for BH (Z4e) had slightly lower correlation, suggesting that BOD and BH are of similar effectiveness. The inclusion of T2-BH was critical to performance as its omission in composites Z3e and Z2a compared to Z4f and Z3d demonstrates.

Discussion

The argument for inclusion of T2 can be extended to other measures, including quantitative T1 values [36], diffusion tensor imaging (DTI) [37], magnetization transfer [38,39], spectroscopy [39,40] and lesion analysis in spinal cord [41] and/or gray matter [38,42]. However, each of these measurements requires additional imaging sequences, which leads to increased patient scan times. Therefore, a practical limit exists on how many measures can be included in a composite score within a clinically relevant scan time.

The strength of the correlation of the individual z-scores to EDSS is an important consideration in constructing a robust composite. However, even if all the individual z-score components strongly correlate to EDSS, the overall correlation to EDSS of the composite may still suffer if the component z-scores are not independent, or negatively correlated (which was not the case here, but plausible for more severely disabled patients). For example, the Z4f and Z3d composites differed only by inclusion of BOD, but had near-identical correlation to EDSS ($r = 0.344$, $p < 0.0001$), because of the high cross-correlation of BOD to BH ($r = 0.92$). Hence, adding BOD as a component did not add significantly new information to the composite score, so there was no improvement in the overall correlation to disability. Generally speaking,

inclusion of more measures to the composite does not necessarily improve the correlation, and potentially could weaken it.

In addition to choosing components, another potential avenue for optimization of composite scores is in the component weighting. In our study we used equal weighting for all the MRI-derived measures, but in principle a more optimal weighting could be derived using regression analysis. Such optimization could make the performance of the composite scores more dependent on the patient sample. A thorough analysis of this sort is nontrivial and beyond the scope of the present work, but is the subject of ongoing investigation.

There may also be an inherent limit to any optimization efforts in finding a composite. For one thing, EDSS is an ordinal score, not a continuous variable, and thus will not necessarily represent the complete range of disabilities a patient may have. In addition, cognitive deficit that is known to be a part of the clinical disability is nearly ignored in the EDSS. Furthermore, there are practical (i.e., scan time) limits on the diversity of components we might include, so any characterization of the disease process we might attempt is necessarily incomplete. Therefore, a ceiling for correlation with EDSS is likely. However, we also note that the number of patients with high EDSS scores was limited in our sample, and the extremes tend to drive the strength of the correlation.

In practical terms, even when using optimized composite scores, the correlation with EDSS may well be lower than what we have reported here due to increased patient variability and differences in methodology. The present study had the advantage of a relatively homogenous patient population, imaged at a single research center with rigorous quality assurance procedures during acquisition and analysis. This level of attention may not be feasible in a larger clinical setting or be replicable between imaging centers, and less stringent quality control would likely have a negative impact upon the strength of the correlations. On the other hand, a larger spectrum of patients at all levels of the EDSS might tend to increase the correlation as well as having sufficient sample size to evaluate how therapy impacts these evaluations. To investigate how the composite scores fare under less ideal acquisition and demographic scenarios we are presently collecting suitable data from ongoing multicenter trials for future analysis.

We also caution that the specific composite(s) with best correlation with EDSS may not be constant across all patient groups. Our patient population was exclusively RRMS, but the pathologic substrates among different MS phenotypes might well result in different correlations of z-scores to EDSS and cross-correlations with each other. In general, the process of constructing and optimizing a composite from the available z-scores must be performed anew for each patient cohort.

The published literature has been inconsistent in reporting correlations between EDSS and MRI measures. One possible reason for this inconsistency could be differences in the population size on which the correlations were based. We have demonstrated that the correlation of any measure (composite or otherwise) decreases as the population size increases [43]. Since increasing patient variability and other general confounding factors act to undermine the correlation to disability, the size of the cohort must be taken into consideration when comparing correlations between studies. In our study of 126 patients using composite scores, the highest observed correlation of 0.344 exceeded that of any study of comparable size [44-46]. Hence, composite scores appear to be more robust against the degrading effect of increased population size, which is particularly critical for large studies and multicenter trials.

The two best-performing z-scores were related to T1 lesions, both the volume (BH) and the mean T2 value (T2-BH). This result agrees with published literature, which has generally observed that T1 lesion load strongly correlates with EDSS [11]. However, an editorial by

Naismith and Cross [47] suggests that correlation between T1 lesions and disability in the RRMS subtype alone are poor, though they only cite one study with a large population sample ($r = 0.19$, $n = 239$) [48]. It is possible that some of the inconsistency between studies is due to the same degrading effect of large population size discussed above.

As noted in the Results, ZGD did not have a statistically significant correlation with EDSS. One reason was simply that only a handful of patients exhibited gadolinium-enhancing lesions, so the GD z-score (which relies on the mean and standard deviation of GD across all patients) was inherently compromised, and may not be conceptually consistent with the concept of normalizing the variables into units of standard deviation. For this and performance-related reasons we did not include GD in any of our composites, with the exception of Z4a, which was previously used in the North American Linomide Trial [24]. Further, gadolinium-enhancing lesion load has been correlated with relapse activity [49-51]. In our cross-sectional study, the relevance of gadolinium-enhancement to correlation with EDSS was limited, as perhaps should be expected of a phenomena that while central to the development of new lesions, is quite transient relative to the disease course.

Indeed, the primary limitation of this study was its single-center, cross-sectional nature. Validation of our results requires repeating the analysis on data from multi-center, longitudinal trials, to assess the stability of the correlations derived from larger patient populations and sources of variation, as well as allowing for predictive correlations from one time point into the future. This is the major motivation of our work. We are participating in a number of multi-center clinical trials, including the NIH-sponsored CombiRx study (ClinicalTrials.gov, NCT00211887). Once data collection is complete (estimated July 2012), we will evaluate the performance of composites as outcome measures to validate the correlation with EDSS on the larger data set, initially cross-sectionally and ultimately longitudinally.

Another limitation was our reliance on EDSS. Even though EDSS is widely used in evaluating clinical disability in MS, there is general agreement that EDSS does not have adequate sensitivity, is excessively weighted to ambulation, and suffers from considerable inter-observer and intra-observer variability [52,53], as well as not being a truly ordinal scale. In response to these limitations, the Multiple Sclerosis Functional Composite (MSFC) score was introduced as an alternative to EDSS by the National MS Society Clinical Outcomes Assessment Task Force [54,55]. In the literature, experience with MSFC is encouraging, but limited. The MSFC was not evaluated in our current patient population, but is being evaluated in the ongoing multi-center clinical trial. Once available, we will investigate correlation of our composite measures with MSFC.

Acknowledgments

Funding Sources: NIH R01 EB02095, S10 RR19186, and U01-NS045719

We would like to acknowledge Vipulkumar Patel, RT (MR) for his help with image scanning and protocol optimization. This work was supported by the NIBIB/NIH grant EB02095 and NINDS/NIH grant U01-NS045719. The MRI scanner was partially funded by NCR/NIH grant S10 RR19186; some costs of patient scanning were offset by a gift from the Band Against MS. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIBIB, NINDS or NIH.

References

1. Kurtzke JF. A new scale for evaluating disability in multiple sclerosis. *Neurology* 1955;5(8):580-3. [PubMed: 13244774]
2. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983;33(11):1444-52. [PubMed: 6685237]

3. Miller DH, Thompson AJ. Nuclear magnetic resonance monitoring of treatment and prediction of outcome in multiple sclerosis. *Philos Trans R Soc Lond B Biol Sci* 1999;354(1390):1687–95. [PubMed: 10603620]
4. Wolinsky JS, et al. Linomide in relapsing and secondary progressive MS: part II: MRI results. MRI Analysis Center of the University of Texas-Houston, Health Science Center, and the North American Linomide Investigators. *Neurology* 2000;54(9):1734–41. [PubMed: 10802777]
5. McFarland HF, et al. The role of MRI as a surrogate outcome measure in multiple sclerosis. *Mult Scler* 2002;8(1):40–51. [PubMed: 11936488]
6. Li DK, et al. The use of MRI as an outcome measure in clinical trials. *Adv Neurol* 2006;98:203–26. [PubMed: 16400836]
7. Erickson BJ, Noseworthy JH. Value of magnetic resonance imaging in assessing efficacy in clinical trials of multiple sclerosis therapies. *Mayo Clin Proc* 1997;72(11):1080–9. [PubMed: 9374985]
8. Comi G, Filippi M. Clinical trials in multiple sclerosis: methodological issues. *Curr Opin Neurol* 2005;18(3):245–52. [PubMed: 15891407]
9. Bar-Zohar D, et al. Magnetic resonance imaging metrics and their correlation with clinical outcomes in multiple sclerosis: a review of the literature and future perspectives. *Mult Scler* 2008;14(6):719–27. Epub 2008 Apr 18. [PubMed: 18424478]
10. Zivadinov R. Can imaging techniques measure neuroprotection and remyelination in multiple sclerosis? *Neurology* 2007;68(22 Suppl 3):S72–82. discussion S91-6. [PubMed: 17548573]
11. Zivadinov R, Leist TP. Clinical-magnetic resonance imaging correlations in multiple sclerosis. *J Neuroimaging* 2005;15(4 Suppl):10S–21S. [PubMed: 16385015]
12. Barkhof F. The clinico-radiological paradox in multiple sclerosis revisited. *Curr Opin Neurol* 2002;15(3):239–45. [PubMed: 12045719]
13. Barkhof F. MRI in multiple sclerosis: correlation with expanded disability status scale (EDSS). *Mult Scler* 1999;5(4):283–6. [PubMed: 10467389]
14. Goodin DS. Magnetic resonance imaging as a surrogate outcome measure of disability in multiple sclerosis: have we been overly harsh in our assessment? *Ann Neurol* 2006;59(4):597–605. [PubMed: 16566022]
15. Pulizzi A, et al. Determinants of disability in multiple sclerosis at various disease stages: a multiparametric magnetic resonance study. *Arch Neurol* 2007;64(8):1163–8. [PubMed: 17698707]
16. Benedict RH, et al. Diffusion-weighted imaging predicts cognitive impairment in multiple sclerosis. *Mult Scler* 2007;13(6):722–30. Epub 2007 Mar 15. [PubMed: 17613599]
17. Tavazzi E, et al. Quantitative diffusion weighted imaging measures in patients with multiple sclerosis. *Neuroimage* 2007;36(3):746–54. Epub 2007 Apr 10. [PubMed: 17498974]
18. Sormani MP, et al. MRI metrics as surrogate markers for clinical relapse rate in relapsing-remitting MS patients. *Neurology* 2002;58(3):417–21. [PubMed: 11839841]
19. Rudick RA, et al. Significance of T2 lesions in multiple sclerosis: A 13-year longitudinal study. *Ann Neurol* 2006;60(2):236–42. [PubMed: 16786526]
20. Kalkers NF, et al. Concurrent validity of the MS Functional Composite using MRI as a biological disease marker. *Neurology* 2001;56(2):215–9. [PubMed: 11160958]
21. Kalkers NF, et al. Optimizing the association between disability and biological markers in MS. *Neurology* 2001;57(7):1253–8. [PubMed: 11591845]
22. Neema M, et al. MRI in multiple sclerosis: what's inside the toolbox? *Neurotherapeutics* 2007;4(4):602–17. [PubMed: 17920541]
23. Martinelli Boneschi F, et al. The use of magnetic resonance imaging in multiple sclerosis: lessons learned from clinical trials. *Mult Scler* 2004;10(4):341–7. [PubMed: 15327027]
24. Wolinsky JS, Narayana PA, Johnson KP. United States open-label glatiramer acetate extension trial for relapsing multiple sclerosis: MRI and clinical correlates. Multiple Sclerosis Study Group and the MRI Analysis Center. *Mult Scler* 2001;7(1):33–41. [PubMed: 11321192]
25. Weisstein, EW. z-Score. 2008. cited; Available from: <http://mathworld.wolfram.com/z-Score.html>
26. Wolinsky, JS., et al. American Academy of Neurology. San Diego, USA: 2006. The z4 score appears to have a predictive correlation as well that suggests that it might have the potential as a surrogate outcome measure.

27. Riahi F, et al. Improved correlation between scores on the expanded disability status scale and cerebral lesion load in relapsing-remitting multiple sclerosis. Results of the application of new imaging methods. *Brain* 1998;121(Pt 7):1305–12. [PubMed: 9679782]
28. Duncan JS, Bartlett P, Barker GJ. Technique for measuring hippocampal T2 relaxation time. *AJNR Am J Neuroradiol* 1996;17(10):1805–10. [PubMed: 8933861]
29. Hoque R, et al. The role of quantitative neuroimaging indices in the differentiation of ischemia from demyelination: an analytical study with case presentation. *Int Rev Neurobiol* 2007;79:491–519. [PubMed: 17531856]
30. Neema M, et al. T1- and T2-based MRI measures of diffuse gray matter and white matter damage in patients with multiple sclerosis. *J Neuroimaging* 2007;17(Suppl 1):16S–21S. [PubMed: 17425729]
31. Barbosa S, et al. Magnetic resonance relaxation time mapping in multiple sclerosis: normal appearing white matter and the “invisible” lesion load. *Magn Reson Imaging* 1994;12(1):33–42. [PubMed: 8295506]
32. Stevenson VL, et al. Variations in T1 and T2 relaxation times of normal appearing white matter and lesions in multiple sclerosis. *J Neurol Sci* 2000;178(2):81–7. [PubMed: 11018698]
33. WTCf Neuroimaging. Statistical Parametric Mapping (SPM2). London: 2005.
34. Sajja BR, et al. Unified approach for multiple sclerosis lesion segmentation on brain MRI. *Ann Biomed Eng* 2006;34(1):142–51. [PubMed: 16525763]
35. Datta S, et al. Segmentation and quantification of black holes in multiple sclerosis. *Neuroimage* 2006;29(2):467–74. [PubMed: 16126416]
36. Parry A, et al. White matter and lesion T1 relaxation times increase in parallel and correlate with disability in multiple sclerosis. *J Neurol* 2002;249(9):1279–86. [PubMed: 12242554]
37. Hasan KM, et al. Diffusion tensor fractional anisotropy of the normal-appearing seven segments of the corpus callosum in healthy adults and relapsing-remitting multiple sclerosis patients. *J Magn Reson Imaging* 2005;21(6):735–43. [PubMed: 15906348]
38. Vrenken H, et al. Magnetization transfer ratio measurement in multiple sclerosis normal-appearing brain tissue: limited differences with controls but relationships with clinical and MR measures of disease. *Mult Scler* 2007;13(6):708–16. Epub 2007 Mar 15. [PubMed: 17613597]
39. Bellmann-Strobl J, et al. MR spectroscopy (MRS) and magnetisation transfer imaging (MTI), lesion load and clinical scores in early relapsing remitting multiple sclerosis: a combined cross-sectional and longitudinal study. *Eur Radiol* 2009;19(8):2066–74. Epub 2009 Mar 24. [PubMed: 19308417]
40. Narayana PA. Magnetic resonance spectroscopy in the monitoring of multiple sclerosis. *J Neuroimaging* 2005;15(4 Suppl):46S–57S. [PubMed: 16385018]
41. Rashid W, et al. Increasing cord atrophy in early relapsing-remitting multiple sclerosis: a 3 year study. *J Neurol Neurosurg Psychiatry* 2006;77(1):51–5. [PubMed: 16361592]
42. Nelson F, et al. Improved identification of intracortical lesions in multiple sclerosis with phase-sensitive inversion recovery in combination with fast double inversion recovery MR imaging. *AJNR Am J Neuroradiol* 2007;28(9):1645–9. Epub 2007 Sep 20. [PubMed: 17885241]
43. Poonawalla A, et al. Inverse dependence between patient population and correlation of composite MRI scores with EDSS in Multiple Sclerosis. *Proc Intl Soc Magn Reson Med* 2009;17
44. Fisher E, et al. Eight-year follow-up study of brain atrophy in patients with MS. *Neurology* 2002;59(9):1412–20. [PubMed: 12427893]
45. Kappos L, et al. Predictive value of gadolinium-enhanced magnetic resonance imaging for relapse rate and changes in disability or impairment in multiple sclerosis: a metaanalysis. *Gadolinium MRI Meta-analysis Group. Lancet* 1999;353(9157):964–9. [PubMed: 10459905]
46. Paty DW, Li DK. Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. II. MRI analysis results of a multicenter, randomized, double-blind, placebo-controlled trial. *UBC MS/MRI Study Group and the IFNB Multiple Sclerosis Study Group. Neurology* 1993;43(4):662–7. [PubMed: 8469319]
47. Naismith RT, Cross AH. Multiple sclerosis and black holes: connecting the pixels. *Arch Neurol* 2005;62(11):1666–8. [PubMed: 16286537]
48. Rovaris M, et al. Short-term correlations between clinical and MR imaging findings in relapsing-remitting multiple sclerosis. *AJNR Am J Neuroradiol* 2003;24(1):75–81. [PubMed: 12533330]

49. Petkau J, et al. Magnetic resonance imaging as a surrogate outcome for multiple sclerosis relapses. *Mult Scler* 2008;14(6):770–8. Epub 2008 Jun 5. [PubMed: 18535021]
50. Sormani MP, et al. Magnetic resonance imaging as a potential surrogate for relapses in multiple sclerosis: a meta-analytic approach. *Ann Neurol* 2009;65(3):268–75. [PubMed: 19334061]
51. Daumer M, et al. MRI as an outcome in multiple sclerosis clinical trials. *Neurology* 2009;72(8):705–11. Epub 2008 Dec 10. [PubMed: 19073945]
52. Noseworthy JH, et al. Interrater variability with the Expanded Disability Status Scale (EDSS) and Functional Systems (FS) in a multiple sclerosis clinical trial. The Canadian Cooperation MS Study Group. *Neurology* 1990;40(6):971–5. [PubMed: 2189084]
53. Goodkin DE, et al. Inter- and intrarater scoring agreement using grades 1.0 to 3.5 of the Kurtzke Expanded Disability Status Scale (EDSS). Multiple Sclerosis Collaborative Research Group. *Neurology* 1992;42(4):859–63. [PubMed: 1565242]
54. Cutter GR, et al. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain* 1999;122(Pt 5):871–82. [PubMed: 10355672]
55. Fischer JS, et al. The Multiple Sclerosis Functional Composite Measure (MSFC): an integrated approach to MS clinical outcome assessment. National MS Society Clinical Outcomes Assessment Task Force. *Mult Scler* 1999;5(4):244–50. [PubMed: 10467383]

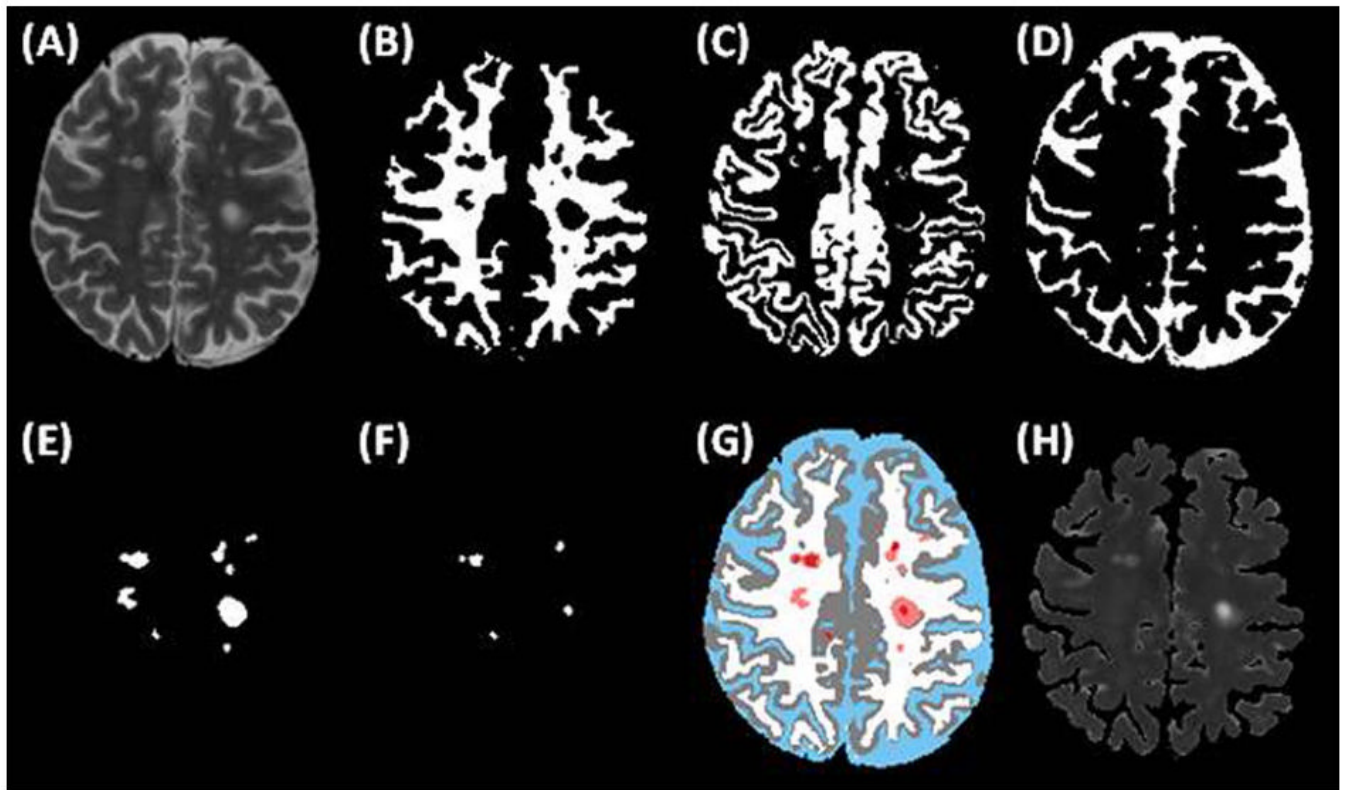


Figure 1.

Segmentation algorithm output for a representative patient anatomic slice: (a) normal T2-weighted image, (b) WM compartment, (c) GM compartment, (d) CSF compartment, (e) T2 lesions, (f) T1 lesions, and (g) a colorized composite map of all compartments, with white for WM, gray for GM, blue for CSF, pink for T2-lesions, and red for T1-lesions. The corresponding T2 map (excluding CSF) is also shown in (h).

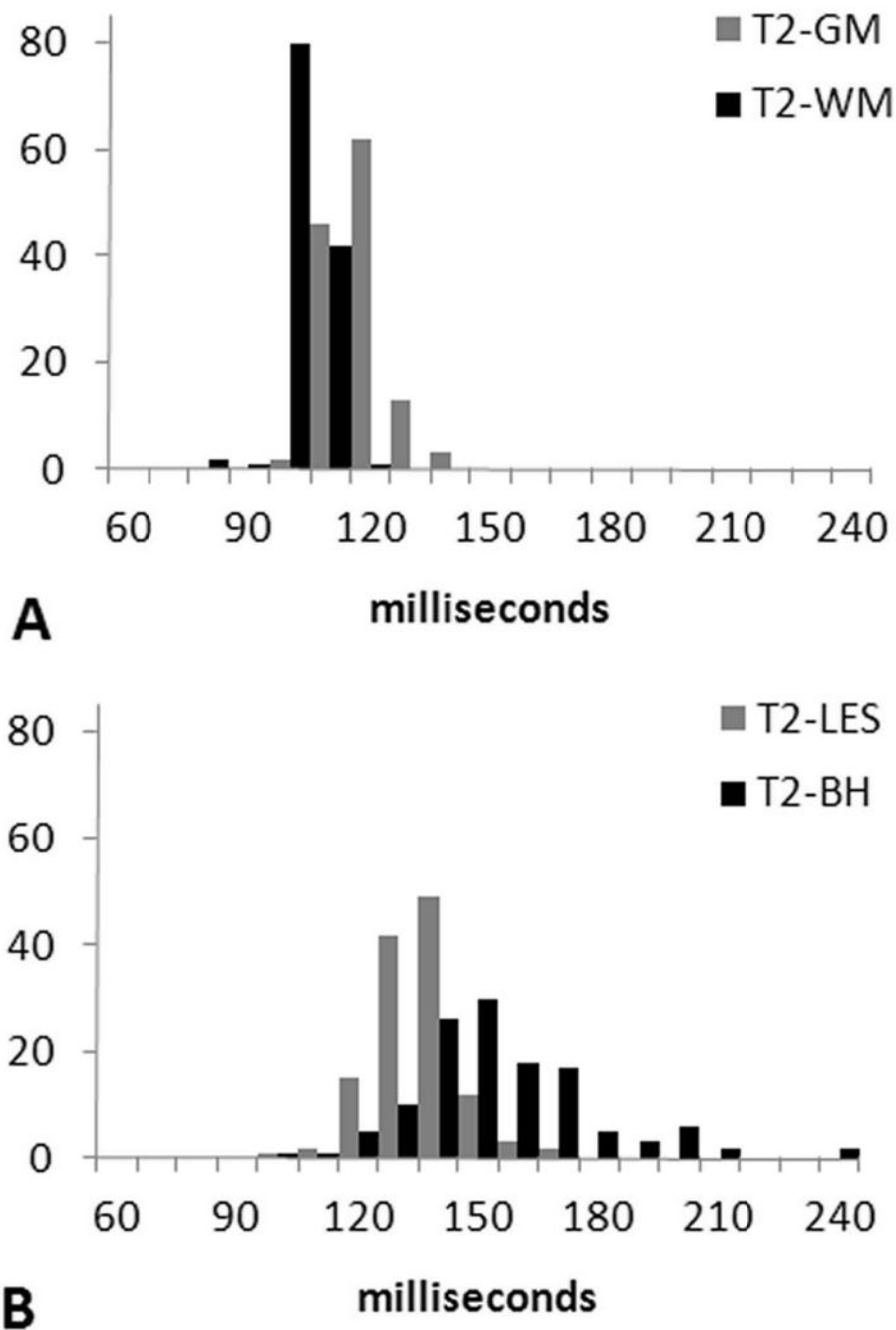


Figure 2. Histograms of mean T2 values for (a) WM and GM, and (b) T2-lesion, and T1-lesion compartments. Note that the histograms for WM, GM and T2-lesions are fairly narrow, whereas the histogram for T1-lesions (i.e., black holes) shows more broadening, suggesting a more heterogeneous microarchitecture of the tissue within this lesion component across patients. The mean values of T2 are 98.3 ± 5.0 ms for WM, 112.2 ± 6.4 ms for GM, 130.4 ± 10.6 ms for T2-lesions, and 150 ± 23.1 ms for black holes.

Table 1

Table 1a: Definition of volumetric MRI measures and their correlation with EDSS			
Measure	Definition	rho	p
GD	Gadolinium-enhancing lesion volume (cc)	0.147	0.101
BOD	“Burden of Disease” T2-hyperintense lesion volume (including T1-lesion component) (cc)	0.276	< 0.002
BH	“Black Hole” T1-hypointense lesion volume (cc)	0.298	< 0.001
nCSF	Ratio of CSF volume to total intracranial volume	0.221	0.013

Table 1b. Definition of relaxation MRI measures and their correlation with EDSS			
Measure	Definition	rho	p
T2-GM	Mean T2 value of gray matter (ms)	0.188	0.035
T2-WM	Mean T2 value of white matter (ms)	0.109	0.224
T2-LES	Mean T2 value of T2-hyperintense lesions (excluding T1-lesion component) (ms)	0.126	0.161
T2-BH	Mean T2 value of T1-lesions (ms)	0.294	< 0.001

Table 2
Pearson correlation coefficients between pairs of individual z-scores

	Z _{GD}	Z _{BOD}	Z _{BH}	Z _{nCSF}	Z _{T2-GM}	Z _{T2-WM}	Z _{T2-LES}	Z _{T2-BH}
Z _{GD}	1.000	0.669	0.623	-0.294	0.099 [†]	0.177	0.285	0.232
Z _{BOD}	0.669	1.000	0.917	0.172 [†]	0.443	0.473	0.369	0.611
Z _{BH}	0.623	0.917	1.000	0.189	0.477	0.434	0.488	0.707
Z _{nCSF}	-0.294	0.172 [†]	0.189	1.000	0.592	0.271	0.212	0.403
Z _{T2-GM}	0.099 [†]	0.443	0.477	0.592	1.000	0.758	0.458	0.534
Z _{T2-WM}	0.177	0.473	0.434	0.271	0.758	1.000	0.463	0.414
Z _{T2-LES}	0.285	0.369	0.488	0.212	0.458	0.463	1.000	0.611
Z _{T2-BH}	0.232	0.611	0.707	0.403	0.534	0.414	0.611	1.000

[†] correlation failed to meet statistical significance (p > 0.05).

Table 3
Definition of various composite z-scores and correlation to EDSS

Score	Definition	rho	p
Z4f	$(Z_{\text{BOD}} + Z_{\text{BH}} + Z_{\text{nCSF}}) + Z_{\text{T2BH}}$	0.344	< 0.0001
Z3d	$(Z_{\text{BH}} + Z_{\text{nCSF}}) + (Z_{\text{T2BH}})$	0.344	< 0.0001
Z3c	$(Z_{\text{BOD}} + Z_{\text{nCSF}}) + (Z_{\text{T2BH}})$	0.341	< 0.0001
Z3e	$Z_{\text{BOD}} + Z_{\text{BH}} + Z_{\text{nCSF}}$ (volumetry)	0.337	0.0001
Z2a	$Z_{\text{BOD}} + Z_{\text{nCSF}}$ (volumetry)	0.325	< 0.001
Z4a [†]	$Z_{\text{GD}} + Z_{\text{BOD}} + Z_{\text{BH}} + Z_{\text{nCSF}}$ (volumetry)	0.322	< 0.001
Z4c	$Z_{\text{T2GM}} + Z_{\text{T2WM}} + Z_{\text{T2LES}} + Z_{\text{T2BH}}$ (relaxation)	0.221	<0.05

[†]Used in the North American Linomide Trial, see reference [24].