



Published in final edited form as:

JAMA Neurol. 2013 March 1; 70(3): 338–344. doi:10.1001/2013.jamaneuro.211.

Reliability of Classifying Multiple Sclerosis Disease Activity Using Magnetic Resonance Imaging in a Multiple Sclerosis Clinic

Ebru Erbayat Altay, MD, Elizabeth Fisher, PhD, Stephen E. Jones, MD, PhD, Claire Hara-Cleaver, MSN, Jar-Chi Lee, MS, and Richard A. Rudick, MD

Mellen Center, Neurological Institute, (Drs Erbayat Altay and Rudick and Ms Hara-Cleaver), Neuroradiology, Imaging Institute (Dr Jones), and Departments of Biomedical Engineering (Dr Fisher) and Quantitative Health Sciences (Ms Lee), Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio.

Abstract

Objective—To assess the reliability of new magnetic resonance imaging (MRI) lesion counts by clinicians in a multiple sclerosis specialty clinic.

Design—An observational study.

Setting—A multiple sclerosis specialty clinic.

Patients—Eighty-five patients with multiple sclerosis participating in a National Institutes of Health–supported longitudinal study were included.

Intervention—Each patient had a brain MRI scan at entry and 6 months later using a standardized protocol.

Main Outcome Measures—The number of new T2 lesions, newly enlarging T2 lesions, and gadolinium-enhancing lesions were measured on the 6-month MRI using a computer-based image analysis program for the original study. For this study, images were reanalyzed by an expert neuroradiologist and 3 clinician raters. The neuroradiologist evaluated the original image pairs; the clinicians evaluated image pairs that were modified to simulate clinical practice. New lesion counts were compared across raters, as was classification of patients as MRI active or inactive.

Results—Agreement on lesion counts was highest for gadolinium-enhancing lesions, intermediate for new T2 lesions, and poor for enlarging T2 lesions. In 18% to 25% of the cases, MRI activity was classified differently by the clinician raters compared with the neuroradiologist or computer program. Variability among the clinical raters for estimates of new T2 lesions was

© 2013 American Medical Association. All rights reserved.

Correspondence: Richard A. Rudick, MD, Cleveland Clinic Mellen Center for Multiple Sclerosis Treatment and Research, 9500 Euclid Ave, JJ36, Cleveland, OH 44195-5244 (rudickr@ccf.org).

Author Contributions: *Study concept and design:* Erbayat Altay, Fisher, Jones, Hara-Cleaver, Lee, and Rudick. *Acquisition of data:* Erbayat Altay, Fisher, Hara-Cleaver, and Rudick. *Analysis and interpretation of data:* Erbayat Altay, Fisher, Jones, Lee, and Rudick. *Drafting of the manuscript:* Erbayat Altay, Lee, and Rudick. *Critical revision of the manuscript for important intellectual content:* Erbayat Altay, Fisher, Jones, Hara-Cleaver, Lee, and Rudick. *Statistical analysis:* Lee and Rudick. *Obtained funding:* Rudick. *Administrative, technical, and material support:* Fisher, Jones, and Hara-Cleaver. *Study supervision:* Fisher.

Conflict of Interest Disclosures: Dr Fisher received consulting fees and research funds from Biogen Idec and Genzyme. Dr Rudick received consulting fees or honoraria from Biogen Idec, Novartis, Genzyme-Bayer, Wyeth, and Pfizer and royalties from Informa Healthcare. Ms Hara-Cleaver has received personal compensation as a consultant or speaker from Biogen Idec, Novartis, Teva, Serono, and Acorda.

Online-Only Material: The eAppendix is available at <http://www.jamaneuro.com>.

affected most strongly by the image modifications that simulated low image quality and different head position.

Conclusions—Between-rater variability in new T2 lesion counts may be reduced by improved standardization of image acquisitions, but this approach may not be practical in most clinical environments. Ultimately, more reliable, robust, and accessible image analysis methods are needed for accurate multiple sclerosis disease-modifying drug monitoring and decision making in the routine clinic setting.

Magnetic resonance imaging (MRI) has a high sensitivity for focal white matter lesions that characterize multiple sclerosis (MS). Magnetic resonance imaging lesion assessment has taken a central role in diagnosis,^{1,2} prognosis,^{3–5} and drug development.^{6,7} In fact, treatment effect on lesion activity has been used to screen all currently approved MS disease-modifying drugs. Treatment effect on MRI lesions parallels treatment effects on relapses and worsening disability, as measured by the Kurtzke Expanded Disability Status Scale.^{7–9}

It follows logically that MRI may be more sensitive than clinical events as a disease activity measure in clinical practice. Longitudinal studies have provided evidence to support MRI to monitor the effectiveness of disease-modifying drugs. However, there are no standardized, validated methods to quantify lesions in the clinical setting. Comparison of serial MRI studies is generally done by neuroradiologists or neurologists using visual inspection to compare serial images. There is a significant literature on interrater reliability for lesion detection in the context of controlled clinical trials,^{10–13} where numerous investigators have emphasized the need to use the same MRI scanner and image acquisition protocol for individual patients, the importance of patient positioning, the use of validated image analysis software, and the value of training the observers when lesions are quantified by visual inspection. Such conditions rarely exist in clinical practice.

Furthermore, to our knowledge, no assessment of interrater variability of lesion counts under typical clinical conditions has been reported. Therefore, the reliability of using serial MRI in a practice setting is unknown. Since treatment decisions are being made by MS specialty clinicians based on MRI, it seemed important to determine the reliability of lesion assessment in the clinical setting.

METHODS

SUBJECTS

Eighty-five patients who had started taking intramuscular interferon beta-1a as their first disease-modifying drug were entered into an observational study to identify biological correlates of treatment response to interferon beta-1a.¹⁴ The patients were followed up between 2005 and 2009 at Mellen Center for Multiple Sclerosis Treatment and Research, Cleveland Clinic. As part of the study, treatment response was categorized as “good” or “poor” based on the number of new or enlarging brain MRI lesions on the month 6 MRI compared with month 0, when the patient started taking interferon beta. The observational study was supported by the National Institutes of Health.

MAGNETIC RESONANCE IMAGING

Image acquisition was standardized and all MRI studies were done on a 1.5-T Siemens Symphony scanner. Images were acquired axially and consisted of T2-weighted fluid-attenuated inversion recovery; proton-density/T2-weighted, dual-echo spin echo images; and T1-weighted spin echo images before and after injection of a standard dose (0.1 mmol/kg of body weight) of gadolinium-diethylenetriaminepentaacetate. All images had a voxel size of

0.9 mm × 0.9 mm × 3 mm except the fluid-attenuated inversion recovery image, which had a voxel size of 0.9 mm × 0.9 mm × 5 mm.

IMAGE ANALYSIS AND REVIEW

For the National Institutes of Health longitudinal study, the number of new T2 lesions and enlarging T2 lesions at 6 months were determined using a semiautomated image analysis program that determines new or enlarging lesions based on coregistration and subtraction of the baseline proton density–weighted image from the 6-month proton density–weighted image. Enhancing lesions at 6 months were also determined using a semiautomated segmentation program. For this study, a neuroradiologist (S.E.J.) examined the same images acquired for the National Institutes of Health study, under optimal conditions. Specifically, image pairs (baseline and month 6) were compared side by side to determine the presence of new or enlarging lesions.

To assess reliability of clinician ratings in a clinic setting, 3 raters with different levels of experience—a neurologist (R.A.R.), a neurology resident (E.E.A.), and a nurse practitioner (C.H.C.)—quantified new lesions by visually comparing the month 6 scan with baseline. Clinician raters viewed images that were modified to simulate clinical practice. For each pair of scans, 1 time point was randomly assigned to 1 of 5 modification groups: (1) images left in their original form, viewable on a picture archiving and communication system: “not modified” (n=16); (2) images not modified, viewable only on hard copy film: “filmed” (n=23); (3) images modified with random noise to reduce the signal to noise ratio and simulate poor-quality images: “low quality” (n=11); (4) images resampled with a different slice thickness to simulate changes in acquisition protocol: “slice thickness” (n=16); or (5) images rotated out of plane to simulate differences in patient position: “rotated” (n=19). Examples of image modifications are shown in Figure 1. To further simulate image analysis in a busy clinical practice, clinical raters were limited to no more than 10 minutes per study to assess lesions.

ANALYSIS

Each rater counted the number of new T2 lesions, newly enlarging T2 lesions, and the number of enhancing lesions. Agreement between raters on counts was compared using Lin concordance correlations. Magnetic resonance imaging disease activity (MRI active) was defined as 3 or more new/enlarging lesions, and patients were individually classified as MRI active or MRI stable. Agreement on classification of the 85 cases was determined using statistics.

To determine which of the image modifications contributed most to variability, variability in lesion counts between clinical raters viewing images was analyzed for each type of image modification described earlier. The set of images that was not modified provided an opportunity to compare variability in images modified to simulate clinical practice with images collected under optimal conditions. Three different methods were used to analyze this question (details are provided in the eAppendix, <http://www.jamaneuro.com>). For each method, variance across the 3 clinical raters for the modified images was compared with variance across the 3 clinical raters for the unaltered images. Method 1 computed the mean variance across the 3 clinical raters. Method 2 computed the mean sum of squared differences between the 3 clinical raters and the neuroradiologist. Method 3 computed the mean sum of squared differences between the 3 clinical raters and the computer-generated lesion counts. Mean variance or sum of squared differences was computed for each of the imaging parameters (eg, new T2 lesions) and for each of the image modification groups (eg, rotated). Variance or sum of squared differences from the neuroradiologist or computer counts were compared between the modified images and the unaltered images. Analysis was

conducted separately for gadolinium-enhancing lesions, new T2 lesions, and enlarging T2 lesions. For each comparison, mean differences observed for the altered images were compared with the unaltered images using the *t* test.

RESULTS

Characteristics of the 85 patients are provided in Table 1. Twenty-seven patients had clinically isolated syndrome and 58 had relapsing-remitting MS. Fifteen patients (18%) were classified in the original protocol as being MRI active at 6 months (3 new or enlarging lesions measured using image analysis software). Characteristics of the MRI active and MRI inactive cases were generally similar at baseline, with the exception of more enhancing lesions and higher T2 lesion volume in the MRI active patients.

Table 2 shows results from new T2 lesion counts and Table 3 shows results from enhancing lesion counts on the month 6 MRI as determined by the neuroradiologist, clinician raters, and the computer program. Lin concordance correlation coefficients for enlarging T2 lesions ranged from 0.00 to 0.14, indicating a very high level of variability in counting enlarging T2 lesions (data not shown).

There was better agreement for new T2 lesions. Lin concordance correlations for new T2 lesions ranged from 0.60 to 0.70. The first column in Table 2 shows results generated by the neuroradiologist, the next 3 columns show numbers generated by the clinician raters, and the fourth column shows results using the computer program. Compared with the neuroradiologist, the neurologist's counts were identical in 60% of the cases, and the neurology resident's and nurse practitioner's counts were identical in 53%. Counts generated by the computer program agreed with the neuroradiologist in 71% of the cases.

Reproducibility was best for enhancing lesions. Lin concordance correlation coefficients for enhancing lesions ranged from 0.80 to 0.96. The number of enhancing lesions on the 6-month MRI is shown in Table 3. Compared with the neuroradiologist, the neurologist enhancing lesion counts were identical in 82% of the cases; the neurology resident's and nurse practitioner's enhancing lesion counts were identical in 86% of the cases. Lesion counts generated by the computer program agreed with the neuroradiologist in 84% of the cases.

By defining MRI active scans as 3 or more new or enlarging T2 lesions or gadolinium-enhancing lesions at 6 months, we were able to determine consistency of classifying the patients across clinical raters. Table 4 shows the number of cases in which the patient was classified the same by each clinician rater, the neuroradiologist, and the computer program. Compared with the neuroradiologist, classification by the neurologist was the same in 82%; classification by the neurology resident was the same in 75%; classification by the nurse practitioner was the same in 81%; and classification by the computer program was the same in 88% of the cases. Thus, misclassification by clinician raters compared with the neuroradiologist ranged from 18% to 25%.

Figure 2 shows the proportion of cases where new T2 or enhancing lesion counts differed by the clinical raters or computer program compared with the neuroradiologist and the proportion of misclassified cases. Discrepant results for T2 lesion number were most common for the neurology resident and nurse practitioner (47%) and least common for the computer program (28%). Discrepant results for enhancing lesions were lower for all raters and lowest for the neurology resident and nurse practitioner (12%). Discrepancy for MRI activity category was highest for the neurology resident (25%) and lowest for the computer program (12%).

Detailed results for analyses exploring which image modifications were associated with the most variability across raters are reported in the eAppendix. All of the image modifications resulted in more variability across raters than was observed for unaltered image pairs. Compared with the unaltered images, increased variability with altered images was not statistically significant, although there were consistent trends indicating that low image quality and image alignment differences had the highest impact on variability.

COMMENT

It is widely agreed that MRI is more sensitive as a disease activity marker than clinical assessment. This is commonly observed in clinical practice, where patients with MS often have new MRI lesions without new clinical symptoms. The sensitivity of MRI for the MS disease process is the basis for making a diagnosis of MS earlier than would be possible using relapses.¹ Similarly, analysis of lesion activity on serial MRI scans is widely used as a measure of treatment effects. Enhancing lesions indicate foci of active inflammation, while T2 lesions persist and accumulate; therefore, the presence of enhancing lesions indicates currently active disease, while the overall T2 disease burden provides a marker for cumulative disease progression. Early in the disease, T2 disease burden and accumulation of T2 lesions predict future disease severity.^{15–17} Therefore, MRI scanning in patients with MS is used for diagnosis, counseling patients about disease severity and prognosis, and deciding the need for disease-modifying drug therapy and as a tool to screen and test new therapies.

While the variability in lesion quantitation as a clinical trial metric has been well studied and its importance emphasized in the literature,^{10–12,18} variability in clinical practice has been less well studied but is increasingly important. The practice of monitoring the effects of MS drugs in clinical practice is supported by independent studies demonstrating a poor outcome in patients with MS with new MRI lesions despite treatment with interferon beta.^{19–22} One of the first studies to demonstrate this relationship was a post hoc analysis of the phase 3 study of intramuscular interferon beta-1a. In that study, patients were classified as being responders or nonresponders based on new T2 lesions or enhancing lesions during the 2-year clinical trial.¹⁹ Intramuscular interferon beta-1a recipients who were classified as nonresponders based on the presence of 3 or more new T2 lesions (the median of placebo-treated patients) during 2 years of receiving treatment had significantly more worsening on the Expanded Disability Status Scale and Multiple Sclerosis Functional Composite and more brain atrophy compared with patients with 2 or fewer new lesions. Other observational studies also suggested that patients with new T2 lesions or enhancing lesions developing while taking interferon beta had poor outcomes.^{20,21,23} These studies suggested that the occurrence of MRI lesions in patients receiving disease-modifying drug therapy, particularly interferon beta, could be used to supplement clinical assessment while monitoring patients receiving disease-modifying drug therapy for breakthrough disease and possibly allow earlier treatment decisions.

Recently, the concept of disease activity–free status has been advanced as a therapeutic target in MS. The definition of disease-free status includes absence of new T2 lesions, enhancing lesions, relapses, or Expanded Disability Status Scale score worsening. Treatment with natalizumab increased the proportion of patients with disease activity–free status in the AFFIRM study,²⁴ oral cladribine increased the proportion in the CLARITY study,^{25,26} and fingolimod increased the proportion in the FREEDOMS study.²⁷ So far, the concept of disease-free status has been restricted to analysis of clinical trial results. If the same definition were to be used as an outcome measure in clinical practice, then lesion analysis will be increasingly used to monitor the effects of therapy in clinical practice.

Assuming that MRI disease activity will be increasingly used to decide whether to continue or change therapy, the reliability of quantifying MRI disease activity in a practice setting will be increasingly important. Magnetic resonance imaging monitoring in the clinical setting typically involves comparing image pairs under suboptimal conditions. For example, patients may have follow-up images acquired at different imaging centers, the image acquisition protocols likely differ, and serial scans may be stored on different media. In addition, the neurologist caring for the patient may compare serial images in the middle of a hectic patient schedule. Under these conditions, the accuracy of new lesion assessment may be compromised.

This study was designed to simulate these conditions to determine the reliability of counting lesions and classifying MRI disease activity in the clinical setting. Consistent with results observed by Molyneux and colleagues,¹³ we found strong agreement between clinical raters and the neuroradiologist for enhancing lesions, moderate agreement for new T2 lesions, and extremely poor agreement for enlarging T2 lesions. Even for new T2 lesion and enhancing lesion counts, however, results between the clinical raters and the neuroradiologist were frequently discrepant (Table 2 and Table 3; Figure 2). Defining MRI active scans as 3 or more active lesions on the 6-month follow-up scan, 18% to 25% of the patients were classified differently by the clinical raters compared with the neuroradiologist. This not only documents variability between clinical raters, but also suggests that a significant number of patients might be misclassified. This would contribute to variable and suboptimal treatment decisions, which could degrade outcomes.

Though not the primary purpose of this study, we also noted discrepancy between the computer program and the expert neuroradiologist. In particular, treatment response classification was discrepant in 12% of the cases. We have not yet conducted studies to determine the explanation for the discrepancy.

In exploratory studies, we analyzed which of the image modifications had the maximum impact on variability across the clinical raters. Not surprisingly, when images were modified to simulate low image quality or significant differences in head position in the paired studies, variability across clinical raters increased. These results suggest that special attention should be paid to achieve reproducible head positioning and acquire images with high image quality.

This study has a number of limitations and caveats. First, in an attempt to simulate clinical practice, the clinicians evaluated the MRI scans under contrived conditions. While an attempt was made to simulate clinical practice by manipulating the images, it is not clear how closely the images resembled what would normally be encountered in a practice setting. Variability between the clinician raters and concordance with neuroradiologist ratings would likely improve with standardized image acquisition conditions, as recommended by the Consortium of Multiple Sclerosis Centers,²⁸ or with intensive training, as reported by Barkhof and colleagues.¹⁰ Second, the neuroradiologist ratings were done under ideal conditions, not likely to be replicated in clinical practice, where prior studies may not be readily available or where different machines or image acquisition parameters were used in the paired image sets. Also, cross-training for new lesion detection is not common in radiology practice. Third, about one-third of the cases had clinically isolated syndrome. Many of these cases had minimal T2 lesion burden. Lesion counting is much easier with low T2 lesion burden compared with more severe disease. Therefore, the composition of this cohort may have led to an underestimation of the variability between raters and the neuroradiologist. Lastly, the most important limitation of this study is that there is no absolute gold standard for new lesions, the outcome of interest in this study. Careful inspection of the 10 cases classified differently by the neuroradiologist and computer

program might be informative. If it were determined that the computer program was accurate in quantifying new lesions, application of computer software to quantify new lesions in a practice setting could lead to more reliable patient monitoring and lower the burden on neuroradiologists.

Despite these limitations, the study documents a significant level of variability across clinical raters in counting new lesions in a simulated clinic setting and significant discrepancy when compared with an expert neuroradiologist evaluating paired images under optimal conditions. Data from controlled clinical trials and rigorous longitudinal studies document the value of MRI in monitoring treatment effects and identifying breakthrough disease, which suggests that MRI should be helpful in making treatment decisions in a practice setting. For consistent treatment decisions, however, practical, reproducible methods will be required to optimize patient care and allow MRI to reach its potential as a clinical monitoring tool for personalized use of disease-modifying drugs in MS.

Acknowledgments

Funding/Support: This study was supported by National Institutes of Health grants P50NS38667 and UL1 RR024989. Drs Fisher and Rudick and Ms Lee received support (paid to institution) for this project via National Institutes of Health–National Institute of Neurological Disorders and Stroke grant P50 NS 38667 (2004–2009), Project 4 P01NS38667.

REFERENCES

1. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann Neurol*. 2011; 69(2):292–302. [PubMed: 21387374]
2. Sicotte NL. Magnetic resonance imaging in multiple sclerosis: the role of conventional imaging. *Neurol Clin*. 2011; 29(2):343–356. [PubMed: 21439445]
3. Barkhof F, Filippi M, Miller DH, et al. Comparison of MRI criteria at first presentation to predict conversion to clinically definite multiple sclerosis. *Brain*. 1997; 120(pt 11):2059–2069. [PubMed: 9397021]
4. O’Riordan JI, Thompson AJ, Kingsley DP, et al. The prognostic value of brain MRI in clinically isolated syndromes of the CNS: a 10-year follow-up. *Brain*. 1998; 121(pt 3):495–503. [PubMed: 9549525]
5. Fisher E, Rudick RA, Simon JH, et al. Eight-year follow-up study of brain atrophy in patients with MS. *Neurology*. 2002; 59(9):1412–1420. [PubMed: 12427893]
6. Miller DH. Biomarkers and surrogate outcomes in neurodegenerative disease: lessons from multiple sclerosis. *NeuroRx*. 2004; 1(2):284–294. [PubMed: 15717029]
7. Sormani MP, Bonzano L, Roccatagliata L, Cutter GR, Mancardi GL, Bruzzi P. Magnetic resonance imaging as a potential surrogate for relapses in multiple sclerosis: a meta-analytic approach. *Ann Neurol*. 2009; 65(3):268–275. [PubMed: 19334061]
8. Sormani MP, Bonzano L, Roccatagliata L, Mancardi GL, Uccelli A, Bruzzi P. Surrogate endpoints for EDSS worsening in multiple sclerosis: a meta-analytic approach. *Neurology*. 2010; 75(4):302–309. [PubMed: 20574036]
9. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*. 1983; 33(11):1444–1452. [PubMed: 6685237]
10. Barkhof F, Filippi M, van Waesberghe JH, et al. Improving interobserver variation in reporting gadolinium-enhanced MRI lesions in multiple sclerosis. *Neurology*. 1997; 49(6):1682–1688. [PubMed: 9409366]
11. Filippi M, van Waesberghe JH, Horsfield MA, et al. Interscanner variation in brain MRI lesion load measurements in MS: implications for clinical trials. *Neurology*. 1997; 49(2):371–377. [PubMed: 9270564]

12. Filippi M, Horsfield MA, Rovaris M, et al. Intraobserver and interobserver variability in schemes for estimating volume of brain lesions on MR images in multiple sclerosis. *AJNR Am J Neuroradiol*. 1998; 19(2):239–244. [PubMed: 9504472]
13. Molyneux PD, Miller DH, Filippi M, et al. Visual analysis of serial T2-weighted MRI in multiple sclerosis: intra- and interobserver reproducibility. *Neuroradiology*. 1999; 41(12):882–888. [PubMed: 10639661]
14. Rudick RA, Rani MR, Xu Y, et al. Excessive biologic response to IFN β is associated with poor treatment response in patients with multiple sclerosis. *PLoS One*. 2011; 6(5):e19262. [PubMed: 21602934]
15. Filippi M, Horsfield MA, Tofts PS, Barkhof F, Thompson AJ, Miller DH. Quantitative assessment of MRI lesion load in monitoring the evolution of multiple sclerosis. *Brain*. 1995; 118(pt 6):1601–1612. [PubMed: 8595489]
16. Filippi M, Paty DW, Kappos L, et al. Correlations between changes in disability and T2-weighted brain MRI activity in multiple sclerosis: a follow-up study. *Neurology*. 1995; 45(2):255–260. [PubMed: 7854522]
17. Brex PA, Ciccarelli O, O’Riordan JI, Sailer M, Thompson AJ, Miller DH. A longitudinal study of abnormalities on MRI and disability from multiple sclerosis. *N Engl J Med*. 2002; 346(3):158–164. [PubMed: 11796849]
18. Filippi M, Barkhof F, Bressi S, Yousry TA, Miller DH. Inter-rater variability in reporting enhancing lesions present on standard and triple dose gadolinium scans of patients with multiple sclerosis. *Mult Scler*. 1997; 3(4):226–230. [PubMed: 9372504]
19. Rudick RA, Lee JC, Simon J, Ransohoff RM, Fisher E. Defining interferon beta response status in multiple sclerosis patients. *Ann Neurol*. 2004; 56(4):548–555. [PubMed: 15389896]
20. Río J, Rovira A, Tintoré M, et al. Relationship between MRI lesion activity and response to IFN- β in relapsing-remitting multiple sclerosis patients. *Mult Scler*. 2008; 14(4):479–484. [PubMed: 18562504]
21. Durelli L, Barbero P, Bergui M, et al. Italian Multiple Sclerosis Study Group. MRI activity and neutralising antibody as predictors of response to interferon beta treatment in multiple sclerosis. *J Neurol Neurosurg Psychiatry*. 2008; 79(6):646–651. [PubMed: 17986500]
22. Rudick RA, Polman CH. Current approaches to the identification and management of breakthrough disease in patients with multiple sclerosis. *Lancet Neurol*. 2009; 8(6):545–559. [PubMed: 19446274]
23. Prosperini L, Gallo V, Petsas N, Borriello G, Pozzilli C. One-year MRI scan predicts clinical response to interferon beta in multiple sclerosis. *Eur J Neurol*. 2009; 16(11):1202–1209. [PubMed: 19538207]
24. Havrdova E, Galetta S, Hutchinson M, et al. Effect of natalizumab on clinical and radiological disease activity in multiple sclerosis: a retrospective analysis of the Natalizumab Safety and Efficacy in Relapsing-Remitting Multiple Sclerosis (AFFIRM) study. *Lancet Neurol*. 2009; 8(3):254–260. [PubMed: 19201654]
25. Giovannoni G, Cook S, Rammohan K, et al. CLARITY study group. Sustained disease-activity-free status in patients with relapsing-remitting multiple sclerosis treated with cladribine tablets in the CLARITY study: a post-hoc and subgroup analysis. *Lancet Neurol*. 2011; 10(4):329–337. [PubMed: 21397565]
26. Giovannoni G, Comi G, Cook S, et al. CLARITY study group. A placebo-controlled trial of oral cladribine for relapsing multiple sclerosis. *N Engl J Med*. 2010; 362(5):416–426. [PubMed: 20089960]
27. Kappos L, Radue EW, O’Connor P, et al. FREEDOMS Study Group. A placebo-controlled trial of oral fingolimod in relapsing multiple sclerosis. *N Engl J Med*. 2010; 362(5):387–401. [PubMed: 20089952]
28. Simon JH, Li D, Traboulsee A, et al. Standardized MR imaging protocol for multiple sclerosis: Consortium of MS Centers consensus guidelines. *AJNR Am J Neuroradiol*. 2006; 27(2):455–461. [PubMed: 16484429]

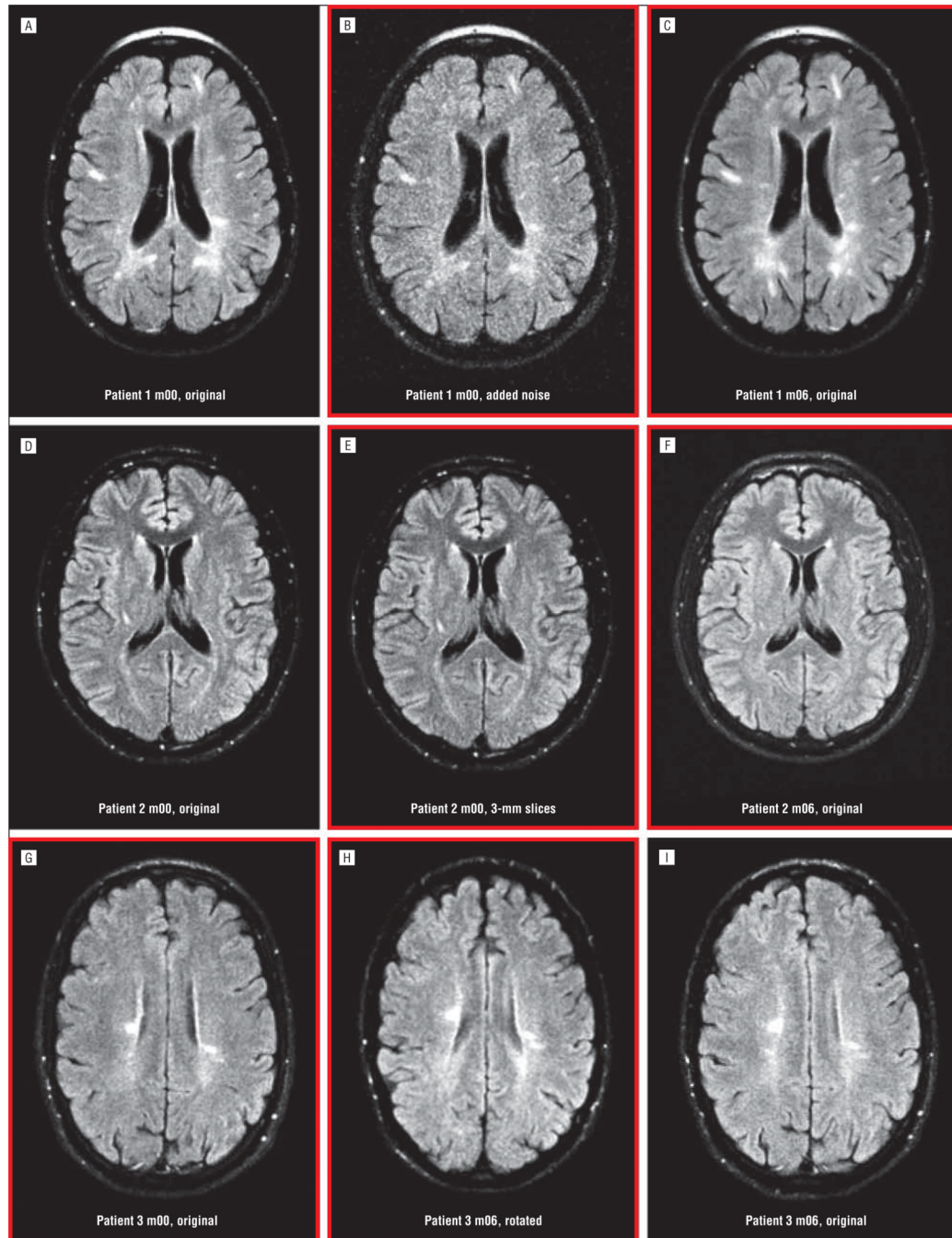


Figure 1.

Examples of image pairs including original and modified images for 3 different subjects. Pairs that were compared are outlined in red. For patient 1 (A-C), the baseline image (m00) was modified by adding noise (B). For Patient 2 (D-F), the baseline image was modified by resampling the data to have a slice thickness of 3 mm instead of 5 mm (E). For Patient 3 (G-I), the month 6 image (m06) was modified by applying a rigid transformation to rotate the volume (H).

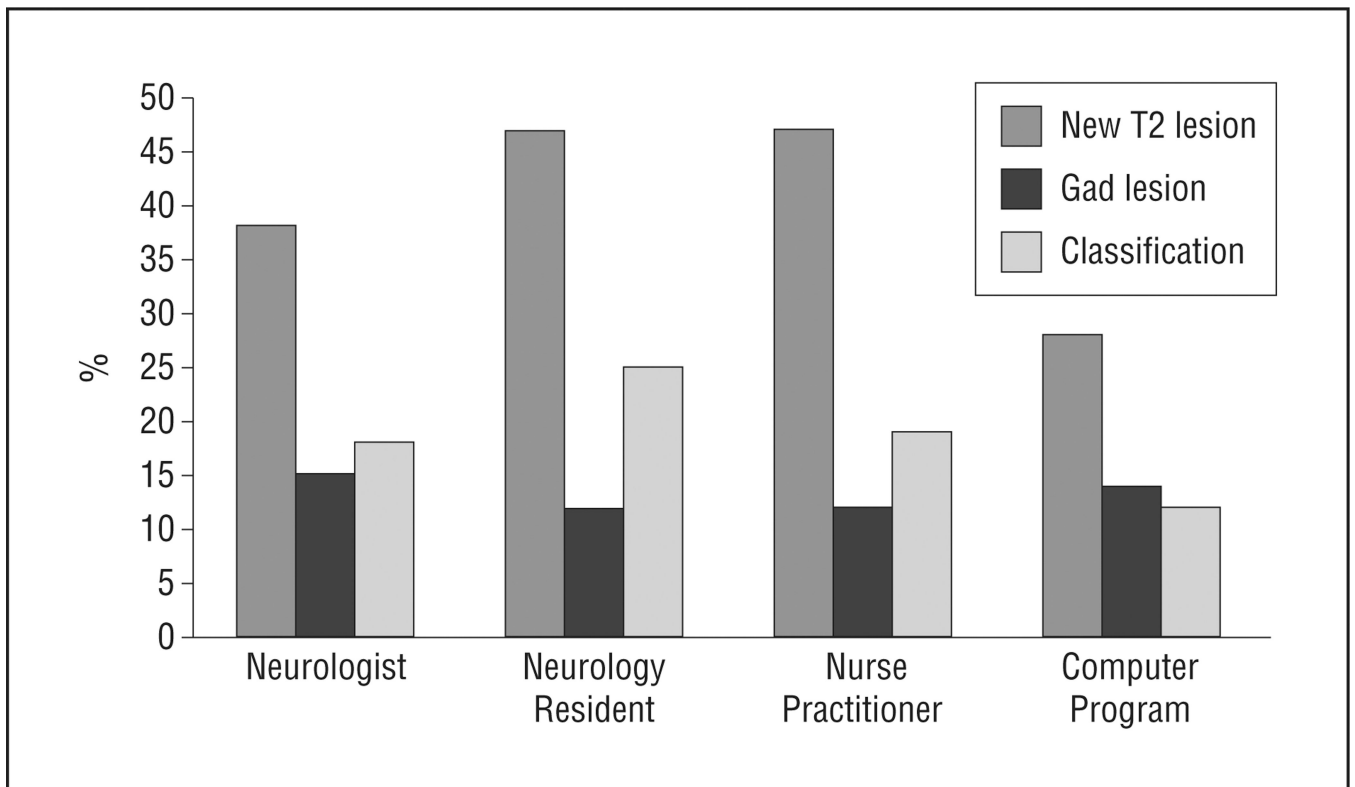


Figure 2. Percentage of discrepant counts for new T2 lesions and gadolinium-enhancing (Gad) lesions and percentage of discrepant classification for magnetic resonance imaging activity status.

Table 1

Baseline Characteristics

| | % | | | <i>P</i> Value (Active vs Inactive) |
|------------------------------------|--------------------------|---------------------------------|-------------------------------|---|
| | All Patients (n = 85) | Inactive MRI at F/U (n = 70) | Active MRI at F/U (n = 15) | |
| Age, y, mean (SD) | 35.7 (9.7) | 36.3 (9.4) | 33 (11.2) | .30 |
| Duration of symptoms, y, mean (SD) | 2.4 (2.9) | 2.5 (3.0) | 1.2 (1.7) | .39 |
| Female | 65 | 69 | 47 | .11 |
| White | 91 | 93 | 80 | .14 |
| EDSS score, mean (SD) | 1.6 (1) | 1.6 (1) | 1.6 (1.2) | .91 |
| Patients with enhancing lesions | 29.4 | 24.3 | 53.3 | .03 |
| T2 volume, mL, mean (SD) | 3.5 (3.8) | 3.0 (3.7) | 5.8 (3.9) | .02 |
| T1 volume, mL, mean (SD) | 0.61 (0.77) | 0.55 (0.75) | 0.87 (0.82) | .19 |

Abbreviations: CIS, clinically isolated syndrome; EDSS, Expanded Disability Status Scale; F/U, follow-up; MRI, magnetic resonance imaging; RRMS, relapsing-remitting multiple sclerosis.

Table 2

Number of New T2 Lesions at 6 Months Compared With Baseline

| Neuroradiologist | Neurologist | | | | Neurology Resident | | | | Nurse Practitioner | | | | Computer Program | | | | | |
|------------------|-------------|----|---|---|--------------------|---|----|---|--------------------|---|---|----|------------------|---|---|----|---|---|
| | 0 | 1 | 2 | 3 | 4 | 0 | 1 | 2 | 3 | 4 | 0 | 1 | 2 | 3 | 4 | | | |
| 53 | 0 | 43 | 6 | 1 | 2 | 1 | 37 | 8 | 7 | 0 | 1 | 38 | 10 | 4 | 1 | 45 | 7 | 1 |
| 18 | 1 | 10 | 5 | 1 | 2 | 7 | 4 | 3 | 1 | 3 | 8 | 4 | 4 | 2 | 6 | 11 | 1 | 1 |
| 6 | 2 | 3 | 1 | 1 | 1 | 2 | 0 | 1 | 3 | 1 | 0 | 2 | 3 | 1 | 2 | 1 | 2 | 1 |
| 3 | 3 | 1 | 1 | 0 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| 5 | 4 | 1 | 1 | 1 | 2 | 2 | 3 | 3 | 1 | 1 | 3 | 1 | 1 | 3 | 3 | 1 | 1 | 1 |

Table 3

Number of Gadolinium-Enhancing Lesions on the 6-Month Scan^a

| Neuroradiologist | Neurologist | | | | | Neurology Resident | | | | | Nurse Practitioner | | | | | Computer Program | | | | |
|------------------|-------------|----|---|---|---|--------------------|---|---|---|---|--------------------|---|---|---|---|------------------|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 0 | 1 | 2 | 3 | 4 | 0 | 1 | 2 | 3 | 4 | 0 | 1 | 2 | 3 | 4 |
| 71 | 0 | 65 | 4 | 1 | 1 | 67 | 3 | 1 | | | 68 | 3 | | | | 66 | 4 | 1 | | |
| 9 | 1 | 6 | 2 | 1 | | 4 | 4 | 1 | | | 4 | 4 | 1 | | | 5 | 3 | 1 | | |
| 1 | 2 | | 1 | | | 1 | | | | | 1 | 0 | | | | 1 | 0 | | | |
| 1 | 3 | | | 1 | | 1 | 0 | | | | 1 | 0 | | | | | | | | 1 |
| 1 | 4 | | | | 1 | | | | 1 | | | | | 1 | | | | | | 1 |

^aIn 2 subjects, gadolinium was not administered on the month 6 scan.

Table 4

Agreement Between Raters on Response Category

| Neuroradiologist | Neurologist | | Neurology Resident | | Nurse Practitioner | | Quantitative MRI | | |
|-----------------------|-------------|--------|--------------------|--------|--------------------|--------|------------------|--------|----|
| | Inactive | Active | Inactive | Active | Inactive | Active | Inactive | Active | |
| Inactive ^a | 72 | 65 | 7 | 55 | 17 | 61 | 11 | 66 | 6 |
| Active ^b | 13 | 8 | 5 | 4 | 9 | 5 | 8 | 4 | 9 |
| Total | 85 | 73 | 12 | 59 | 26 | 66 | 19 | 70 | 15 |

Abbreviation: MRI, magnetic resonance imaging.

^a Inactive: MRI demonstrates 2 or less new or newly enlarging T2 lesions at month 6.

^b Active: MRI demonstrates 3 or more new or newly enlarging T2 lesions at month 6.