ABSTRACT

Sensitive and reproducible measures of multiple sclerosis (MS) severity and progression are important in the treatment of patients and in the design of MS clinical trials. Some of the most commonly used MS rating scales are limited in usefulness because of poor sensitivity and considerable physician-to-physician variability in patient ratings. Recent clinical studies have found that quantitative, continuous rating scales, such as the Multiple Sclerosis Functional Composite, are more sensitive to change related to MS over time and are less variable than the qualitative rating scales often used. Despite the effectiveness of these quantitative measures, a significant obstacle to the use of clinical rating scales in MS studies is that disease activity and progression are mostly subclinical during the relapsing-remitting stage of the disease. This is a problem because early treatment is theoretically more effective than delayed therapy. Thus, other biologically based markers of MS progression are required, such as brain volumes or lesion-based metrics. Clinical studies have begun to characterize how these markers change during the course of MS and how they are affected by treatments that are known to improve neurologic function. (Adv Stud Med. 2004;4(4B):S306-S311)

SENSITIVE AND MRI MARKERS OF MS DISEASE PROGRESSION

Richard A. Rudick, MD

Sensitive and reproducible clinical rating scales for multiple sclerosis (MS) are essential to monitoring the course of illness in individual patients and to evaluating potential new therapies in controlled clinical trials. The development of MS clinical rating scales is challenging because the disease is often clinically silent and because it is not obvious which clinical markers best identify disease progression. Several neurologist-rated MS scales have been developed, in which the physician rates disease severity using qualitative ordinal ratings (eg, “mildly affected,” “severely affected”). The most well known of these ordinal scales is the Expanded Disability Status Scale (EDSS); other common examples include the Ambulation Index (AI) and the Scripps Neurologic Rating Scale (SNRS). An important limitation of ordinal rating scales is that by grouping the patients into a relatively small number of distinct categories, the information is “truncated”—important information about variability between and within patients may be lost by confining an individual to one rating step for long periods of time. This can lead to reduced sensitivity in identifying changes over time or between treatment groups. Also, the boundaries between the steps on ordinal scales are rarely defined precisely, which leads to imprecision or high variability between raters.

An alternative to the use of qualitative ratings is the use of quantitative scales in which neurologic function is rated using a continuous numeric scale (eg, timed tasks, distance walked) in a standardized manner. Several quantitative MS assessments are available, including the Quantitative Evaluation of Neurologic Function, Quantitative Motor Testing, and the Multiple Sclerosis Functional Composite (MSFC). Quantitative continuous measures are preferred over ordinal measures for the evaluation of MS-related function in clinical studies because they offer greater precision and sensitivity than ordinal rating scales and because the statistical analysis of the results is more
The precision of a clinical rating scale (the degree to which repeated observations provide similar outcomes in the absence of any biologic change) is of particular concern in the design of clinical trials. Two sources of variability affect the measured values of any test: technical variability, which is caused by the imprecision of the test, and biologic variability, which is the result of changes in the disease state. The goal of designing a test is to reduce the variability due to imprecision, making it easier to detect changes due to biologic variability of the disorder. One challenge, however, is interpreting the clinical meaning of change in continuous measures. How much change is clinically relevant for a particular test?

The advantages of continuous rating scales in clinical research were illustrated in a recent study by Schwid and colleagues, who examined several outcome measures in a prospective clinical trial of 133 patients with secondary progressive multiple sclerosis (SPMS). The patients were evaluated over 6 months in a randomized double-blind trial of roquinimex or placebo. The investigators used several ordinal scales (including the EDSS, AI, SNRS, and others) and continuous measures (the timed 25-ft walk and 9-hole peg tests, both of which use time as their outcome measure). The continuous measures were more sensitive to changes in MS severity compared with the ordinal measures. The EDSS identified 25% of patients as having worsened during the course of the study, and the AI identified 20% as having worsened. The timed 25-ft walk, the 9-hole peg test, or a composite of the two identified between 54% and 69% of the patients as worsened during the course of the study. Statistically significant worsening from baseline was noted for the quantitative scales but not for the qualitative scales. This suggests that in a clinical trial, effective treatment could slow progression of the disease but not be identified as effective using ordinal rating scales because of poor sensitivity. One could argue that only small effects would be missed because of this problem. However, the course of MS is one of very slow evolution over long periods of time. A 20% or 30% change in the slope of disease progression, if maintained, may translate into significant benefits later. The relative merits of "clinically meaningful" but insensitive outcome measures vs change on highly precise and sensitive measures of uncertain clinical significance is a point of debate.

A second problem with qualitative scales was also illustrated in this study. On the SNRS, about 46% of the patients were classified as worsened, a percentage similar to results on the quantitative rating scales. The variability in this measure was quite large, however, presumably because of inconsistencies in the way that different physicians rated patients (mean change from baseline was 0.3 points, but the standard deviation of this measure was 9.5 points). As a result of this variability, change over time in this outcome measure was not statistically significant despite the relatively large number of patients identified as worsened from baseline.

In 1996, a task force of the National Multiple Sclerosis Society recommended the use of the MSFC for future clinical trials of MS treatments. The MSFC is a 3-part composite rating scale that measures arm function (9-hole peg test, measured in seconds), walking (timed 25-ft walk, measured in seconds), and cognitive function (paced auditory serial addition test, measured using the number of correct responses). Scores on the MSFC correlate reasonably well with the EDSS (r = -0.47), with the greatest correlations between the EDSS and the walking subscale of the MSFC. There is moderately good correlation between EDSS and the arm function subscale on the MSFC and virtually no correlation between EDSS and the cognitive function subscale of the MSFC. This indicates that the MSFC includes measures of disease dimensions that are not well represented in the EDSS. The individual component scores are converted to standardized Z-scores, which are then averaged to produce a composite Z-score to indicate how much better or worse the patient's functioning is in comparison with a reference population. The use of Z-scores (number of standard deviations above or below the reference population mean) permits inclusion of measures with different metric properties within a single rating scale.

Use of the MSFC in clinical research is illustrated by the recent International Multiple Sclerosis Secondary Progressive Avonex® Clinical Trial (IMPACT), which was the first study to use the MSFC as the prespecified principal outcome measure. IMPACT was a multicenter, randomized, double-blind, placebo-controlled study of 436 patients with SPMS. The patients were randomized to treatment with either interferon (IFN)-beta-1a 60 µg or placebo administered once per week by intramuscular injection for 24 months. To attenuate the known learning effect observed with MSFC and to establish a true baseline, patients underwent 3 prebaseline MSFC test sessions before their baseline visit.
During the course of the clinical trial, patients in the placebo group declined by an average Z-score of 0.495, approximately equivalent to a decrease of one half of the standard deviation of the entire pooled study population at baseline. Among patients who received IFN, the decrease in the MSFC was 0.36 standard deviation units, and the comparison across study arms showed that the deterioration was significantly less in the IFN group (P = .033). This beneficial effect was primarily attributed to the arm function scale and to some degree by the cognitive function scale. Benefits were also noted for patients who received IFN-beta-1a on several secondary measures: there was a 33% relative reduction in relapse rate, a significant reduction in number of new or enlarging lesions on T2-weighted magnetic resonance imaging (MRI), and a significant reduction in the number of gadolinium-enhancing lesions. However, there was no benefit on time to worsening on EDSS or on EDSS change and no benefit on change in the timed 25-ft walk (a component of the MSFC). This study illustrates several important points: (1) Different rating scales (MSFC vs EDSS) are not equally sensitive—MSFC showed treatment differences whereas EDSS did not; (2) Different neurologic systems may not respond equally to intervention—there were treatment benefits observed in upper extremity measures but not walking measures. This study has been interpreted by some as showing small but statistically significant benefits observed only with a very sensitive measure, and questions have been raised about the clinical significance of a particular amount of change on the MSFC. This is a question of current research and debate.

The clinical significance of changes in MSFC scores has been addressed by determining the relationship between 1- or 2-year change on MSFC and subsequent clinically relevant disability. In one study, a subset of patients from the original Avonex trial (those patients who entered the original trial early enough to reach 2 years on study) and with relapsing-remitting multiple sclerosis (RRMS) was followed up 8 years after the original trial began. The investigators evaluated the relationship between the change in MSFC score during the initial 2-year study with subsequent functional status approximately 6 years later. The patients were classified into quartiles on the basis of their MSFC worsening during the initial 2-year study. A total of 160 patients had follow-up ascertainment (137 with clinic visits, the remainder with telephone interviews) out of an original study population of 172 patients who had been followed for at least 2 years. A strong relationship was noted between the change in MSFC during the first 2 years and the proportion of patients with EDSS scores of 6 or greater at the end of long-term follow-up (Figure 1). This study demonstrated that MSFC changes during the earlier stages of MS were to some degree predictive of clinically relevant disability at a later stage. Another study using the same patient dataset found that scores on the MSFC were more closely related than EDSS scores to the degree of whole-brain atrophy evident on MRI.

A similar study was reported by Cohen and colleagues at the joint meeting of the American and European Committees for Treatment and Research in Multiple Sclerosis in 2002. These investigators related the change in MSFC in patients with progressive MS to ambulatory status an average of 10 years later. This was a follow-up study of patients who were initially enrolled in a randomized, double-blind clinical trial that compared methotrexate with placebo for progressive MS in patients with baseline EDSS scores between 3.0 and 6.5. In the initial clinical trial, the investigators reported that methotrexate improved the function of the upper extremities but did not benefit ambulation. Cohen and colleagues performed a 10-year follow-up study comparing vital status and ambulation for all 60 patients enrolled in the original trial. The patients who were ambulatory at 10-year followup had

![Figure 1. Change in MSFC Over 2 Years Predicts Ambulatory Status 8 Years Later](image-url)
exhibited significantly better MSFC scores at baseline than patients who were not ambulatory at followup. In addition, the 1-year change in MSFC during the initial study was also significantly related to patient status (non-ambulatory or deceased) at 10-year follow-up. Again, this suggests that quantitative measurement with the MSFC during short intervals is related to disease progression of obvious clinical significance at later time points.

**Brain and Spinal Cord Atrophy in MS**

As described previously, several studies have found that quantitative clinical rating scales are at least somewhat predictive of the future course of MS and are also sensitive to the effects of pharmacologic treatment of MS. However, it is also important to consider that a significant portion of the disease is subclinical. In particular, many groups have shown central nervous system (CNS) tissue loss beginning early in the disease and progressing without obvious clinical correlates. Therefore, clinical measures should be supplemented by regional and global measures of brain and spinal cord atrophy. Progressive brain atrophy is observed during the course of MS, and in later stages of the disease it correlates with the degree of neurologic disability. Because a large proportion of the human brain is composed of axons (approximately 46% of brain volume) and myelin (approximately 24%), cerebral atrophy must reflect loss of myelin and axons. Measuring atrophy should be useful in future clinical trials for monitoring neuroprotection or repair.

The rate of atrophy progression over time in MS and how this rate changes during the progression of the disease are important but poorly understood issues. Fox and colleagues measured cerebral atrophy over the course of 1 year in patients with MS. In order to overcome the variability inherent in the location and extent of MS lesions, these investigators measured whole-brain volume rather than lesion volume within particular prespecified brain regions. Changes in brain volume were examined longitudinally by examining brain volume changes on registered MRI scans. Changes in brain volume for 26 patients with MS were compared with those of a group of healthy control subjects who were matched to the patients on the basis of age and sex. The investigators found that patients with MS exhibited a median loss of brain volume of 0.8%, compared with a median loss of 0.3% for matched controls. The rate of atrophy was similar regardless of the specific subtype of MS. Similarly, Ge and colleagues, who examined changes in the brain parenchymal volume (the ratio of brain matter to the total cerebral volume) over follow-up durations of 1 to 7 years, reported an average loss of 1.6% in the parenchymal fraction per year, with no significant difference between patients with RRMS and SPMS. Kalkers and colleagues, who conducted a larger study (42 patients with RRMS, 21 with SPMS, and 20 with primary progressive MS) also examined the brain parenchymal fraction. With a follow-up period of 2 to 4 years, these investigators described annualized changes in parenchymal fraction of 0.7% per year, again with no difference between the different subgroups of patients.

When reviewing these studies of longitudinal changes in brain volume, it is important to distinguish these changes from normal brain atrophy associated with aging. In normal healthy volunteers, brain imaging studies have suggested that atrophy rates are very low before the age of 50 to 60 years. A study in progress to compare atrophy rates prospectively in 18 healthy control subjects, 14 patients with clinically isolated syndromes, 34 patients with RRMS, and 21 patients with SPMS. The initial findings from this study (the results of which have not yet been published) suggest minimal atrophy in healthy subjects or patients with clinically isolated syndromes but significant atrophy over the course of 12 months in patients with MS. Another way to evaluate the effects of normal aging is by examining the proportion of the variance in brain volume that is explained by patient age. In healthy control subjects, 44.5% of the variance in volume was explained by age, whereas in patients with RRMS or SPMS, only 6.4% and 1.3%, respectively, was explained by age. This is strong indirect evidence that atrophy in MS is caused by the disease and that age contributes little.

Several clinical trials have examined the effects of treatment on atrophy in patients with MS. A post-hoc analysis of the clinical study that examined IFN-beta-1a for the treatment of MS was conducted to determine whether IFN-beta-1a treatment, which slowed the rate of deterioration of MS, also slowed the progression of atrophy during the course of the 2-year study. Atrophy was examined in a subgroup of 140 patients for whom MRI scans were available at baseline and after 1 and 2 years of randomized treatment. These investigators found the atrophy progressed at similar rates in the IFN-beta-1a and placebo groups during the first year. However, during the second year, atrophy continued to progress in the placebo patients at the same rate but...
slowed significantly in the group of patients who received IFN-beta-1a (Figure 2). The authors argued that inflammation that was present before randomization into the study resulted in atrophy during the early part of the clinical trial and that anti-inflammatory therapy affected the rate of atrophy after a lag period and was therefore evident in the second year of the study. Several other reports have also described significant reductions in atrophy during 2 to 3 years of treatment with IFN-beta-1a in patients with RRMS. Zivadinov and colleagues reported that pulse intravenous (IV) methylprednisolone was associated with preservation of brain volume over a 5-year period, whereas volume declined significantly in control patients. This study has not yet been reproduced but suggests that pulses of IV methylprednisolone in the earlier stages of MS may be neuroprotective. Not all clinical trials have reported benefits in atrophy progression with treatment, and at least 2 studies that have examined IFN-beta-1a did not report any beneficial effect of treatment on brain volume. In contrast to these findings in RRMS, numerous studies have been conducted in SPMS, but none has shown any significant reduction in atrophy in these patients.

The future of monitoring for destructive pathology in MS may lie with MRI as the primary measure, supplemented by clinical measures. More specific measures that describe changes in CNS tissue, particularly normal-appearing white matter, would be a significant advance. Measures linked to demyelination and axonal integrity or to inflammation would also be extremely valuable. In addition, measures that more strongly predict atrophy would be very useful. Researchers have completed studies using conventional markers of MS lesions, but at present, even the best models can account for only 20% to 25% of the variance in atrophy progression at future time points. Therefore, there are major factors that we are currently missing in our current traditional MRI measures that we need to capture in newer techniques. Regional atrophy measures (e.g., cortex, thalamus, frontal lobes) are likely to be quite informative, and new clinical trial designs are required to adequately assess the effects of treatment on atrophy. Interestingly, Hardmeier and colleagues reported that atrophy was measurable over a 3-month period in patients with RRMS. This suggests that it should be possible to conduct clinical trials in which brain atrophy is a primary endpoint in periods as short as 1 year. Patients could have an untreated observation period of 3 to 4 months, then be randomized to 2 or more treatments and atrophy rates calculated several months later. Sample size calculations for this type of study suggest that relatively large numbers of patients would be needed using this approach, but such a study might provide extremely important data.

**Summary and Conclusions**

Progress in clinical trials and in evaluating evidence for new treatments of the disease is largely a matter of the availability of meaningful measurements that have good metric properties and have been validated. Quantitative clinical measures and precise measures of brain and spinal cord atrophy should be combined in future trials of neuroprotection strategies; eventually, these should be supplemented by more specific MRI-based markers of tissue integrity, demyelination, axonal integrity, and inflammation. It will be extremely important to demonstrate the predictive validity of all techniques used to measure neuroprotective strategies, because of the desire to apply these interventions early in the disease to prevent eventual brain destruction. However, because MS is largely subclinical during its early years, one may not see a clinically apparent benefit for effective neuroprotection.
during a 2- to 3-year period in RRMS. At present, there is an urgent need for better outcome measures and better study designs, which will be a prerequisite for meaningful testing of therapies aimed at neuroprotection.

REFERENCES

7. Cohen JA, Goodman AD, Heidenreich FR, et al. Relationship between the MS functional composite and MRI in IMPACT. Presented at the 7th Annual Meeting of the American Committee for Treatment and Research in Multiple Sclerosis and the 18th Congress of the European Committee for Treatment and Research in Multiple Sclerosis; September 18-21, 2002; Baltimore, Md.