Large, Nonplateauing Relationship Between Clinical Disability and Cerebral White Matter Lesion Load in Patients With Multiple Sclerosis

Zografos Caramanos, MA; Simon James Francis, MSc; Sridar Narayanan, PhD; Yves Lapierre, MD; Douglas Lorne Arnold, MD

Objective: To better characterize the relationship between cerebral white matter lesion load (CWM-LL) and clinical disability by (1) covering the entire range of the Kurtzke Expanded Disability Status Scale (EDSS), (2) minimizing nonbiological sources of variability, and (3) increasing pathologic specificity by studying CWM lesions that are hypointense on T1-weighted magnetic resonance imaging.

Design: Cross-sectional, retrospective study.

Setting: Hospital-based multiple sclerosis (MS) clinic.

Patients: A total of 110 patients with untreated MS were recruited and studied from June 1, 1997, through June 30, 2003.

Main Outcome Measures: Cube-rooted CWM-LL and EDSS-measured clinical disability scores.

Results: We found a large, nonplateauing relationship between cube-rooted CWM-LL and concurrent EDSS scores, more so for T1-hypointense than T2-hyperintense lesions ($r = 0.619$ vs $0.548$). Correlations between the EDSS scores and CWM-LL diminished when, as typically done in clinical trials, only those patients with EDSS scores of 0 to 6.0 were studied ($n = 92; r = 0.523$ for T1-hypointense lesions and $r = 0.457$ for T2-hyperintense lesions); more important, a series of boot-strapped correlations suggested that this decrease was not simply due to smaller sample size, and these relationships remained even after correcting for disease duration.

Conclusion: A large, nonplateauing relationship exists between CWM-LL and EDSS-measured clinical disability when patients with MS are studied to examine the entire range of disability, minimize nonbiological sources of variability, and increase pathologic specificity.

Arch Neurol. 2012;69(1):89-95

Multiple sclerosis (MS) is an inflammatory, demyelinating, and degenerative autoimmune disorder of the central nervous system.1 Its pathological hallmark is the presence of multifocal, demyelinating, white matter lesions of the brain and spinal cord that are disseminated in space and time.2 All of these white matter lesions are focal regions of inflammation, demyelination, oligodendroglial loss, reactive gliosis, and axonal degeneration.3 Such lesions are readily visible in vivo as hyperintensities on T2-weighted magnetic resonance imaging (MRI) and, more rarely, hypointensities on T1-weighted MRI.4 Depending on their location, such T2- and T1-weighted lesions can differentially result in many of the observed MS-related symptoms.5 Importantly, total cerebral white matter lesion load (CWM-LL) is used as an outcome measure in almost all clinical trials of patients with MS; thus, fully understanding the relationship between such MRI measures and clinical disability is of great importance.6

Despite numerous criticisms,7,9 clinical disability in patients with MS is typically evaluated using the Kurtzke Expanded Disability Status Scale (EDSS).10,11 with scores ranging from 0 (normal neurologic examination results) to 10 (death due to MS). Recent reviews7,12,13 have suggested that correlations between EDSS-measured clinical disability and T2-weighted CWM-LL are variable, ranging from 0.09 to 0.60 (median, 0.33) (based on 16 comparisons, each including between 39 and 718 patients; median, 58 patients). (Note that, according to Cohen’s14 definitions of effect sizes, a
correlation of 0.10 is considered small, 0.30 moderate, and 0.50 large.)

Consistent with this are the results by Li et al., who found a Spearman rank-order $\rho$ value of 0.35 ($P < 0.001$) between T2-weighted CWM-LL and EDSS in 1312 patients with MS (and EDSS scores between 0 and 6.5) who had been placebo controls in 1 of 11 randomized controlled trials included in the Sylvia Lawry Centre for MS Research database. Importantly, they found EDSS scores to be positively related with T2-weighted CWM-LL only in patients with scores between 1.5 and 4.5, after which, they suggest, the relationship reaches a plateau. This plateauing relationship was recently reassessed by Sormani et al., who studied the results of 877 patients with MS (also with EDSS scores between 0 and 6.5) who were placebo controls in 2 other recent clinical trials. They also found evidence of a moderate relationship between T2-weighted CWM-LL and EDSS (Pearson product moment $r = 0.39$, $P < 0.001$) but did not, however, find statistical evidence of a plateauing relationship. Similarly, Fisniku et al. also did not detect a plateauing T2-weighted CWM-LL relationship with clinical disability for 20 years.

The study by Li et al. points out a number of sources of variability associated with multicenter clinical trials and even more so with studies that combine data from many such trials. First, the data included in their analysis came from 11 different clinical trials, each with potentially different entry criteria regarding the following: (1) EDSS scores, (2) relapse rates, (3) lesion status on gadolinium-enhancing T1-weighted MRI, and (4) disease phase (eg, relapsing-remitting MS [RR-MS] or secondary-progressive MS [SP-MS]). Any of these could have an effect on the range of EDSS scores and T2-weighted CWM-LL values that might be expected in the patients ultimately included in these clinical trials. Second, depending on the specific trial, the T2-weighted CWM-LL data were (1) generated from MRI data collected on different scanners using potentially different acquisition parameters and (2) processed and analyzed by different people at different centers using potentially different protocols. Each of these are potential sources of nonbiological noise that could increase the variability associated with the MRI-measured T2-weighted CWM-LL values. Third, EDSS scores were generated by many different health care professionals at many different clinical centers, which are additional sources of variability associated with these scores. Importantly, any or all of these sources of increased noise and variability could (1) combine to decrease the strength of the relationship observed by Li et al. between their patients' T2-weighted CWM-LL values and EDSS scores and (2) result in the high degree of variability in the relationship between the T2-weighted CWM-LL values and the EDSS scores that was described previously.

In addition to the aforementioned factors, the typically moderate correlations observed may also be partially attributed to potentially low statistical power associated with only examining a restricted range of EDSS-measured disability. For example, group studies and clinical trials usually sample only a limited range of the EDSS scores (usually including only patients with scores between 0 and 6.5 or less), truncation that could lead to underestimating the strength of the true overall correlation. Furthermore, they usually include only relatively small samples (eg, <60) even though it is known that, allowing for type I and II errors of 5% and 20%, a sample size of 84 is necessary to consistently find a statistically significant moderate-sized correlation. Moreover, they usually include MRI estimates of only T2-weighted CWM-LL and not those of the more pathologically specific T1-weighted CWM-LL, which have been shown in one recent review13 to have larger and less variable correlations with EDSS, ranging from 0.48 to 0.74 (median, 0.53) (based on 4 comparisons, each including 38 to 82 patients [median, 47 patients]).

The present cross-sectional, retrospective study of a large, representative group of patients with untreated MS was aimed at better examining the relationships among (1) EDSS-measured clinical disability, (2) disease duration and phase (ie, RR-MS or SP-MS), and (3) CWM-LL for both T2 and T1 lesions. We tried to reduce the effect of nonbiological sources of variability by (1) acquiring all of the MRIs on the same scanner and using the same acquisition and postprocessing procedures; (2) having only a small number of highly trained individuals using semiautomated, standardized procedures to quantify our patients' T2-weighted and T1-weighted CWM-LL values; and (3) having only a small number of highly experienced neurologists from the same MS clinic assess our patients' EDSS scores. More important, we did this for both patients with the full range of scores observed on the EDSS (the full EDSS group) and a subset of patients with a limited range of scores similar to that typically studied (ie, 0-6; the limited EDSS subgroup).

### METHODS

#### STUDY PARTICIPANTS

A total of 110 patients with clinically definite MS were recruited via the MS clinic at the Montreal Neurological Hospital and studied from June 1, 1997, through June 30, 2003: 82 with RR-MS (62 females) and 28 with SP-MS (15 females). All 110 patients were included in the full EDSS group. Ninety-two patients with EDSS scores of 0 through 6.0 were included in the limited EDSS subgroup: 79 with RR-MS (59 females) and 13 with SP-MS (6 females). All patients were untreated and relapse free at the time of study, and none had a history of immunomodulatory therapy before study. The ethics committee of the Montreal Neurological Hospital approved the study, and informed consent was obtained from all participants. Patients' EDSS scores were assigned by the same highly experienced neurologist who had been observing them within the MS clinic. For each patient, disease duration was calculated as follows: ([date of scan) – (date of symptom onset)]/365.25.

#### BRAIN MRI

Brain MRIs were acquired on a 1.5-T Philips Gyroscan ACS II scanner (Philips Medical Systems) using a body-coil transmitter and a quadrature head-coil receiver. Fifty contiguous, 3-mm-thick, T2- and proton density–weighted images were acquired parallel to the callosal line using a dual-turbo spin-echo sequence (repetition time, 2073 milliseconds; echo times, 31.6 and 90 milliseconds; 256 × 256 matrix; and a 250-mm field of view). T1-weighted images were acquired with the same ma-
trix using a 3-dimensional gradient-echo sequence (repetition time, 35 milliseconds; echo time, 10.2 milliseconds; and a 40° excitation angle). Each MRI volume was corrected for image-intensity inhomogeneity using N3,18 and the T2- and proton density–weighted volumes were registered to the T1-weighted volume using a mutual information-based approach.19

DETERMINATION OF T2- AND T1-WEIGHTED CWM-LL VALUES

As shown in Figure 1, T2- and T1-weighted CWM lesions were segmented using a manually corrected, automatic, Bayesian, tissue classification approach.20 On the basis of these segmentations, T2- and T1-weighted CWM-LL values were generated automatically for each patient. (A detailed report of the extremely high intrarater and interrater reliability associated with these T2- and T1-weighted CWM-LL values is available in the “Methods” section of the online Supplementary Material available at http://www.bic.mni.mcgill.ca/PersonalCaramanosZografos/SupplementaryMaterial.)

STATISTICAL ANALYSIS

As in the studies by Li et al15 and Sormani et al,6 cube-rooted transformations were used to eliminate the skew that is typically associated with T2- and T1-weighted CWM-LL data, thereby yielding a more suitable variable for parametric analyses. Scores on the EDSS were treated both as ordinal data, which were tested with nonparametric statistics and are appropriate given the nonlinear nature of the scale,5,21 and continuous data, which were tested with parametric statistics, which are typically used in studies of patients with MS (and as was done in the studies by Li et al15 and Sormani et al6). Differences in group means (ie, RR-MS vs SP-MS) were evaluated using 1-way analysis of variance, which also generate $r^2$ values that indicate how much of the variability in the dependent variable is explained by the grouping variable. Group dif-
Inferences in EDSS scores were also evaluated nonparametrically using Mann-Whitney tests. Bivariate relationships were evaluated nonparametrically (using the Spearman rank-order correlation $r$ values) and parametrically (using Pearson product moment correlation $r$ values). Furthermore, a boot-strapping approach was used to determine whether any differences in the $r$ values for the 92 patients in the limited EDSS group and the 110 patients in the full EDSS group were due to differences in their sample sizes or differences in their range of EDSS scores: 100 random samples of 92 patients each were drawn from the full EDSS group, and bootstrapped $r$ values were calculated for each of these 100 sets of data. Finally, partial correlations were used to examine bivariate relationships while controlling for the patients’ symptom durations.

Scatterplots were used to illustrate the bivariate relationships. These scatterplots were superimposed with the 95% CI of the associated linear smoother and a LOWESS (Locally Weighted Scatterplot Smoothing) smoother,22 which was used to try to detect any nonlinear relationship in the variables. (A detailed description of these LOWESS smoothers is available in the “Methods” section of the online Supplementary Material.)

For the reader to better judge the statistical significance and clinical meaningfulness of our results, we present all of the observed test statistics, all the uncorrected $P$ values, and any associated degrees of freedom. $P < .05$ was nominally considered to be statistically significant; however, all statistically significant findings in the full EDSS group were associated with $P < .001$.

SPSS statistical software, version 12.0, for Windows (SPSS, Inc) was used to generate the partial correlations. SYSTAT 10.2 for Windows (Systat Software) was used for all other statistical and graphic analyses.

## RESULTS

### GROUP DIFFERENCES

Demographic, clinical, and cube-rooted CWM-LL data are summarized in Figure 2. Compared with the patients with RR-MS, those with SP-MS had statistically greater mean ages, mean symptom durations, mean EDSS scores, mean cube-rooted T2-weighted CWM-LL values, and mean cube-rooted T1-weighted CWM-LL values. Importantly, the patients’ MS subgroup (ie, RR-MS vs SP-MS) was consistently able to account for more of the variance in the full EDSS group’s data than the limited EDSS subgroup’s data, sug-
suggesting that sampling patients across the entire range of the EDSS increases the ability of each of these measures to discriminate between patients with RR-MS and SP-MS. Finally, regardless of the EDSS subgroup, the MS subgroup accounted for more of the variance in the patients’ mean cube-rooted T2-weighted CWM-LL values than in their mean cube-rooted T2-weighted CWM-LL values, suggesting that patients with RR-MS can be better discriminated from those with SP-MS by their cube-rooted T1-weighted CWM-LL values than by their cube-rooted T2-weighted CWM-LL values.

**BIVARIATE RELATIONSHIPS WITH EDSS SCORES**

Bivariate relationships between EDSS scores and demographic, clinical, and cube-rooted CWM-LL data are summarized in Figure 3. The ρ and r values were consistent and statistically significant for all comparisons, suggesting that EDSS had a moderate correlation with age, somewhat larger correlation with symptom duration, even larger correlation with cube-rooted T2-weighted CWM-LL, and larger still correlation with cube-rooted T1-weighted CWM-LL.

Importantly, the correlations in the full EDSS group were consistently larger than those in the limited EDSS subgroup, but, as indicated by the mean boot-strapped r values in the full EDSS group being practically the same as their r values, this seemed to be due to the greater range of EDSS scores rather than simply the larger sample size. These findings suggest that sampling the entire range of the EDSS can lead to a better capture and description of these measures’ relationships to MS-related clinical disability.

Interestingly, the correlation of the patients’ EDSS scores and ages was no longer statistically significant after partialing out the effect of symptom duration. On the other hand, statistically significant correlations with the cube-rooted CWM-LL values were maintained even after controlling for symptom duration, remaining larger with cube-rooted T1-weighted CWM-LL than with cube-
rooted T2-weighted CWM-LL and, thereby, further suggesting that T1-weighted CWM-LL values are more clinically relevant than T2-weighted CWM-LL values in patients with MS.

With respect to the plateauing relationship between CWM-LL and EDSS-measured clinical disability scores, inspection of the LOWESS smoother in the limited EDSS subgroup suggests a possible plateau after an EDSS score of approximately 4.0 is reached. Importantly, however, inspection of the data from the full EDSS group confirms that a positive relationship persists throughout the middle and upper range of the EDSS for both cube-rooted T2-weighted CWM-LL and cube-rooted T1-weighted CWM-LL (this is true even despite a small number of cases with cube-rooted CWM-LL values that were much lower than might have been expected given their high scores on the EDSS). Together, these findings suggest that increases in EDSS scores are positively related to increases in CWM-LL values, more so to increases in T1-weighted CWM-LL values than to T2-weighted CWM-LL values, and this relationship does not seem to plateau if the entire range of possible scores on the EDSS is examined. (For results and discussion of the bivariate relationships among the demographic, clinical, and cube-rooted CWM-LL data from these patients, which further suggest that symptom duration and cube-rooted T1-weighted CWM-LL are more clinically relevant in patients with MS than age and cube-rooted T2-weighted CWM-LL, please refer to the “Results” section of the online Supplementary Material.)

**COMMENT**

We found evidence of a large, nonplateauing relationship between our patients’ MRI-measured CWM-LL values and their EDSS-measured clinical disability scores, even after controlling for disease duration. Importantly, this relationship was largest when the entire range of possible scores on the EDSS is examined. (For results and discussion of the bivariate relationships among the demographic, clinical, and cube-rooted CWM-LL data from these patients, which further suggest that symptom duration and cube-rooted T1-weighted CWM-LL are more clinically relevant in patients with MS than age and cube-rooted T2-weighted CWM-LL, please refer to the “Results” section of the online Supplementary Material.)

The latter notion is supported, for example, by findings that EDSS scores in patients with RR-MS show a larger correlation with T2-weighted CWM-LL values within and around the corticospinal tracts than with their total T2-weighted CWM-LL values.23

Despite not finding a plateauing relationship between EDSS and CWM-LL, we found some initial evidence to suggest that this relationship is relatively flat in patients with low scores on the EDSS (ie, ≤2.0). Interestingly, the degree of CWM-LL in these patients with low EDSS scores seemed to be greater than would be expected given a linear relationship across the entire range of the EDSS scale, results that are similar to those shown but not commented on in the studies by Li et al15 and Sormani et al.7 Together, these findings, which need to be confirmed in a larger sample, suggest that patients with MS might be able to compensate for the presence of CWM disease until a certain threshold is reached, which is consistent with functional MRI evidence for adaptive brain changes in the initial stages of MS.24 On the other hand, this flat relationship with EDSS scores of 2.0 or less might also reflect the poor responsiveness of the EDSS, particularly with respect to the lower end of the EDSS.

In the present study, we aimed to better examine the relationship between EDSS-measured clinical disability and MRI-measured CWM-LL in a large group of representative patients with untreated MS. We tried to minimize nonbiological sources of variability, increase statistical power and generalizability by examining this relationship across the entire range of the EDSS, and increase pathological specificity by also examining the relationship of EDSS to T1-weighted CWM-LL. (For a discussion of the importance and implications of this approach, please refer to the “Comment” section of the online Supplementary Material.) Despite the large correlations resulting from these optimized circumstances, we were, at best, able to account for only approximately 38% of the variance in our patients’ EDSS-measured clinical disability using their conventional MRI-measured CWM-LL. It is our belief that the amount of the variance accounted for could be increased by combining such CWM-LL data with findings regarding extracerebral lesions (eg, from the spinal cord and optic nerves) and nonlesional central nervous system disease measured using nonconventional MRI techniques.25-31 Ideally, we hope that such data could be combined to predict not only patients’ concurrent degree of clinical disability but also, and more important, their future clinical course22 and their likely response to a particular given treatment.

**Accepted for Publication:** April 9, 2011.

**Correspondence:** Zografos Caramanos, MA, Montreal Neurological Institute, McGill University, 801 University St, Montreal, Quebec, Canada H3A 2B4 (zografos.caramanos@mcgill.ca).

**Author Contributions:** Study concept and design: Caramanos, Narayanan, and Arnold. Acquisition of data: Caramanos, Francis, and Narayanan. Analysis and interpretation of data: Caramanos, Narayanan, Lapierre, and Arnold. Drafting of the manuscript: Caramanos and Arnold. Critical revision of the manuscript for important intellectual content: Caramanos, Francis, Narayanan,

Financial Disclosure: None reported.

Funding/Support: This work was supported by grant MOP-84367 (Dr Arnold) from the Canadian Institutes of Health Research and from the Multiple Sclerosis Society of Canada. Mr Caramanos was supported by a Neuromagnetism Strategic Training Program Award from the Canadian Institutes of Health Research, a Mayseis MacSparran Graduate Studentship from McGill University's Faculty of Medicine, and a Biomedical Research PhD Studentship from the Multiple Sclerosis Society of Canada.

Additional Contributions: The neurologists at the MS clinic of the Montreal Neurological Hospital provided clinical data regarding their patients, the technicians at the Magnetic Resonance Spectroscopy Unit of the Montreal Neurological Institute segmented our patients' T2- and T1-weighted lesions, and the coordinators at the Magnetic Resonance Spectroscopy Unit provided help in scheduling visits. Most important, we thank all the study participants for their time and cooperation.

REFERENCES