Large, Non-Plateauing Relationship between Clinical Disability and Cerebral-White-Matter Lesion-Load in Patients with Multiple Sclerosis

*Zografos Caramanos[1], M.A., Simon James Francis[1], M.Sc.,
Sridar Narayanan[1], Ph.D., Yves Lapierre[2], M.D., and Douglas Lorne Arnold[1], M.D.

[1]Magnetic Resonance Spectroscopy Unit, McConnell Brain Imaging Centre
[2]Clinical Research Unit

Montreal Neurological Institute, McGill University
3801 University Street
Montreal, Quebec, Canada
H3A 2B4

*Corresponding Author
Tel: (514) 398-8185
Fax: (514) 398-2975
E-Mail: zografos.caramanos@mcgill.ca

Online Supplementary Material

Text Word Count (excluding references): 1,325 words
Total Word Count (including text and references): 1,647 words

Originally submitted to *Archives of Neurology* on February 15th, 2011
to be considered for publication as an "Original Contribution"

Revision submitted electronically to *Archives of Neurology* on April 8th, 2011

ONLINE SUPPLEMENTARY METHODS

Brain MRI

T2-LL and T1-LL Values: Reader Inter- and Intra-Rater Reliability

The seven readers who generated the lesion labels in this study had previously undergone: *(i)* extensive training on similar MRI data from a large set of patients, and *(ii)* post-training testing on a "**Test**" set from 10 representative patients. As determined by their trainer (SJF), these Test scans had: *(i)* manually-corrected, total cerebral-white-matter T2-lesion-load (**T2-LL)** values of between of 1.2 and 62.0 cc, and *(ii)* manually-corrected, total cerebral-white-matter T2-lesion-load (**T2-LL)** values of between 0.2 and 12.9 cc. The readers' manually-corrected Test-scan segmentations showed excellent inter-rater reliability relative to their trainer's reference segmentations, both for: *(i)* T2 lesions [*Dice kappa values*[1] *across the 10 pairs of scans* (**kappa**): mean (range) = 0.94 (0.92 – 0.95); and *intra-class correlations amongst the two sets of 10 scans* (**ICC**): mean (range) = 0.99 (0.99 – 1.00)], and *(ii)* T1 lesions [*kappa* = 0.88 (0.81 – 0.92); *ICC* = 0.99 (0.99 – 1.00)]. The readers' manually-corrected Test-scan segmentations also showed excellent intra-rater reliability with a lag of about one week between reads, again both for *(i)* T2 lesions [*kappa*: 0.94 (0.91 – 0.96); *ICC*: 0.99 (0.99 – 1.00)], and *(ii)* T1 lesions [*kappa*: 0.89 (0.81 – 0.93); *ICC*: 0.99 (0.99 – 1.00)].

Statistical Analysis

Bivariate-Relationships: LOWESS Smoothers

A LOWESS smoother is created by running along the x-values and finding predicted values from a weighted average of nearby y-values – the surface being allowed to flex locally in order to better fit the data. In all of the plots shown in Figure 2, a tension of 0.5 was used, which means that a maximum of 50% of the points were included in each running window used to generate a predicted value; please note, however, that practically identical results were obtained using tensions of 0.2, 0.3, 0.4, and 0.5).

ONLINE SUPPLEMENTARY RESULTS

Bivariate relationships

The **eTable** below presents Pearson-product-moment-correlation matrices showing *r*-values (and associated *p*-values) for all of the bivariate relationships amongst the patients' demographic, clinical, and cube-rooted (**cr**) T2 and T1 lesion-load (**LL**) data. Spearman rank-order correlations were completely consistent with these and are not presented here. Correlations that are statistically significant at $p < 0.05$ are bolded, and those that are large (*i.e.*, $r > 0.50$) are underlined.

Correlation Matrix for the Limited-EDSS Subgroup (n = 92; EDSS 0 - 6.0)

r (*p*)	Age	Duration	EDSS Score	cr-T2-LL
Duration	**0.600 (< 0.00001)**	-	-	-
EDSS Score	**0.254 (0.0147)**	**0.337 (0.00101)**	-	-
cr-T2-LL	0.033 (0.75692)	**0.324 (0.00164)**	**0.457 (< 0.00001)**	-
cr-T1-LL	0.118 (0.26256)	**0.346 (0.00072)**	**0.523 (< 0.00001)**	**0.890 (< 0.00001)**

Correlation Matrix for the Full-EDSS Group (n = 110; EDSS 0 - 9.5)

r (*p*)	Age	Duration	EDSS Score	cr-T2-LL
Duration	**0.628 (< 0.00001)**	-	-	-
EDSS Score	**0.325 (0.00054)**	**0.456 (< 0.00001)**	-	-
cr-T2-LL	0.135 (0.16082)	**0.393 (0.00002)**	**0.548 (< 0.00001)**	-
cr-T1-LL	**0.230 (0.01579)**	**0.442 (< 0.00001)**	**0.619 (< 0.00001)**	**0.904 (< 0.00001)**

Regardless of EDSS Subgroup (*i.e.*, Limited-EDSS vs. Full-EDSS), we found that: *(i)* even though the patients' ages and symptom durations were strongly correlated, their symptom durations consistently showed larger correlations with both their cr-LL values and their EDSS scores than did their ages; and *(ii)* even though the patients' cr-T2-LL values and cr-T1-LL values were strongly correlated, their cr-T1-LL values consistently showed

larger correlations with their EDSS scores than did their cr-T2-LL values. Together, these findings suggest that the patients' symptom durations and T1-LL values are more clinically-relevant than their ages and T2-LL values.

ONLINE SUPPLEMENTARY COMMENT

The benefits of decreasing variability while increasing statistical power and pathological specificity

In the present study, we attempted to minimize the amount of non-biological sources of variability that may have contributed to previous findings of highly-variable and, on average, only-moderate correlations between EDSS-measured clinical disability and T2-LL in patients with MS.[2-4] Furthermore, we attempted to maximize our statistical power by: *(i)* examining the entire range of the EDSS, and *(ii)* including data from a large sample of patients with MS. Finally, we tried to increase the pathological specificity of the patients' MRI-measured cerebral-WM-LL values by also quantifying their T1-LL, which has been shown to be more indicative of the extent of destructive lesional-pathology in patients with MS.[5] In so doing, we were able to find statistically-significant evidence for: *(i)* a large relationship between our patients' EDSS scores and their T2-LL values, and *(ii)* an even-larger relationship with their T1-LL values. Importantly, we were also able to show that consideration of scores across the entire range of the EDSS: *(i)* allows for a more-complete description of the relationship between patients' clinical disability and their cerebral-WM-LL values, and *(ii)* suggests that this large relationship is maintained across the middle and the upper ranges of the EDSS scale.

Whereas we attempted to minimize any sources of non-biological and extra-patient variability, the data in the Li *et al.*[6] study was associated with the many potential sources of increased variability that are present when data from many clinical trials are combined (as detailed in the Introduction). Importantly, it is not clear: *(i)* what the inclusion criteria were for the various clinical trials that contributed data to the Li *et al.*[6] analysis; or *(ii)* whether or not these may have resulted in any selection bias towards patients with clinical and imaging characteristics that were not representative of those from an unselected sample of patients with a similar range of scores on the EDSS. Furthermore, as noted previously, Li *et al.*[6] only studied patients with EDSS scores limited to the range of 0–6.5, which – as suggested by our bs-*r* findings – can lead to underestimating the true, overall relationship between EDSS scores and cr-T2-LL values in patients with MS. On the other hand, our sample of 110 patients with EDSS scores of 0–9.5 was representative of those seen in the MS Clinic of the Montreal Neurological Hospital and, we assume, representative of the untreated MS population in general.

The impact of increasing statistical power and of reducing many of the aforementioned causes of variability can be seen not only in our findings, but also in the results of a recent study by Fisniku *et al.*,[7] who studied the relationship between T2-LL and clinical disability in 74 patients who had presented with a clinically-isolated syndrome (**CIS**) suggestive of MS twenty years previously (of which 44 went on to have clinically-definite MS). They found evidence for a large, non-plateauing relationship between T2-LL and disability at 20-year follow-up, both as measured by: *(i)* scores on the EDSS ($\rho = 0.50$, $p < 0.001$; $n = 74$, with EDSS scores >= 6 in 27 cases); and *(ii)* scores on the Multiple Sclerosis Functional Composite scale ($\rho = -0.53$, $p < 0.001$; $n = 62$),[8, 9] which provides a quantitative measure of three key clinical dimensions of MS: ambulation, fine hand-motor coordination, and cognitive function. Importantly, in their study: *(i)* a reasonably-large sample size was studied, *(ii)* the entire range of EDSS-measured clinical disability was sampled, *(iii)* all of the MRI data was collected on the same scanner, *(iv)* all of the lesions were identified by one experienced neuroradiologist who was blinded to clinical details, and *(v)* T2-LL was quantified using a semiautomatic local-threshold contour technique[10] that had been proven to have high inter- and intra-rater reliability[11]

In conclusion, our findings have a number of important implications for future studies and clinical trials. First, they emphasize the importance of trying to minimize the effects of non-biological factors that influence variability in both the MRI measures and the clinical measures that are being studied: indeed, this is becoming more common in clinical trials[12] with the use of, for example: *(i)* central MRI-reading centers that are involved in the development of appropriate MRI protocols, as well as in the assurance that data from each of the contributing sites and scanners are acquired, processed, and analysed in a valid and reliable manner;[13] and *(ii)* the adoption of formal training and testing programs that ensure consistent clinical testing results from each of the contributing sites and testers (*e.g.*, the Neurostatus program, www.neurostatus.net). Second, they emphasize the importance of quantifying not only the patients' T2-LL (which is a relatively non-specific measure), but also their T1-LL (which seems to be more related to their degree of EDSS-measured clinical disability). Finally, the present findings point out the importance of: *(i)* studying patients with scores that span the entire range of the EDSS if the aim of a study is to understand what happens across the entire spectrum of disability that is found in MS, or *(ii)* remembering that findings in patients with EDSS scores of 0–6.0 are not necessarily representative of the entire course of the disease.

ONLINE SUPPLEMENTARY REFERENCES

1. Zijdenbos AP, Forghani R, Evans AC. Automatic "pipeline" analysis of 3-D MRI data for clinical trials: application to multiple sclerosis. *IEEE Trans Med Imaging.* Oct 2002;21(10):1280-1291.

2. Barkhof F. MRI in multiple sclerosis: correlation with expanded disability status scale (EDSS). *Mult Scler.* Aug 1999;5(4):283-286.

3. 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.* Mar 24 2006;59(4):597-605.

4. Zivadinov R, Leist TP. Clinical-Magnetic Resonance Imaging Correlations in Multiple Sclerosis. *J Neuroimaging.* Oct 2005;15(4_suppl):10S-21S.

5. Barkhof F, Karas GB, van Walderveen MA. T1 hypointensities and axonal loss. *Neuroimaging Clin N Am.* Nov 2000;10(4):739-752.

6. Li DK, Held U, Petkau J, et al. MRI T2 lesion burden in multiple sclerosis: a plateauing relationship with clinical disability. *Neurology.* May 9 2006;66(9):1384-1389.

7. Fisniku LK, Brex PA, Altmann DR, et al. Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. *Brain.* Mar 2008;131(Pt 3):808-817.

8. Fischer JS, Rudick RA, Cutter GR, Reingold SC. 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.* Aug 1999;5(4):244-250.

9. Rudick RA, Cutter G, Reingold S. The multiple sclerosis functional composite: a new clinical outcome measure for multiple sderosis trials. *Mult Scler.* Oct 2002;8(5):359-365.

10. Sailer M, O'Riordan JI, Thompson AJ, et al. Quantitative MRI in patients with clinically isolated syndromes suggestive of demyelination. *Neurology.* Feb 1999;52(3):599-606.

11. Grimaud J, Lai M, Thorpe J, et al. Quantification of MRI lesion load in multiple sclerosis: a comparison of three computer-assisted techniques. *Magn Reson Imaging.* 1996;14(5):495-505.

12. Hauser SL, Waubant E, Arnold DL, et al. B-cell depletion with rituximab in relapsing-remitting multiple sclerosis. *N Engl J Med.* Feb 14 2008;358(7):676-688.

13. Gedamu EL, Collins DL, Arnold DL. Automated quality control of brain MR images. *J Magn Reson Imaging.* Aug 2008;28(2):308-319.