Inverse dependence between patient population and correlation of composite MRI scores with EDSS in Multiple Sclerosis

A. H. Poonawalla[1], S. Datta[1], V. Juneja[1], F. Nelson[2], J. Wolinsky[2], G. Cutter[3], and P. Narayana[1]

[1]Diagnostic and Interventional Imaging, University of Texas Medical School at Houston, Houston, TX, United States, [2]Neurology, University of Texas Medical School at Houston, Houston, TX, United States, [3]Biostatistics, University of Alabama at Birmingham School of Public Health, Birmingham, AL, United States

INTRODUCTION. In multiple sclerosis (MS), patient disability is assessed according to the qualitative, ordinal Expanded Disability Status Scale (EDSS). In order to monitor treatment efficacy and disease progression, quantitative MRI metrics are sought that will correlate with EDSS [1]. In the relapsing-remitting MS (RRMS) literature, r-values for correlation with EDSS of various metrics have ranged from 0.31 – 0.61 [2-9], with a strong inverse trend with population sample size (Figure 1) implying that studies with fewer patients have biased correlation results. It is well-known that using small samples overestimates the true correlation, but a rigorous analysis would provide useful insight into whether some metrics are more resistant to the degrading effect of population size than others, what minimum population size is needed for robust results, and whether any fundamental limitations affect correlation with EDSS overall.

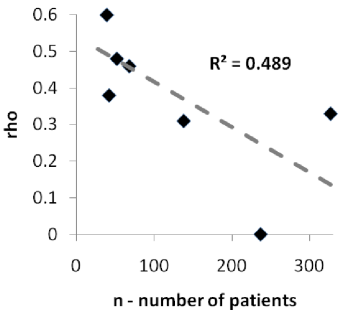

Figure 1. Reported correlations with EDSS of MRI metrics in RRMS patients follow a strong inverse trend with population sample size.

METHODS. 139 consenting RRMS patients (113 female, 26 male) with median age 43 years (20 – 64), and median EDSS score of 2.0 (0 – 6) were imaged on a 3T MRI scanner using a 6-channel SENSE head coil (Philips Medical Systems, Best, Netherlands). The acquisition included dual-echo FSE, FLAIR, and pre/post-Gd contrast T1-weighted images. All images had in-plane resolution of 0.94 mm x 0.94 mm and 44 contiguous axial slices of 3.0 mm thickness.

MRI images were segmented by an automated algorithm [10] to derive Gd-enhancing tissue volume (GD), total lesion volume "burden of disease" (BOD), lesion-only components of T2-hyperintense and black hole (BH) volumes, and normalized CSF volume fraction (nCSF). Voxel-wise T2 values were computed from the dual-echo FSE images as $(TE_2\text{-}TE_1) / \ln(S_1/S_2)$, where TE_1 & TE_2 are the echo times and S_1 & S_2 the respective signal intensities. Mean T2 values for the white-matter (WMT2), gray-matter (GMT2), black hole (BHT2), and lesion (LST2) compartments were then computed using the segmentation masks. All measures were standardized to dimensionless units of deviation from the mean using the z-transform $(x - u) / \sigma$, where u and σ are the mean and standard deviation of each metric over the patient population. Various composite metrics were constructed by combining these measures, and baseline correlations with EDSS (using the full data set of 139 patients) were calculated for each composite using Pearson's coefficient.

Numerical simulations were performed to evaluate the dependence of each composite metric's correlation with EDSS as a function of population size. For each of 1000 iterations, a random ordering of the n = 139 patients was generated, from which subsets of the first 25, 30, … 125 patients were successively extracted. From each subsample, the individual measures and composite metrics were calculated anew, and their correlation to EDSS determined as described above. Thus, a unique correlation vs. sample size history was derived for each iteration. The entire process was repeated for successive iterations, and then the mean and standard deviation of the correlation history across all 1000 iterations was calculated.

RESULTS. The numerical simulations confirmed the inverse dependence of the correlation with EDSS upon population size. As expected, the mean value of the simulated correlation with EDSS converged on the baseline correlation value as the number of subsamples approached the maximum value of n = 139 (Figure 2, below), with corresponding monotonic decrease in standard deviation. At the lowest values of n, the correlation was inflated for all scores, with a maximum of r = 0.55 which applied to all composite scores irrespective of the baseline performance. The population size threshold for convergence varied by composite score, with higher thresholds for weaker composites (n ~ 80, e.g. Figure 2a) and lower thresholds for stronger composites (n ~ 80. e.g. Figure 2d).

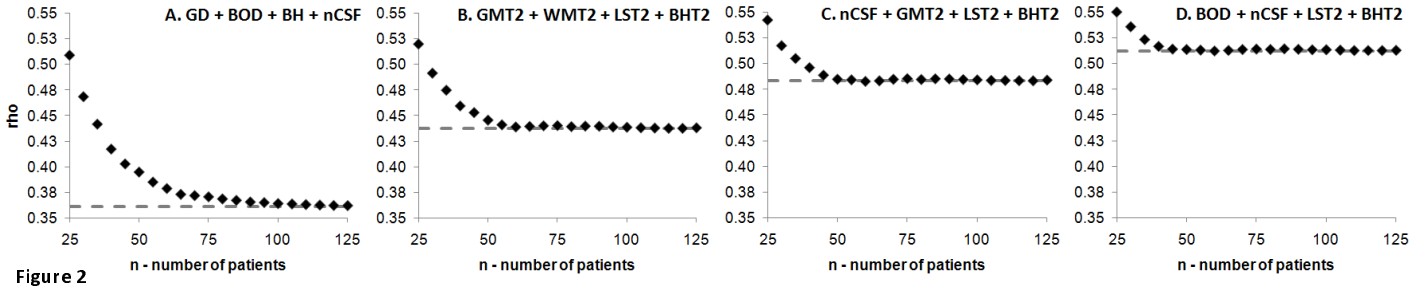

Figure 2

DISCUSSION. The choice of metric dramatically influences the degree of correlation bias, as well as baseline correlation value. However, the maximum observed correlation is independent of metric, and likely represents a "hard ceiling" to what correlation can be extracted from the data. Comparisons of correlations between studies must take population size into account, therefore, with the understanding that the reported correlation may be overstated for small studies (due to bias) and understated for large ones (due to metric). A threshold of 100 patients seems adequate to ensure that the effects of bias, at least, are minimized, irrespective of metric. The existence of a hard ceiling to correlation with EDSS suggests an inherent limit to any optimization efforts in choosing a metric. This may be partly due to the ordinal nature of EDSS, which does not necessarily typify the full range of disabilities a patient may have. Furthermore, practical limits (i.e., scan time) on the diversity of component measures used for constructing a composite metric may fundamentally constrain the characterization of the disease process.

REFERENCES. [1] Zivadinov R et al. J Neuroimaging. 2005;15:10S-21S. [2] Paolillo A et al. Eur J Neurol. 2002; 9: 645-55. [3] Morgen K et al. Mult Scler. 2001;7:167-171. [4] Riahi F et al. Brain. 1998;121:1305-1312. [5] Paolillo A et al. J Neurol Sci. 2000; 174: 85-91. [6] Zivadinov R et al. Neurology. 2001; 57: 1239-47. [7] Fisher E et al. Neurology. 2002;59:1412-1420. [8] Kappos L et al. Lancet. 1999; 353: 964-9. [9] Paty DW, Li DK. Neurology. 1993; 43: 662-7. [10] Datta S et al. Neuroimage. 2006;29:467-474.