ON-LINE APPENDIX A: PARTIAL LEAST-SQUARES REGRESSION DETAILS

The performance of the PLSR model with respect to generalizability to the rest of the population was assessed using leave-one-out or jackknife cross-validation.[1] More specifically, we removed 1 subject from the data and observation and calculated the PLSR model for a given number of components on the remaining data. We then calculated the prediction error for the data point that was removed and calculated the error, repeating this until all of the data points had been removed. The Predicted Residual Sum of Squares (PRESS)[2] for the model with that number of components was then the sum of all of the errors. One issue particular to PLSR is to how to choose the correct number of components to be included in the final model so that prediction is optimized without fitting noise (ie, overfitting). One way to do this is to observe the value of PRESS as you add components. Initially the value of PRESS decreases as the model includes more information about the data and the prediction becomes more accurate. As noise begins to affect model estimation, PRESS increases because the generalizability of each model decreases. Therefore, we chose the number of components for each model to be the value at which adding 1 more component caused an increase in PRESS.[2,3]

After we chose the number of components in the model, bootstrapping (resampling the data with replacement)[4] was used to assess the effect of population variation on the final results and to get a measure of stability for each of the regression coefficients. Bootstrapping in a population of this size is especially important to assess the effect of the particular characteristics of the population on the final model so that it has better generalizability to the whole population. Confidence intervals for the regression coefficients and the R^2 of the model fit were calculated using the bias-corrected and accelerated percentile method.[5] Confidence intervals were calculated to perform a hypothesis test—in this case, to test whether a regression coefficient was nonzero. A hypothesis test was performed for each of the 173 input variables (ChaCo in 86 brain regions, atrophy in 86 brain regions, and age) in each of the 2 WM biomarker-based models; the Šidák method[6] was used to correct for all tests. If the confidence interval of the regression coefficient of a particular variable did not include zero, then it was considered a significant predictor of the SDMT.

ON-LINE APPENDIX B: VARIABILITY OF ChaCo SCORES

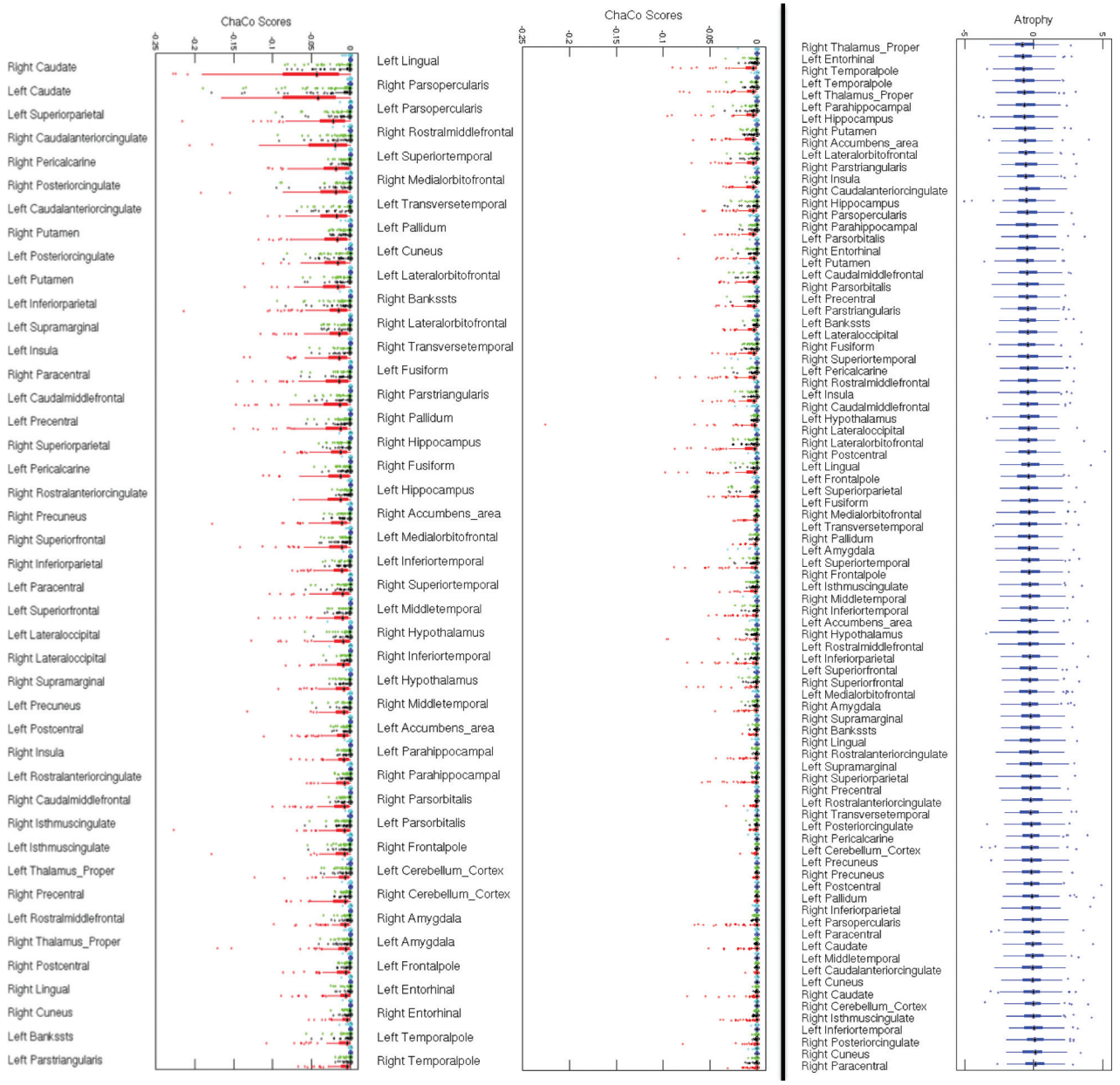
Some aspects of the reliability and validity of the ChaCo measure have been explored in our article that introduced the NeMo Tool (see the on-line supplemental information in Kuceyeski et al[7]). In that work, it was shown that the reliability of the scores evaluated in the space of the original tractography versus common Montreal Neurological Institute space was very high: ChaCo scores evaluated in native space versus Montreal Neurological Institute space had a Pearson correlation of 0.95. In addition, we removed certain well-known WM regions and calculated their ChaCo scores to show that the results were what we expected on the basis of normal anatomy.

However, the location and topology of WM tracts vary from individual to individual. Therefore, the impact of a given WM abnormality may also vary from individual to individual. The NeMo tool attempts to capture this variability by calculating the impact of WM abnormality across 73 healthy controls and reporting the average. This idea is not a new one: Any atlas-based method makes the same assumptions. It is important, however, to inspect the level of variability across healthy controls to get an idea of the confidence that can be assigned to the final ChaCo scores. Therefore, we calculated a measure of dispersion (ie, the quartile coefficient of dispersion $(Q_3 - Q_2)/(Q_3 + Q_2)$ where Q_i is the i^{th} percentile) of the T2 FLAIR–based ChaCo scores across healthy controls. On-line Fig 4 gives the distribution of dispersion measures over the 121 subjects with MS for each GM region. Values that are larger correspond to those regions that had greater dispersion in the healthy population and thus may correspond to less reliable ChaCo scores. In this figure, we only included regions with >1% of tracts disrupted (ChaCo < -0.01) because the dispersion measure is unreliable for values close to zero. Those regions with relatively high dispersion, such as the entorhinal and parahippocampal regions, were areas that, in general, were not important in our analysis.

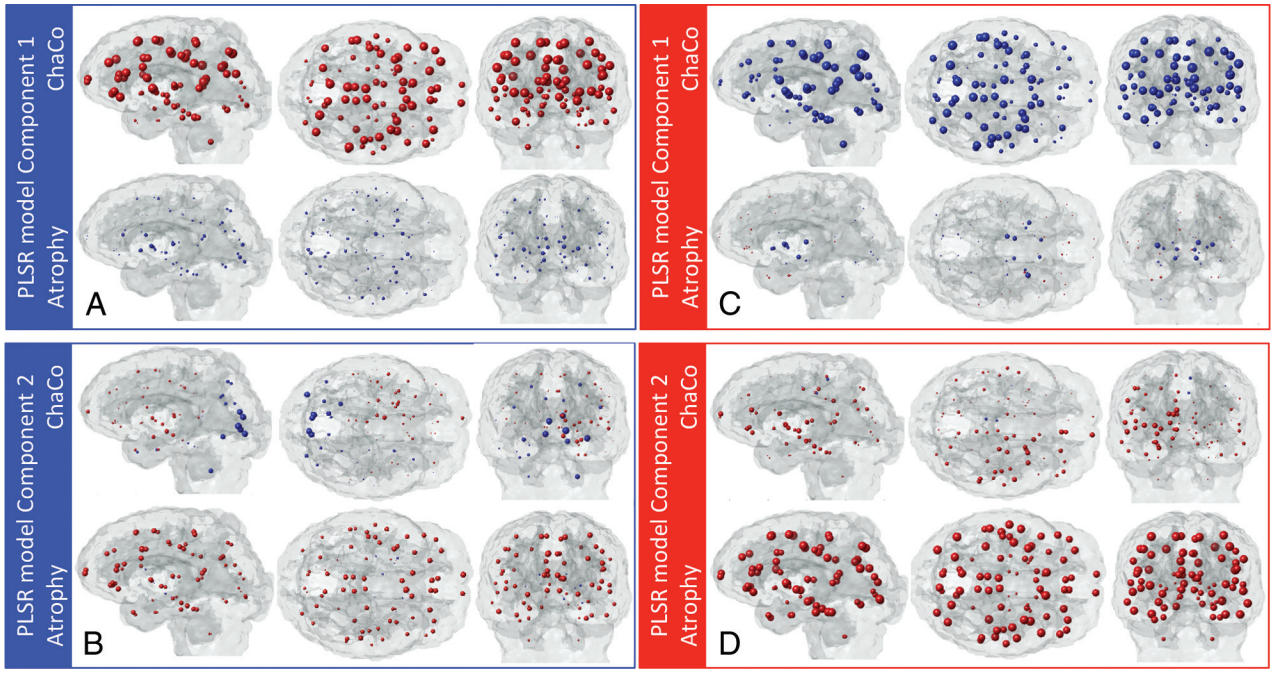
Finally, we measured the agreement of the ChaCo scores in the 73 healthy controls by calculating each subject's pair-wise Pearson correlation coefficient with the others (for a total of $[73 \times 72]/2 = 2628$ pairs) for each of the 121 subjects with MS. The histogram of all of the pair-wise values for all of the 121 subjects with MS (total of 2628×121 values) is given in On-line Fig 5, along with their averages over each of the subjects with MS (1×121 values) on the right. We see that there is a high level of agreement among the healthy controls, with the median pair-wise correlation being 0.86 and the median of the average pair-wise correlations over each subject with MS being 0.83 (denoted with red lines).

REFERENCES

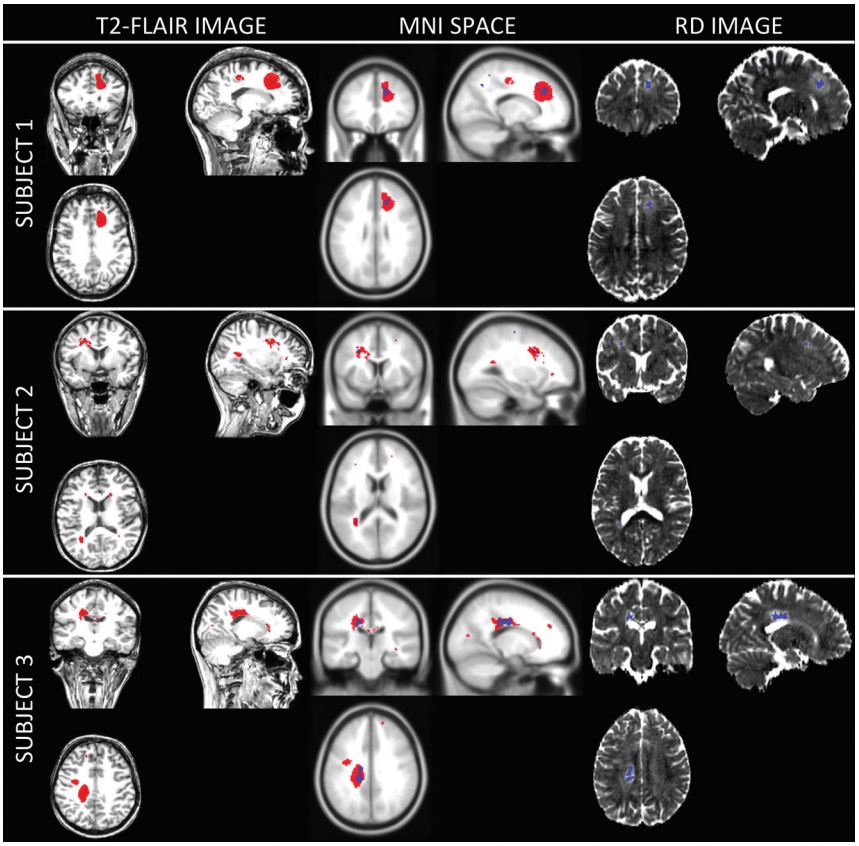
1. Wold S, Sjöström M, Eriksson L. **PLS-regression: a basic tool of chemometrics.** *Chemometrics and Intelligent Laboratory System* 2001; 58:109–30
2. Abdi H. **Partial least squares regression and projection on latent structure regression (PLS Regression).** *WIREs Comp Stat* 2010;2: 97–106
3. Abdi H, Dunlop JP, Williams LJ. **How to compute reliability estimates and display confidence and tolerance intervals for pattern classifiers using the bootstrap and 3-way multidimensional scaling (DISTATIS).** *Neuroimage* 2009;45:89–95
4. Efron B. *The Jackknife, the Bootstrap, and Other Resampling Plans.* Philadelphia: Society for Industrial and Applied Mathematics; 1982: 100
5. Efron B, Tibshirani R. *An Introduction to the Bootstrap.* London: Chapman and Hall/CRC; 1994:456
6. Šidák Z. **Rectangular confidence regions for the means of multivariate normal distributions.** *J Amer Statist Assn* 1967;62:626–33
7. Kuceyeski A, Maruta J, Relkin N, et al. **The Network Modification (NeMo) Tool: elucidating the effect of white matter integrity changes on cortical and subcortical structural connectivity.** *Brain Connect* 2013;3:451–63



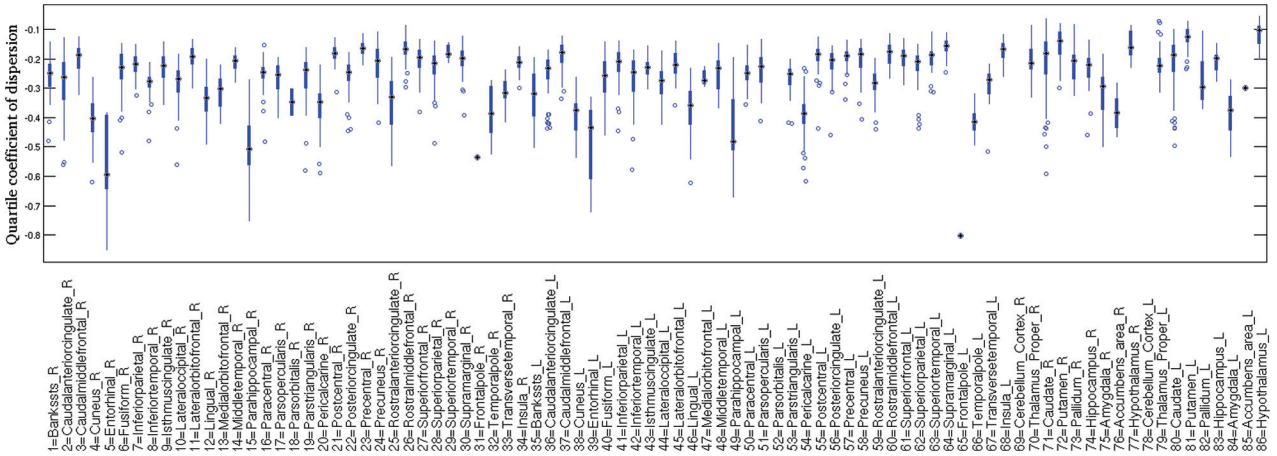
ON-LINE FIG 1. Barplots of ChaCo scores and atrophy. These barplots illustrate the distributions of ChaCo scores (*left 2 columns*) and atrophy (*right column*) over the population of 121 subjects with MS for each of the 86 GM regions considered. The ChaCo score plots are color-coded by the imaging modality that was used to produce the abnormality masks. Cyan indicates axial diffusivity maps; blue, fractional anisotropy maps; green, mean diffusivity maps; black, radial diffusivity maps; and red, T2 FLAIR hyperintensities.



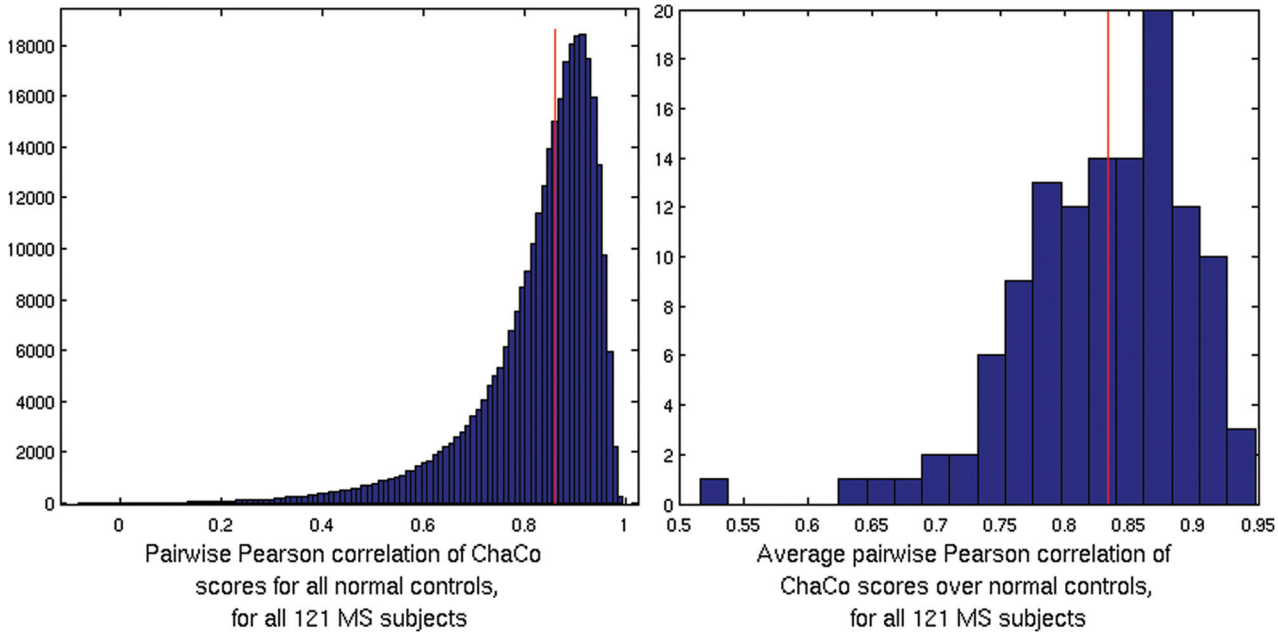
ON-LINE FIG 2. Partial least-squares regression components. These glass brain displays show the 2 components for each of the 2 models (*A* and *B*, T2 FLAIR–based ChaCo; *C* and *D*, RD abnormality–based ChaCo). Red values indicate negative component coefficients, and blue values indicate positive component coefficients; however, the actual sign is not important because PLSR components are invariant to change in sign. Model component coefficients indicate which variables in the input data are most important for a given component.



ON-LINE FIG 3. Assessment of coregistration and agreement of the T2 FLAIR– and RD-based abnormality masks. Here we see 3 examples of the subject's native space (*left column*) and Montreal Neurological Institute space (*middle row*) T2 FLAIR lesion masks. We also overlay the RD-based lesion masks in blue for comparison with the T2 FLAIR lesion masks. Because the image resolution and brain orientations are not identical across the different modalities/orientations, it is difficult to compare the locations directly on a section-by-section basis. However, we see that the Montreal Neurological Institute space masks respect the ventricular boundary and generally appear to be indicating the same general anatomic location. We also see that in most cases, the RD-based masks are either circumscribed by or in close proximity to the T2-FLAIR lesion masks. These locations increase confidence that our coregistration and lesion-creation algorithms are adequate.



ON-LINE FIG 4. Histograms of the quartile coefficient of dispersion for the ChaCo scores of the 121 subjects with MS across the 73 healthy controls in the NeMo Tool.



ON-LINE FIG 5. Histogram of the Pearson correlation of the ChaCo scores over each pair of healthy controls in the NeMo tool, for each of the 121 subjects with MS (*left*) and the average pair-wise Pearson correlation over healthy controls for all 121 subjects with MS (*right*). Median values are denoted with a vertical red line.