Disease correlates of rim lesions on quantitative susceptibility mapping in multiple sclerosis

Supplemental File

Melanie Marcille, BA[1], Sandra Hurtado Rúa, PhD[2], Charles Tyshkov, MD[3], Abhishek Jaywant, PhD[4], Joseph Comunale, MD[5], Ulrike W. Kaunzner, MD, PhD[1], Nancy Nealon, MD[1], Jai S. Perumal, MD[1], Lily Zexter, BA [1], Nicole Zinger, BS[1], Olivia Bruvik [1], Yi Wang, PhD [5], Elizabeth Sweeney, PhD[7], Amy Kuceyeski, PhD[5,6], Thanh D. Nguyen, PhD [5] and Susan A. Gauthier, DO, MPH [1,5,6]

[1] Department of Neurology, Weill Cornell Medicine, New York, NY, USA

[2]Department of Mathematics and Statistics, Cleveland State University, Cleveland, OH, USA

[3]Department of Pediatrics, Weill Cornell Medicine, New York, NY, USA

[4]Department of Psychiatry and Rehabilitation Medicine, Weill Cornell Medicine, New York, NY, USA

[5]Department of Radiology, Weil Cornell Medicine, New York, NY, USA

[6]Feil Family Brain and Mind Institute, Weill Cornell Medicine, New York, NY USA

[7]Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA

Corresponding Author:
Susan Gauthier, DO, MPH
Judith Jaffe Multiple Sclerosis Center
1305 York Avenue
New York 10021, NY
Tel: (646) 962-9800
Fax: (646) 962-0390
E-mail: sag2015@med.cornell.edu

Running title: Rim lesions associate with disability

Appendix 1. Multivariate regression model and MANOVA table

Multivariate regression is a technique that models more than one outcome variable in a single regression model framework. Our cross-sectional retrospective study has four cognitive outcome variables per patient (SDMT, CVLT-II, BVMT-R, and EDSS) as well as several covariates of interest (current treatment duration, Age, Gender, No.rim.p, and logTLV.flair). Instead of running four independent univariate regression analyses, we performed a multivariate analysis (see equation 1). The left-hand side of equation 1 denotes the vector of cognitive scores (multiple outcome variables) for patient i, and the right-hand side shows the linear model with two-way interactions. Equation 1 models the linear association between all outcomes (as a vector) and covariates of interest.

Note that our goal in estimating Equation 1 is to quantify the association between cognitive scores and the presence/absence of rim+ lesions while accounting for clinical patient-level covariates and overall lesion size. We did not include atrophy and cortical thinning measurements as we did not intend to model the marginal association (after accounting for atrophy and cortical thinning) between the number of rim+ lesions and cognitive scores.

Our primary goal is to estimate each vector of beta parameters ($\boldsymbol{\beta_j}$) as they quantify the association between each cognitive score and the number of rim+ lesions per patient (No.rim.p) while adjusting for other patient-level covariates such as logTLV.flair (x_2), sex (x_3), age (x_4), disease duration (x_5), and current treatment (x_6), as well as covariate interactions of the form x_{ji} No. rim. p_i, j=2,…6).

$$\begin{pmatrix} SDMT_i \\ CVLT_i \\ BVMT_i \\ log\,(EDSS_i + 0.1) \end{pmatrix} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}\, No.rim.p_i + \sum_{j=2}^{6} \boldsymbol{\beta_j}\, x_{ji} + \sum_{j=7}^{10} \boldsymbol{\beta_j}\, x_{ji}\, No.rim.p_i + \boldsymbol{\varepsilon_i}, \qquad \textit{(Eq. 1)}$$

Where each $\boldsymbol{\beta_j}$**; j=0,1,2,…,10** is a 1 × 4 vector of parameters quantifying the linear association between each covariate and each cognitive score; $\boldsymbol{\varepsilon_i}$ is the 1 × 4 vector of error terms, and it is

assumed to be multivariate normal with a mean zero vector and a covariance-variance matrix given by Σ; thus, the error terms associated with different outcomes may be correlated.

In this paragraph, we will discuss some of the advantages of multivariate regression models over univariate regression models (Johnson and Wichern, 2007). First, a multivariate regression model allows us to quantify the relationship between all four disability scores as outcomes and all patient-level covariates in one framework. Multivariate regression controls over the family-wise error rate and gives a more realistic modeling framework than looking at a single variable. It also provides a more powerful test of significance compared to univariate techniques. Lastly, many statistical software, including R, estimate the parameters in equation (1) as well as perform hypothesis testing (Fox and Weisberg, 2019).

Multiple Analysis of Variance (MANOVA) table extends the Analyses of Variances (ANOVA) table to studies with two or more related outcomes/dependent variables while controlling for the correlations among them. The MANOVA table displays a p-value based on a multivariate test statistic. We reported a p-value based on an F approximation of the Pillai test statistics for each parameter. Although the MANOVA p-values provide a more powerful test of significance compared to univariate techniques, we also reported the ANOVA tables, which are based on t-statistics and look at each score independently.

Appendix 2. Model Selection

We selected the final multivariate multiple regression model via model selection using the AIC and BIC selection criteria. In particular, we implemented the stepwise backward selection procedure by specifying the full model with all possible two-way interactions (see equation 1) and the lower bound model as the model without interactions (Cheng, 2017). All final models were selected using a stepwise backward procedure with a 0.10 significance level (Cheng, 2017).

Table 1 below displays the p-values from the Multivariate Analysis of Variance (MANOVA) table. The full multivariate model for SDMT, CVLT-II, BVMT-R, and EDSS as a vector of response variables included current Treatment Duration, Age, Gender, No. RIM+ lesions, log

T2wFLAIR lesion volume, and all two-way interaction terms between No. RIM+ lesions and other covariates.

	MANOVA (p-values)
(Intercept)	<0.001
Current Treatment Duration	0.026
Age	<0.001
Gender	0.040
logT2wFLAIR.lesion.volume	<0.001
No. RIM (0 rim+ versus 1+ rim+)	0.002
No. RIM * logT2wFLAIR.lesion.volume	0.013
No. RIM * Gender	0.115
No. RIM * Age	0.619
No. RIM * Current Treatment Duration	0.186

Table 1. Summary of p-values from the full Multivariate Analysis of Variance (MANOVA). The initial multivariate model for SDMT, CVLT-II, BVMT-R, and EDSS as a vector of response variables included the following covariates: Current Treatment Duration, Age, Gender, No. RIM+ lesions, log T2wFLAIR lesion volume, and all interaction terms between No. RIM and other covariates. This report gives the p-values associated with the MANOVA Pillai test.

Table 2 summarizes the p-values obtained from the multivariate analysis of variance (MANOVA) table. The final model (MANOVA) included a statistically significant interaction effect of rim+ lesions and total lesion volume on FLAIR imaging on disability outcome measures (p=0.006). Furthermore, the number of rim+ lesions (0 versus at least 1) and log-TLV were also significant (p=0.010 and p<0.001, respectively). Other statistically significant patient-level covariates were treatment duration, age, and gender with p-values of 0.012, <0.001, and 0.038, respectively.

	MANOVA (p-values)
	(p-values)
(Intercept)	<0.001
Current Treatment Duration	0.012
Age	<0.001
Gender	0.038
No. RIM (0 rim+ versus 1+ rim+)	0.010
logT2wFLAIR.lesion.volume	<0.001
No. RIM * logT2wFLAIR.lesion.volume	0.006

Table 2. Summary of p-values from the Multivariate Analysis of Variance (MANOVA) table and subsequent Analysis of Variance (ANOVA) tables for each score. The final multivariate model for SDMT, CVLT-II, BVMT-R, and EDSS as a vector of response variables included current Treatment Duration, Age, Gender, No. RIM of+ lesions, log T2wFLAIR lesion volume,

and the interaction term No. RIM * logT2wFLAIR.lesion.volume as covariates. This report gives the p-values associated with the MANOVA Pillai test and the approximated F-statistics for each score (ANOVA).

Appendix 3. List of most relevant R packages used in this analysis

A list of the R packages used is below:

- {stats} : Multivariate Model fitting
- {qtlmt}: Select a multivariate multiple regression model via model selection.
- {car}: Anova/MANOVA tables
- {effects}: Contrasts, effects
- {ggplot2, gridExtra, cowplot}: Plots

References

Fox, J., Weisberg, S., 2019. An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage.
Johnson, R., Wichern, D., 2007. Applied Multivariate Statistical Analysis. Upper Saddle River, NJ: Pearson Prentice-Hall.
Cheng, R., 2017. Tools for Mapping Multiple Complex Traits. R package version 0.1-6.