

Anna Markella Antoniadis^{1,2}, Miriam Galvin³, Mark Heverin³, Orla Hardiman^{1,3,4}, Catherine Mooney^{1,2}

1 FutureNeuro, the SFI Research Centre for Chronic and Rare Neurological Diseases;
2 School of Computer Science, University College Dublin, Belfield, Dublin 4, Ireland;
3 Academic Unit of Neurology, Trinity Biomedical Sciences Institute, Trinity College Dublin, Dublin 2, Ireland;
4 Department of Neurology, National Neuroscience Centre, Beaumont Hospital, Dublin 9, Ireland;

1. Introduction

Amyotrophic Lateral Sclerosis (ALS) is a fatal neurodegenerative disease that leads to the progressive death of motor neurons. There is no known cure for the disease, so treatment is focused on alleviating symptoms and improving quality of life (QoL).

Aim: to identify patient and caregiver factors that are related to a patient's quality of life.
Purpose: (a) knowledge of the features can have an impact on the healthcare and support that a patient receives,
(b) a subset of these features can be used to develop a Clinical Decision Support System that will alert clinicians on a patient's risk of low QoL.

Metrics: Two overall QoL scores from the McGill QoL Questionnaire[1] for comparison, namely MQoL and MQoL-SIS. The scores were split in two classes to create a binary problem.

2. Data

Source

- Irish ALS Register and questionnaires on patients and their primary caregivers (demographics, socio-economic, resource use, QoL, anxiety and depression, burden) collected at 3 time-points
- 90 patient-caregiver dyads at baseline

Outcomes split in two classes based on the Metric's median value:
MQoL: calculated total QoL from sub-scores [1]
MQoL-SIS: Single Item Score

Size

- 167 and 176 entries for MQoL and MQoL-SIS, 136 features

3. Methods

- Ensemble Feature Selection (EFS) [2]:** Ensemble of six different methods for the elimination of biases: median, Pearson- and Spearman-correlation, logistic regression (LR), and two variable importance measures embedded in the "randomForest" (RF) [3] implementation in R
- Evaluation:** ROC Curves comparison and Student's t-test to compare the created model (LR or RF) with a model that uses a permutation of the class variable

4. Results

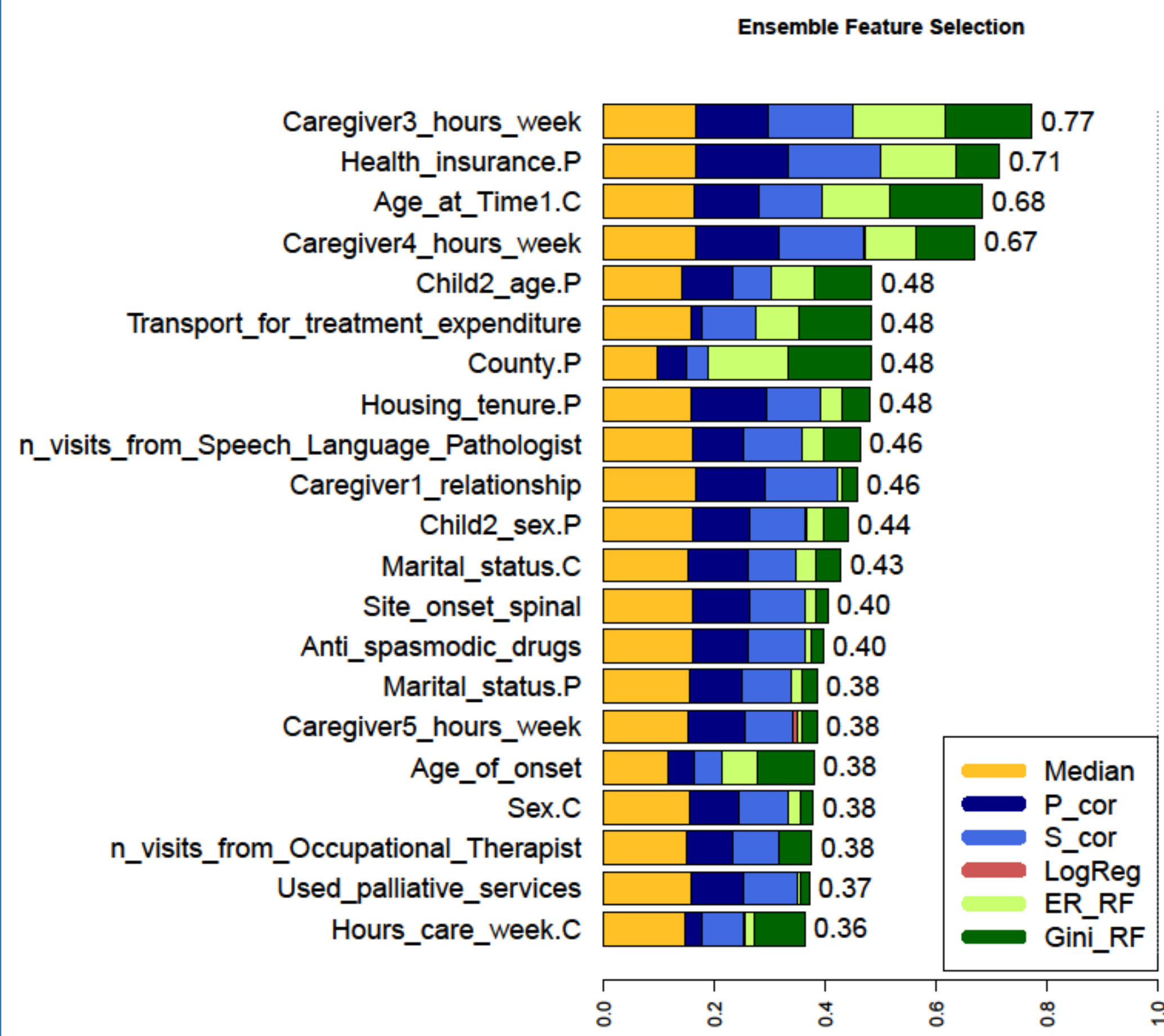


Fig. 1. The 20 most important features for the prediction of MQoL-SIS.

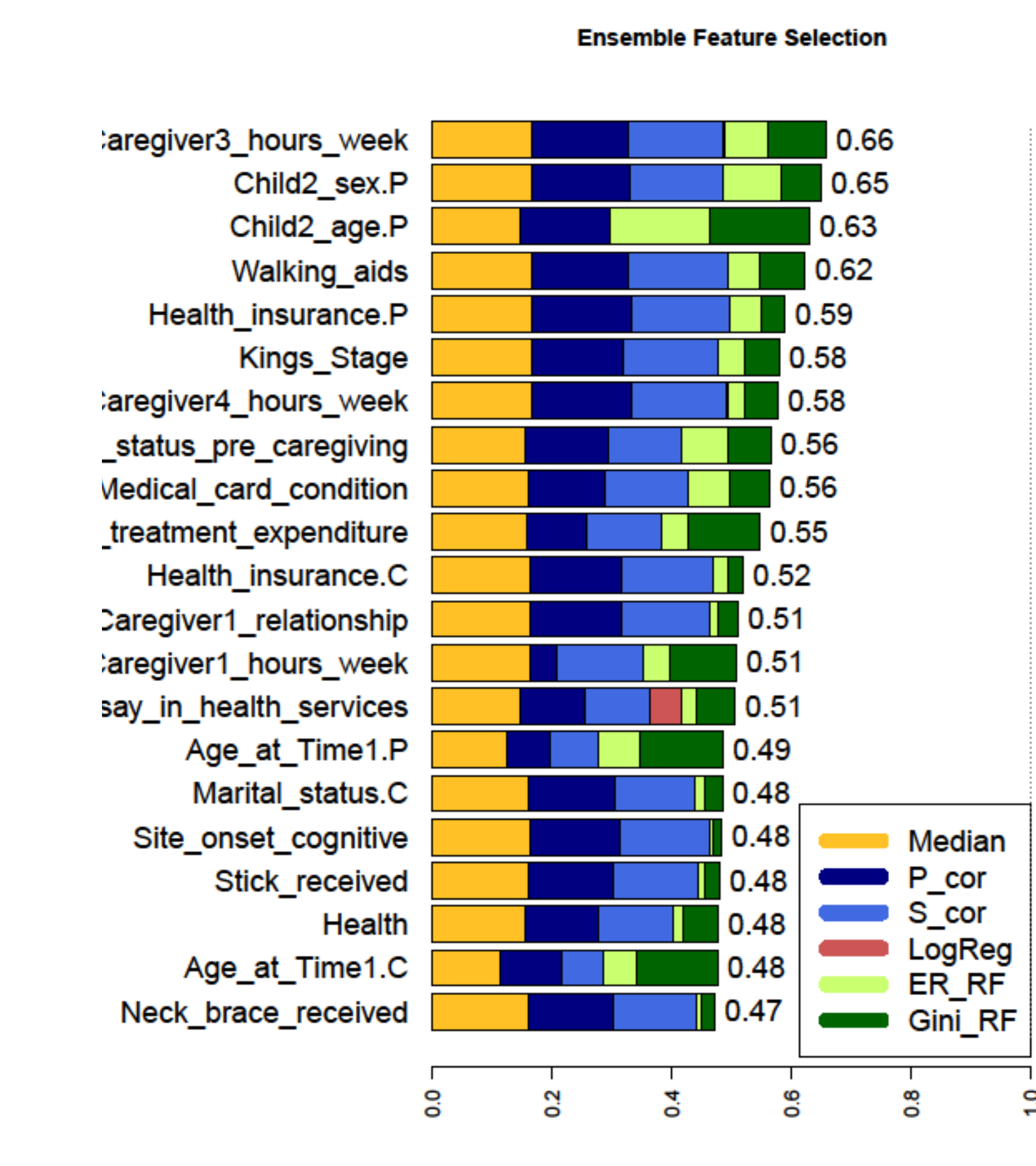


Fig. 2. The 20 most important features for the prediction of MQoL.

Evaluation: The RF and permutation tests in both QoL scores are significant at the $\alpha=0.05$ level of significance, but the LR-ROC tests are not. (Table 1)

Table 1. The evaluation metrics for the feature selection of the most predictive features of the two QoL outcomes.

Metrics	MQoL-SIS	MQoL
AUC of LR	0.67	0.64
AUC of RF	0.71	0.75
P-value of LG-ROC test (vs LG with all features)	0.35	0.09
P-value of RF-ROC test	<0.001	0.011
P-value of permutation test	<0.001	<0.001

5. Conclusion

- Hours of caregiving duties** by informal caregivers predictive of both QoL scores. Also, the patient's health insurance, age of primary caregiver and information related to children.
- Use of aids and appliances, stage of disease** according to the King's Staging System [4] highly predictive of the MQoL score.
- Use of healthcare services** highly predictive of the MQoL-SIS.

5. References

[1] S. R. Cohen, "McGill Quality of Life Questionnaire," in: Michalos A.C. (eds) Encyclopedia of Quality of Life and Well-Being Research. Springer, Dordrecht, 2014.
[2] U. Neumann, M. Riemenschneider, J. Sowa, T. Baars, J. K'alsch, et al., "Compensation of feature selection biases accompanied with improved predictive performance for binary classification by using a novel ensemble feature selection approach," BioData Mining, vol. 9(36), pp. 1-14, 2016.
[3] L. Breiman, "Random forests," Machine Learning, vol 45(1), pp. 5-32, 2001.
[4] C. Roche, R. Rojas-Garcia, K. M. Scott, W. Scotton, C. E. Ellis, et al., "A proposed staging system for amyotrophic lateral sclerosis," Brain, vol 135(3), pp. 847-52, 2012.