Estimation of Disability Status from Blood Serum Protein Concentrations

T. Chitnis[1] , J. Foley[5] , C. Ionete[2] , N. El Ayoubi[3], S. Saxena[1], P. Gaitan-Walsh[1], H. Lokhande[1], A. Paul[1], F. Saleh[1], H. Weiner[1], T. Hoyt[5], F. Qureshi[4], F. Zhang[4], F. Rubio da Costa[4], V. M. Gehman[4], S. J. Khoury[3]

[1]Brigham and Women's Hospital, Harvard Medical School, Boston, MA; [2]University of Massachusetts Medical School, Worcester, MA; [3]American University of Beirut, Beirut, Lebanon; [4]Octave Bioscience, Menlo Park, CA; [5]Rocky Mountain Multiple Sclerosis Clinic, Salt Lake City, UT

Background

The evaluation of multiple sclerosis (MS) disability status (DS) currently relies on qualitative assessments of radiographic and clinical evidence. Expanded Disability Status Scale (EDSS) [1] and Patient Determined Disease Steps (PDDS) [2] are two broadly adopted semi-quantitative tools for describing disability status of MS patients.

Objectives

Explore development of a blood-based multiplex proteomic test associated with EDSS and PDDS using serum samples obtained from 4 sites: Brigham and Women's Hospital (BWH - CLIMB Study), University of Massachusetts (UMASS - FSDD Study), American University of Beirut (AUB), and Rocky Mountain Multiple Sclerosis Clinic (RMMSC).

Cohort

- 595 serum samples with disability status scores were assayed in a custom proteomic immunoassay panel.
- The cohort and the assay panel were optimized for the assessment of MS disease activity (DA). The DS analysis presented in this report represents a secondary, exploratory study.
- *After correction for age and sex*, the data were used to build models for both regression of disability score and classification of samples as having either mild/moderate or severe disability.
- The threshold for severe disability was taken as each score's association with the ability to walk unassisted (EDSS = 5.0, PDDS = 4.0)[1, 2].

Tab. 1: Cohort summary table, split by disability status.

		Whole Cohort	Mild/Moderate	Severe
Cohort	Sample Count	595 (100%)	541 (90.9%)	54 (9.1%)
	Sex [% Female]	70.2	74.1	70.6
	Age [$\mu \pm \sigma$]	41.9 ± 13.0	40.6 ± 12.2	55.4 ± 12.3
Site	AUB	33.6% (200)	36.2% (196)	7.4% (4)
	BWH	30.4% (181)	31.2% (169)	22.2% (12)
	RMMSC	28.2% (168)	25.1% (136)	59.3% (32)
	UMASS	7.7% (46)	7.4% (40)	11.1% (6)
DMT Class	Anti-CD20	6.9% (41)	6.7% (36)	9.3% (5)
	Dimethyl Fumarate	10.6% (63)	10.9% (59)	7.4% (4)
	Fingolimod	12.8% (76)	13.9% (75)	1.9% (1)
	Glatiramer Acetate	10.3% (61)	10.4% (56)	9.3% (5)
	Interferons	15.0% (89)	15.9% (86)	5.6% (3)
	Natalizumab	21.7% (129)	19.8% (107)	40.7% (22)
	Other	5.4% (29)	5.4% (29)	5.6% (3)
	Unknown	17.5% (93)	17.2% (93)	20.4% (11)

Analysis: Univariate

Methods:

- Classification models were trained against a common binary endpoint using the definition of severe disability defined above. For regression, we mapped PDDS into "effective EDSS" [3], for a combined analysis of both endpoints.
- Univariate associations between measured protein concentrations in the MSDA test assay panel for classification and regression endpoints.

Results:

- Calculated AUROC or Spearman's correlation coefficient and generated nonparametric p-values (by counting the number of times the signal data was exceeded by randomly generated/null hypothesis data).
- We found the following statistically significant proteins ($p \leq 0.05$):
  - Classification: GFAP, CDCP1, TNFRSF10A
  - Regression: GFAP, FLRT2, CXCL9, CXCL13, CCL20, APLP1, IL-12B, TNFRSF10A, OPG

Analysis: Multivariate

Feature selection methods:

- All significant univariate features were included in the optimized feature set.
- Selected more features for each analysis using greedy forward selection.
- Maximized area under the receiver operator characteristic curve (AUROC) and R^2 between predicted and effective EDSS.
- Train/test split (two thirds/one third) to control overfitting, performed 100 bootstrap simulations, averaging over all splits to minimize the effect of outlier samples.
- Forward selection search output shows the maximized metric as the most performant next protein is added to the feature set.

Feature selection results:

- Models include the protein at that tick and all the ones to the left.
- Shaded region around the traces from variation across the train/test splits.
- Also plot fractional change in optimized metric as proteins are added to the feature set (1% fractional change marked with a dashed gray line).
- For each analysis (classification and regression), we took two feature sets:
  - Maximum performance metric ("Best"),
  - the "point of diminishing returns" ("PoDR") where the forward selection curve "levels off" (PoDR formally defined as the smallest feature set where performance is within one standard deviation of the neighborhood around the optimal model size).
- Selected proteins for each analysis (with the PoDR features in bold):
  - Classification: **GFAP, CDCP1, CXCL13**, VCAN, APLP1
  - Regression: **CXCL13, GFAP, APLP1, FLRT2, CXCL9, TNFRSF10A, CCL20, CDCP1**, GH

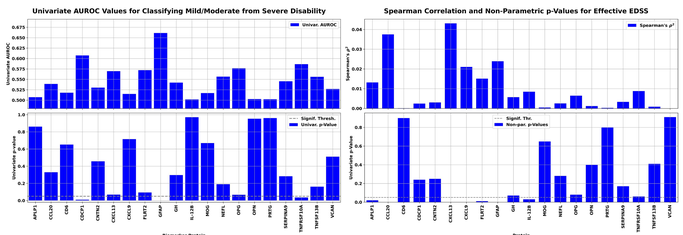


Fig. 1: Univariate performance for classification of disability status (left) and regression of effective EDSS for each protein as specified by optimized metric (top panel of both sides, AUROC for classification and Spearman's ρ^2 for regression) and non-parametric p-value (bottom panel of both sides).

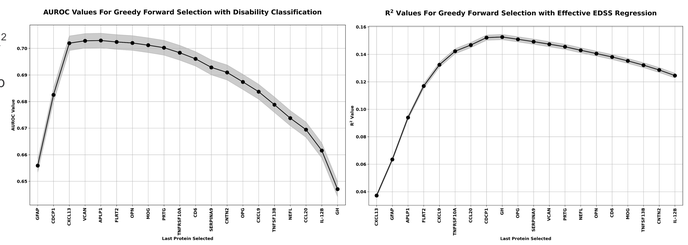


Fig. 2: Forward selection output for classification (left) and regression (right).

T. Chitnis has received compensation for consulting from Biogen, Novartis Pharmaceuticals, Roche Genentech, and Sanofi Genzyme. She has received research support from the National Institutes of Health, National MS Society, US Department of Defense, EMD Serono, I-Mab Biopharma, Mallinckrodt ARD, Novartis Pharmaceuticals, Octave Bioscience, Roche Genentech, and Tiziana Life Sciences. This research was conducted in part with the support of the Department of Defense through the Multiple Sclerosis Research Program under Award No. W81XWH-18-1-0648 (to T. Chitnis). C. Ionete has received research support from Biogen, Serono, Genentech, NMSS, and Department of Defense. She received compensation for advisory board activity from Sanofi-Genzyme. N. Ayoubi has received support to attend scientific educational courses from the following companies: Novartis, Merck Serono, Sanofi, Biologix. Has received speaker honoraria for scientific presentations on Multiple Sclerosis from the following companies: Biologix, Sanofi, Merck Serono, Novartis. S.i Saxena, P. Gaitan-Walsh, A. Paul, F. Saleh and T. Hoyt have no disclosures. H. Lokhande has received research support from the US Department of Defense and Octave Bioscience. H. Weiner has received research support from the Department of Defense, Genentech, Inc., National Institutes of Health, National Multiple Sclerosis Society, Novartis and Sanofi Genzyme. He has received compensation for consulting from Genentech, Inc, IM Therapeutics, IMAB Biopharma, MedDay Pharmaceuticals, Tiziana Life Sciences and vTvTherapeutics. S. Khoury has received compensation for scientific advisory board activity from Merck and Roche. J. Foley has received research support from Biogen, Novartis, Adamas, Octave, Genentech, and Mallinckrodt. He received speakers' honoraria and acted as a consultant for EMD Serono, Genzyme, Novartis, Biogen, and Genentech. He has equity interest in Octave. He is the founder of InterPro Biosciences. F. Qureshi, M. Becich, F. Zhang, F. Rubio da Costa, and V. Gehman are employees of Octave Bioscience.

Analysis: Multivariate (cont.)

Model building methods:

Classification and regression analyses proceeded along the following parallel path:

1. Optimal baseline logistic regression (classification) and ridge regression (regression) model configuration was chosen using a grid search through reasonable parameter space using only the proteins as features.
2. Age and sex then added as features (disease duration not added because of start date ambiguity in MS, and its correlation with age).
3. Trained separate versions of the baseline model for men and women.
4. Used Catboost [4] (a more sophisticated model architecture, better suited for categorical variables) for proteins, age, and sex.
5. Finally, we included disease modifying therapy (DMT) class (from Tab. 1) to the model to examine the importance of DMT information.

Model building results:

- All results compared to a baseline model using only age and sex to predict disability status or effective EDSS.
- Comparison is necessary because disability naturally increases with age–need to quantify added benefit from protein information.
- Feature importance allows us to check the degree to which each feature contributes to the model.
- Feature importance values shown in Fig. 4 are non-parametric and estimated from the error induced by permuting each column of the feature matrix individually ten times and then averaging the resulting the drop in performance (either AUROC or R^2) [5, 6]. Features not included in a particular model or feature set have importance set to zero in Fig. 4.

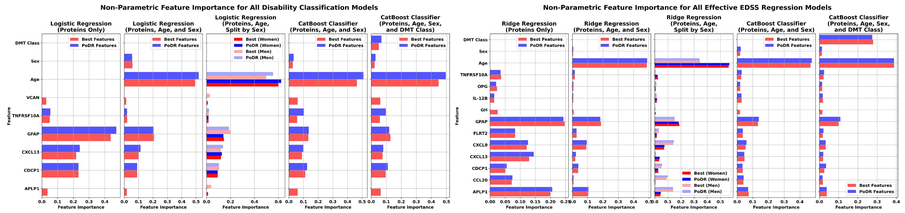
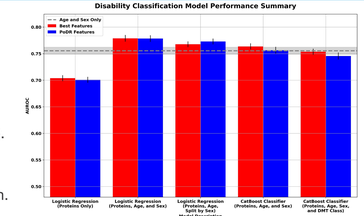


Fig. 4: Non-parametric feature importance (from permutation-induced error) for all models and feature sets. Classification features are shown on the left, regression features on the right. The "best" features have importances drawn in red bars, with PoDR features in blue. For the one study where men and women were treated in separate models, the models for men are drawn in a lighter shade than women. Features not contained in a particular model have their importance set to zero.



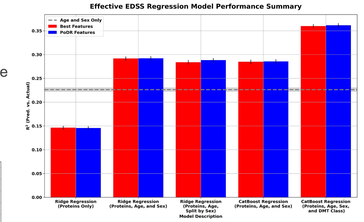


Fig. 3: Model performance for classification (top) and regression (bottom) models for both feature sets. The "best" feature set is shown in red, while the PoDR feature set is shown in blue. In both panels, the age and sex only model is shown as a grey band.

Conclusions

- We see promising initial results for the correlation of several proteins with disability as measured by EDSS/PDDS, key indicators of MS disability status.
- We see a strong dependence of all models on patient age wherever that feature is allowed to contribute to a model. When included in a model, age is the most important feature.
- Demographically corrected protein concentrations alone show inferior performance compared to age and sex alone.
- Protein concentration with age and sex are among the most performant feature sets in this study. Most pronounced with regression, some effect in classification as well.
- DMT class adds little information to classification models, but greatly improves regression. Difference is reflected in the DMT class importance values too.
- Little performance difference between best and PoDR features. Most "statistical work" being done by proteins chosen early in the forward selection (note the flatness of forward selection curves near global maximum).
- GFAP is the most important protein feature. APLP1 is nearly as important for regression, hardly impacts classification. CXCL13 and CDCP1 are also important for both.
- Catboost classification and regression yields little difference in performance but seems to spread importance more evenly across all features.
- When we separate models for men and women there is some interesting difference between sexes in feature importance. Models for women tend to lean more on age, while those for men tend to have a more even distribution across a greater number of features.
- Recall that the patient cohort was not optimized for this endpoint. This study builds on a secondary endpoint for a study using Gd-enhancing lesions to track MS disease activity.
- Additionally, EDSS and PDDS are both quite focused on mobility, making them an incomplete measure of disability in MS patients.
- Plans for future investigations of DS include:
  - Other endpoints, particularly radiographic information and different measures of cognition.
  - Longitudinal/dynamical analysis of patient histories to better account for patient to patient variation.
  - Predictive modeling to assess early in a patient's journey whether they are likely to show high levels of DS or not.

References:

[1] J.F. Kurtzke. "Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS)." Neurology, 33 (11) 1444; DOI:10.1212/WNL.33.11.1444

[2] M.J. Hohol, E.J. Orav, H.L. Weiner. "Disease steps in multiple sclerosis: a simple approach to evaluate disease progression." Neurology. 1995 Feb;45(2):251-5. DOI: 10.1212/wnl.45.2.251. PMID: 7854521.

[3] Y.C. Learmonth, R.W. Motl, B.M. Sandroff, et al. "Validation of patient determined disease steps (PDDS) scale scores in persons with multiple sclerosis." BMC Neurol 13, 37 (2013). DOI:10.1186/1471-2377-13-37

[4] L. Prokhorenkova, G. Gusev, A. Vorobev, A.V. Dorogush, and A. Gulin. "CatBoost: Unbiased Boosting with Categorical Features." Proceedings of the 32nd International Conference on Neural Information Processing Systems. (2018) pp. 6639–6649. DOI:10.5555/3327757.3327770

[5] T. Parr, K. Turgutlu, C. Csiszar, and J. Howard. "Beware Default Random Forest Importances." (2018) https://explained.ai/rf-importance/index.html

[6] C. Strobl, A.L. Boulesteix, T. Kneib, T. Augustin, A. Zeileis. "Conditional variable importance for random forests." BMC Bioinformatics. (2008) 9, 307. DOI:10.1186/1471-2105-9-307. PMID: 18620558; PMCID: PMC2491635.

For questions, please contact Victor Gehman: vgehman@octavebio.com

Presented at the 7th annual Americas Committee for Treatment and Research in Multiple Sclerosis (ACTRIMS) Forum, February 24-26, 2022, West Palm Beach, Florida