A Machine Learning and Statistical Meta-Analysis for Trial Simulations Predicting Transitions from Relapsing-Remitting to Secondary Progressive Multiple Sclerosis

Matthew Wiens[1], Kyle Barrett[1], Amir-Hadi Maghzi[2], Jonathan French[1], Himanshu Naik[3*], Jackson Burton[2]
[1]*Metrum Research Group, Tariffville, CT, USA*, [2]*Biogen Inc., Cambridge, MA, USA*, [3]*Sumitovant Biopharma, New York, NY, USA*
[*] *At Biogen Inc. while contributing to this work*

Introduction

- **Multiple sclerosis** (MS) is an autoimmune disease characterized by distinct episodes of acute neurological worsening (relapse) and formation of brain lesions seen on MRIs.
- Individuals progress from a relapsing-remitting (RRMS) state to a secondary progressive (SPMS) state. The vast majority of current treatments target RRMS and have minimal effect against SPMS [1].
- Therapeutics targeting SPMS and transitions to SPMS is an emerging clinical area of interest, but there is currently limited modeling and simulation to support decision making. Furthermore, SPMS is often diagnosed retrospectively, and there is ongoing research into identifying SPMS in a without relying on a multi-year clinical history [2].
  - An EDSS (Expanded Disability Status Scale) score of 4 or more is a key part of the Lorscheider definition
  - No relapses is also a clinically important marker of SPMS
- The objective of this analysis was to **identify patients likely to transition to SPMS**, as defined by Lorscheider [2] based on baseline patient characteristics using both a **machine learning** (ML) classifier and a **statistical survival model**.
- The models were used to simulate clinical trial power based on varying inclusion/exclusion criteria to improve trial design for potential SPMS trials.

Methods

A gradient-boosted tree machine learning model (xgboost) and a parametric accelerated failure time (AFT) survival model were fit to data from five studies and their long-term extensions to predict SPMS transitions, and then used to simulate the time-to-SPMS transition. Covariates were selected based upon clinical intuition and Shapley values (metric of covariate importance) from the machine learning model. Clinical trials were simulated using both models with specified inclusion/exclusion criteria. The impact of study enrichment through these inclusion/exclusion criteria was explored by comparing sample sizes for fixed levels of power and type 1 error.

Gradient Boosted Trees

- Gradient boosted trees build trees sequentially, each partially correcting the errors of the previous trees
- Predictions are the mean of all the trees
- Unlike random forests, only tens of trees are typically constructed, but not in parallel (nor are the trees all IID)
- Tree-based methods do not suffer from instability with highly correlated features (covariates)
- Simple survival models are available in the xgboost package

Key Machine Learning Concepts and Workflow

- **Class imbalance**: Most patients did not progress to SPMS, so the weight of each observation with a transition to SPMS in the loss function (objective function) was upweighted to make the total weights of each class equal. This ensured the model focused on accuracy patients transitioning to SPMS.
- **ML Model Hyperparameter tuning**: Cross-validation using a grid (max-entropy) was used to optimize hyperparameters, which was important to control overfitting in the final boosted tree. Default values performed much worse than optimized values.
- **Precision-recall AUC**: When evaluating the model, the commonly used ROC (sensitivity vs. specificity) curve was likely misleading because of the class imbalance, and because predicting patients who transitioned to SPMS was much more important than identifying the patients who did not transition. The precision-recall (sensitivity) curve focuses on prediction of the class of interest.
- **Causal interpretation (Figure 3)**: Shapley values and causal graphs were used to understand treatment effects. Treatments which had a greater effect of suppressing relapses were associated with greater rates of transitions to SPMS, likely because relapses masked underlying slow transitions, both clinically and in the Lorscheider definition. This justified not including treatment as a feature, and in fact suggested we should *not* include treatment as a predictor because it could hide effects of other features.
- **Isotonic regression**: Predictions from a model using weighted observations did not correspond to actual probabilities. Therefore, an isotonic regression where a piecewise non-decreasing function was estimated to transform model predictions to probabilities for simulation [3].
- **Shapley values** (Figure 6): Feature importance was assessed through Shapley values, and informed understanding of particular covariate ranges which had particularly large effects [4]. Shapley values assess the importance of each feature for each patient by comparing the effect of using the actual feature value compared to the mean feature value for different sets of features. For tree methods and linear models, exact Shapley values can be computed efficiently.

Figure 1: Precision-recall curve

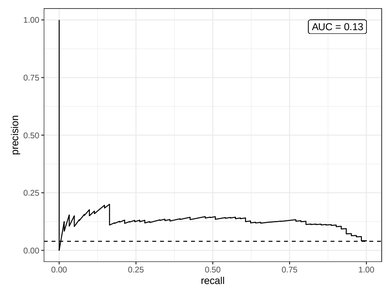


Figure 2: Estimated isotonic regression

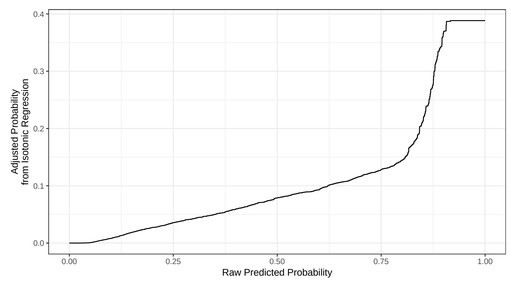
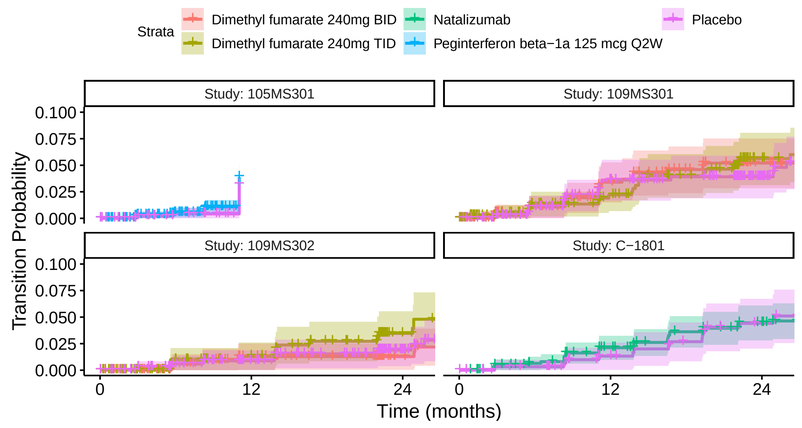


Figure 3: Causal inference: treatment association with transitions to SPMS. Causal mediation likely explains the observed relationship; treatment directly causing an increase in SPMS was assumed to be highly implausible.



Conclusion

- We pooled data from 5 trials and associated long-term extension studies in patients with RRMS for modeling and simulation.
- EDSS and its subscores are the most important factors for predicting progression to SPMS.
- Trends of power and trial successes were similar for time to event model and binary models, although time to event models and analyses had higher power, reflecting incorporation of additional data into the parametric model compared to the machine learning model.
- This workflow demonstrates a complementary approach for leveraging both ML and parametric models for exploratory and predictive modeling.
  - Limited ML survival models are available with the xgboost package, but parametric models allowed for much more rapid development of more complex survival models
  - ML models gave a performance benchmark for parametric models and helped us understand where parametric models could be refined (e.g. choices of hazard function)
  - Using multiple approaches increased confidence in our understanding important features and reliability of predictions
- Many of the ML techniques used in this analysis are applicable to the pharmacometrics community and address challenges when implementing an ML workflow with clinical data, but are not widely discussed in the ML literature or ML tutorials.

References

[1] Ontaneda, D., Fox, R.J. and Chataway, J. Clinical trials in progressive multiple sclerosis: lessons learned and future perspectives. *Lancet Neurol.* **14** (2015):208–223.

[2] Lorscheider, J., Buzzard, K., Jokubaitis, V., Spelman, T., Havrdova, E., Horakova, D., Trojano, M., Izquierdo, G., Girard, M., Duquette, P., Prat, A., Lugaresi, A., Grand'Maison, F., Grammond, P., Hupperts, R., Alroughani, R., Sola, P., Boz, C., Pucci, E., Lechner-Scott, J., Bergamaschi, R., Oreja-Guevara, C., Iuliano, G., Van Pesch, V., Granella, F., Ramo-Tello, C., Spitaleri, D., Petersen, T., Slee, M., Verheul, F., Ampapa, R., Amato, M.P., McCombe, P., Vucic, S., Sánchez Menoyo, J.L., Cristiano, E., Barnett, M.H., Hodgkinson, S., Olascoaga, J., Saladino, M.L., Gray, O., Shaw, C., Moore, F., Butzkueven, H., Kalincik, T. and MSBase Study Group. Defining secondary progressive multiple sclerosis. *Brain* **139** (2016):2395–2405.

[3] Niculescu-Mizil, A. and Caruana, R.A. Obtaining calibrated probabilities from boosting (2012).

[4] Chen, T. and Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16 (ACM, New York, NY, USA, 2016), pages 785–794. URL `http://doi.acm.org/10.1145/2939672.2939785`

Results

The analysis dataset comprised of 6298 patients, 518 (8.2%) of whom transitioned to SPMS during the study and 228 (3.6%) of whom transitioned to SPMS within 2 years. In comparison, the ML model identified a possible trial population of 20% patients who transitioned to SPMS within 2 years, corresponding to a 5.5-fold population enrichment, while maintaining a sensitivity of 15%. EDSS score and the Pyramidal subscore were important features as assessed by Shapley values. Trial simulations showed that without patient population enrichment, trials in similar populations would require large trial sizes and large treatment effects to have power greater than 0.8.

A **Rshiny application** was built for stakeholders of broad background, enabling real-time trial design and population enrichment simulations.

Data Summary

Parent Study Identifier	N	Number of patients progressing to SPMS	Time in Study (years)	Percent with duration > 2 years	Number of EDSS observations after baseline
105MS301	1512	50	3.07 (1.39)	72.4%	13.4 (5.09)
109MS301	1217	78	5.04 (3.74)	63.6%	16.5 (9.65)
109MS302	1410	87	4.95 (3.61)	62.8%	16.5 (9.46)
C-1801	939	106	7.31 (4.09)	93.1%	23.5 (9.79)
C-1802	1196	97	4.29 (3.00)	86.7%	16.5 (7.68)

	Parent Study Identifier					
	105MS301 n = 1512	109MS301 n = 1234	109MS302 n = 1417	C-1801 n = 939	C-1802 n = 1196	Summary n = 6298
Transitioned to SPMS						
N	1462 (96.7)	1156 (93.7)	1330 (93.9)	833 (88.7)	1099 (91.9)	5880 (93.4)
Y	50 (3.3)	78 (6.3)	87 (6.1)	106 (11.3)	97 (8.1)	418 (6.6)
Transitioned to SPMS within 24 months						
N	1477 (97.7)	1181 (95.7)	1374 (97.0)	896 (95.4)	1142 (95.5)	6070 (96.4)
Y	35 (2.3)	53 (4.3)	43 (3.0)	43 (4.6)	54 (4.5)	228 (3.6)
Sex						
F	1071 (70.8)	908 (73.6)	993 (70.1)	657 (70.0)	875 (73.2)	4504 (71.5)
M	441 (29.2)	326 (26.4)	424 (29.9)	282 (30.0)	321 (26.8)	1794 (28.5)
Smoking history						
Never Smoked	1508 (99.7)	1231 (99.8)	1412 (99.6)	930 (99.0)	1163 (97.2)	6244 (99.1)
Current Smoking	3 (0.2)	3 (0.2)	5 (0.4)	7 (0.7)	24 (2.0)	42 (0.7)
Past Smoking	1 (0.1)	0 (0.0)	0 (0.0)	2 (0.2)	8 (0.7)	11 (0.2)
NA	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)	1 (0.0)

Figure 4: Observed rates of transition to SPMS from RRMS by baseline EDSS score. Intervals are 95% confidence intervals. There was a substantial increase in transition rates between a score of 2 and 3, which corresponds to a 1-2-point increase to an EDSS of 4 - a key component of the Lorscheider definition of SPMS.

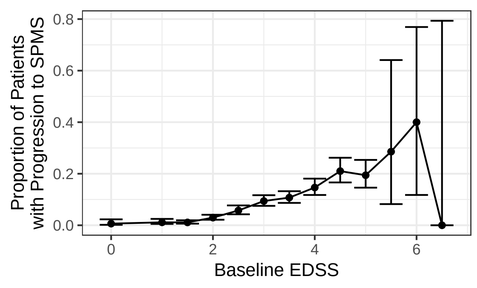


Figure 5: Power curves for inclusion criteria based on EDSS, which was one most predictive features identified. The machine learning model was used to simulate outcomes. 500 patients per arm were assumed.

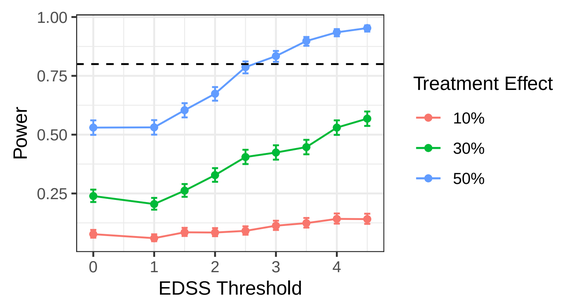


Figure 6: Shapley values for the xgboost model for a binary outcome. EDSS score and the Pyramidal subscore were important features and also part of the Lorscheider definition [2]. Other quantitative measurements were important, such as the 25-foot walk time and the nine-hole peg test. However, smoking history nor imagery data were not important features in this population.

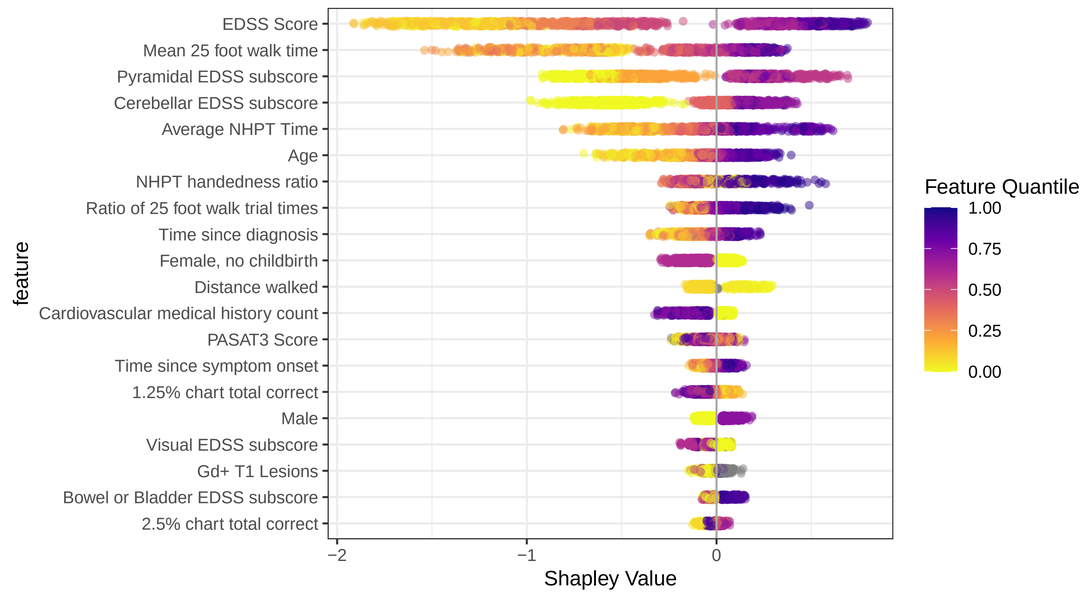
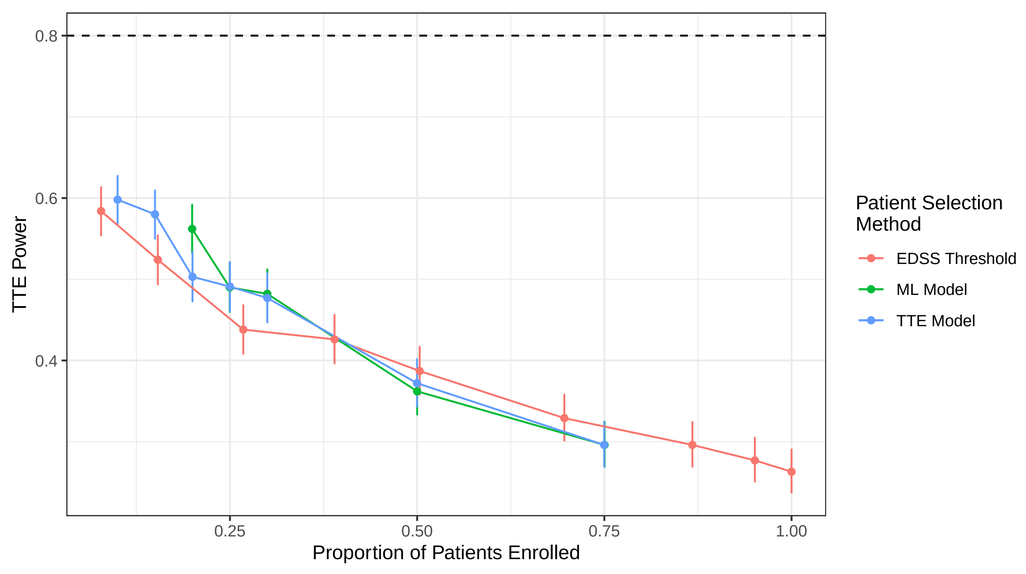


Figure 7: Power curves for TTE simulation with varying inclusion/exclusion criteria strictness and methods; the ML and TTE models both were substantially more powerful than using EDSS score alone. Simulated power from TTE simulations under $\alpha = 0.05$, using three appoaches to define inclusion/exclusion criteria. Bars are 95% confidence intervals for Monte Carlo uncertainty. The trial size per arm was fixed at 300 and the treatment effect was 30%.



Acknowledgements

This work was funded by Biogen Inc.

