Identifying Multiple Sclerosis Relapses from Clinical Notes Using Combined Rule-based and Deep Learning Methodologies

Iris Chin, PhD[1], Heather E. Moss, MD, PhD[1,2], Kathryn Sands[1], Manan Kocher, PhD[1], Oguz Akkas[1], Aracelis Z. Torres, PhD, MPH[1]
[1] Verana Health, San Francisco, California, US,
[2] Departments of Ophthalmology, Neurology & Neurological Sciences, Stanford University, Palo Alto, CA US

Objective

To develop an algorithm to extract multiple sclerosis (MS) relapse events from clinical notes in the Axon de-identified data from the American Academy of Neurology Institute (AANI) Axon Registry®, a neurology-specific patient registry that collects, reports, and analyzes real-world electronic health record (EHR) data.

Background

- Relapse frequency is a key outcome measure for MS patients, indicating disease activity.
- While there have been a number of studies that have leveraged real-world data (RWD) to describe characteristics of MS[1,2], few have focused on relapse.
- Of previous studies aimed at describing MS relapse with RWD, they have often used structured data such as claims.[3]
- However, MS relapses are often documented in unstructured clinical notes rather than structured fields, and mechanisms to automatically extract this information will better enable real-world evidence (RWE) studies.

Design & Methods

A combined rule-based, deep learning (DL) approach was developed to classify, at a given encounter, the relapse status (current relapse, no relapse, discussion of past relapse only, or unknown) of MS patients in the Axon Registry.

At the time of the study (May 2022), there were 18 million patient visits from more than 3 million patients across more than 1,000 registered providers and 150 practices in the Axon Registry. To curate disease-specific data modules (Qdata), Verana Health leverages VeraQ®, Verana's clinician-directed and AI-enhanced population health data engine, on data from the AAN's Axon Registry. Through the Qdata MS module, 46,600 MS patients were identified, with a subset validated via clinical review of patients' notes.

Model Development

- 1,000 notes were randomly sampled from MS patient notes containing relapse phrases, identified via string searches.
- The sampled notes were then labeled by a clinical expert for their relapse statuses to generate training, validation, and testing sets (70-15-15 split).
- Using the training and validation sets, a pre-trained clinical-longformer model[4] was further fine-tuned to classify notes into one of the relapse statuses.
  - To balance the classes, the number of "no relapse" notes were downsampled and additional synthetic samples of short text snippets (e.g., "Patient is having a flare up of MS." → current relapse) were added to the training set for the other classes.

Model Validation

- Performance of the model was assessed on the test set.
- To assess whether the model generated expected clinical patterns, the model was applied on all MS patient notes with relapse mentions. From the outputs, frequency of relapse metrics were calculated.

Results

Number of Examples per Class in the Training Set

Current Relapse	No Relapse	Past Relapse	Unknown
179	400	72	264

Performance Metrics on Test Set (n = 155)

	n	Specificity	Precision	Recall	F1-score
Current Relapse	12	0.97	0.71	0.83	0.77
No Relapse	92	0.90	0.94	0.95	0.94
Past Relapse	11	0.97	0.64	0.82	0.72
Unknown	40	0.97	0.88	0.75	0.81

Due to the small sample size of the "current relapse" class in the test set, an additional evaluation set was created to further assess the model's precision to ensure that the false positive rate was not high. To create the set, predictions for MS patients notes were first generated using the final model. Patients that appeared in the original model development set were excluded. A subset of notes for each predicted class was then randomly selected and reviewed by a clinical expert. Thirty samples each for "current relapse" and "no relapse" (core classes of interest) and 20 each for "past relapse" and "unknown" classes were sampled.

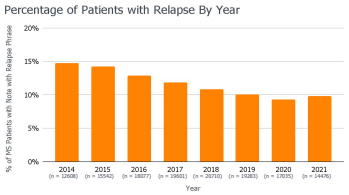
Model Precision of Follow-up Set (n = 100)

Current Relapse	No Relapse	Past Relapse	Unknown
0.80	0.83	0.65	0.80

Comparing the precision between the test set and the additional evaluation set, we observed variability in performance across sampled sets. However, the overall precision performance for the core classes, "current relapse" and "no relapse," are nonetheless acceptable and do not suggest high rates of false positives.

Assessing Clinical Patterns

- Model predictions of all MS patient notes with relapse phrases revealed that patients averaged 0.58 +/- 0.55 relapses/year (median = 0.40, IQR = 0.25 - 0.75).
- The proportion of MS patients with relapses was also found to generally decrease over time concurrent with clinical availability of higher efficacy therapies.
- Both of these findings are consistent with clinical expectations.



Conclusions

The study suggests that a combined rule-based, DL methodology can be leveraged to extract relapses from clinical notes. Performance metrics and clinically consistent patterns found in the results provide support that this may be a scalable algorithm that can be used for RWE studies. Given the small size of some of the classes in the training/test sets and variable performance found across sampled sets, future work will involve continuously maintaining and monitoring the model to ensure representativeness as well as to expand upon the data from the training datasets.

[1] Davis, M. F., Sriram, S., Bush, W. S., Denny, J. C., & Haines, J. L. (2013). Automated extraction of clinical traits of multiple sclerosis in electronic medical records. *Journal of the American Medical Informatics Association, 20*(e2), e334-e340.
[2] Culpepper, W. J., Marrie, R. A., Langer-Gould, A., Wallin, M. T., Campbell, J. D., Nelson, L. M., ... & LaRocca, N. G. (2019). Validation of an algorithm for identifying MS cases in administrative health claims datasets. *Neurology, 92*(10), e1016-e1028.
[3] Chastek, B. J., Oleen-Burkey, M., & Lopez-Bresnahan, M. V. (2010). Medical chart validation of an algorithm for identifying multiple sclerosis relapse in healthcare claims. *Journal of Medical Economics, 13*(4), 618–625.
[4] Li, Y., Wehbe, R. M., Ahmad, F. S., Wang, H., & Luo, Y. (2022). Clinical-longformer and clinical-bigbird: Transformers for long clinical sequences. *arXiv preprint arXiv:2201.11838.*

This study was funded by Verana Health and all authors are employees or consultants of Verana Health.

Verana Health