medRxiv preprint doi: https://doi.org/10.1101/2024.06.20.24309267; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

An Algorithm to Predict the Need for Magnetic Resonance Imaging to Support Disease Monitoring in Multiple Sclerosis

Michael S. Robinette, Karla Gray-Roncal*, Kathryn C. Fitzgerald, Kadija Ferryman, Casey Overby Taylor, John Scott, Elias S. Sotirchos, Peter A. Calabresi, Ellen M. Mowry, William Gray-Roncal*

**Authors contributed equally*

Abstract

People with multiple sclerosis (MS), a chronic neurologic disease, typically undergo brain magnetic resonance imaging (MRI) to assess for new disease activity or to evaluate response to disease-modifying therapy. The required frequency of these imaging studies has not yet been determined, and clinicians rely on expert panel recommendations or intuition to select the interval length for their patients. These former recommendations are often broad and lack the incorporation of individualized information (i.e., age, presence of recent disease activity, type of disease modifying therapy) into the synthesis of their assessment. We developed an algorithm that can predict new disease activity (i.e., lesions) on the next MRI scan. Multiple data sources from the electronic health records of 1045 patients were used to train and test the algorithm which resulted in an accurate solution with an area under the curve (AUC) of 0.8 that minimizes missed lesions while accurately identifying two-thirds of patients as not having a new lesion in subsequent MRI. We believe that this algorithm can be developed into a clinical decision support tool as input into a clinician-patient assessment of the appropriate interval length between MRIs for a specific patient.

Introduction

Multiple sclerosis (MS) is a relatively common disorder and a major cause of neurologic disability in early adults, impacting the ability of many to remain employed or engaged in family or personal care. The earliest “relapsing” phase of MS, for the vast majority (85-90%) of people with the diagnosis, is characterized by the development of autoimmune-induced focal areas of demyelination throughout the central nervous system (CNS) which can be visible as “lesions” on brain and spinal cord magnetic resonance imaging. When they develop in regions most relevant to daily functioning, they produce new neurological symptoms (e.g., numbness, inability to see out of one eye, or trouble walking), which are known as MS exacerbations, relapses, flare-ups, or attacks. At onset, patients have this form of MS (relapsing remitting MS [RRMS]), but later in

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.06.20.24309267; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

their course they often develop slowly worsening disability (secondary progressive MS [SPMS]), which on the average begins at the age of 45 ± 10 years.[1]

Early treatments to prevent ongoing accrual of damage to the central nervous system can mitigate the risk of MS-related inflammatory activity. There are now more than 20 disease-modifying therapies (DMTs) that prevent the autoimmune “attacks” of MS. However, the modern treatment landscape for MS includes DMTs with a wide range of efficacy, where some have greater average likelihoods of suppressing new inflammatory activity than others.[2]

Monitoring the treatment response to MS DMT has not yet been optimized, particularly in the modern treatment era. Current standard of care recommends imaging patients at routine intervals, which results in costs and patient discomfort that may not be necessary. When a person with MS begins a DMT, neurologists clinically evaluate that person with some periodicity (e.g., every 6 months) to ensure no new clinical exacerbations have occurred; they also repeat MRI of the brain and/or spinal cord to evaluate for new lesions that may have formed in the absence of new clinical symptoms. The appropriate frequency of such MRIs for specific patients or patient subgroups (e.g., older patients, those on higher efficacy DMT) has not been determined. Expert panels have suggested recommendations, but these are broad and intended for all MS patients, without offering individually tailored algorithms that support specific patients by incorporating individual patient characteristics (e.g., stability of disease) or thresholds for interval changes changes.[3,4]

Statistical and machine-learning approaches allow for a systematic approach to explore various classification methods to optimize performance while emphasizing relevance to clinical practice (e.g., interpretability). These algorithms can utilize large amounts of patient data (e.g., prior imaging results, relapse history, medication information) to use for training of their prediction models, with outputs that could potentially be read by providers and used as decision-making aids for clinically relevant questions.

We developed an algorithm to predict the presence of new lesions in a subsequent MRI. We designed and implemented an approach that accurately classifies new patient lesions and offers an interpretable result that can help to build trust and intuition surrounding these decisions. Our logistic model threshold can be adjusted to enable different risk settings and minimize missed lesions appropriately. This result can be incorporated into a clinical decision tool framework and tested in a randomized clinical trial to prospectively validate the predictions.

medRxiv preprint doi: https://doi.org/10.1101/2024.06.20.24309267; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Methods

We explain in more detail the data sources that are included as well as the classification approach used to develop our model.

Data Sources

Our tool development includes readily available data from two sources: the MS Performance Test (MSPT) [5] and the MS Smartform[6]. The MSPT is an iPad-based assessment that patients complete when they enter the clinic. This assessment provides insight into a patient's MS status by monitoring physical (e.g., walking speed, dexterity scores) and cognitive (e.g., processing speed) metrics known to be overrepresented in, and indicative of disease burden among, people with MS.[5] We focused on just three fields reflecting common symptoms in MS that may reflect disease burden: anxiety, depression, and fatigue also available as a part of the MSPT. These measures are readily available (as subscales of NeuroQoL[7]) and patient-reported, so their collection is not only standard of care at Johns Hopkins but can easily be incorporated by clinicians (or mapped to other common symptom assessments). This facilitates the sharing of this tool at other centers where MSPT itself is not standard of care.

The MS Smartform is a tool for MS providers that is collected in a standardized data format and is also available in the EHR. Data used from this repository included relapses, number of lesions on MRI, and DMT category (Figure 1) that are then projected to a viewer so patients and clinicians can review it together. The MS Smartform has been widely and freely shared through the Epic Foundation and Epic Library and has been adapted to the Cerner EHR, it is in use at several major MS centers in the US and will remain accessible to external clinicians who wish to use our model.[6]

During the data merging process, we excluded patients that had only had one visit, to provide enough historical information for the prediction algorithm. We also excluded patients who had recently started a higher-efficacy DMT to reduce potential bias; an indication to switch to such medication means the patient may reasonably have developed intervening new lesions, and it is clinically appropriate to obtain a new "baseline" scan after such a switch. Patients with no reported lesions were also excluded.

medRxiv preprint doi: https://doi.org/10.1101/2024.06.20.24309267; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Classification Approach

The output of our classifier predicts if a new lesion will occur (or not) during the next scan. For classification, we use logistic regression to predict the target variable (i.e., new lesion present). Logistic regression was chosen for its ease of use and interpretability across a compact feature set; these qualities are important to build clinical trust and effectively deploy tools. In our development process, we also explored more complex machine learning methods such as deep-learning-based classifiers but found the results to be similar with higher computational requirements, less explainability, and less robustness. We used a 5-fold stratified cross-validation technique, which trains and tests the data by iteratively splitting the set into 80%/20% training/test groups, resulting in each patient encounter being in the test set exactly once.

After obtaining our results, we analyzed different decision thresholds for our patient cohort. However, the decision threshold can be adjusted, especially in cases where the importance of false negatives or false positives may vary for a particular application or pose patient risk. During our data analysis, we identified a threshold that was conservative in selecting patients who are at high risk of developing new lesions (true positives) and should obtain a follow up scan, thus increasing its sensitivity. Because our model is targeted for a future clinical application, we wanted to be confident that patients who would benefit from a scan would be selected while eliminating unnecessary scans; we chose 8% as a threshold based on the distribution of the classification results and expert clinician input (Figure 1). For example, if the classifier predicts a new lesion will occur, the model will recommend a follow-up MRI, along with information about the features supporting this result.

Results

Patient Characteristics

A total of 1045 patients' data were used for development and testing of the classifier in LESION (Table 1). Their average age was 48 years, and there was a 3:1 female-to-male ratio; 807 (77%) self-identified as White; 191 (18%) as Black; and 9 (0.9%) as Asian. Within this cohort, average time since last relapse was 8.7 (SD 9.5) years, 468 were on a high efficacy DMT, and 445 were on a moderate one (Table 2). In aggregate, 95 (9%) had a new lesion on their follow up MRI (which served to test the model outcome).

medRxiv preprint doi: https://doi.org/10.1101/2024.06.20.24309267; this version posted July 5, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

medRxiv preprint doi: https://doi.org/10.1101/2024.06.20.24309267; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 1: Baseline characteristics of patient data used for in LESION development.

Characteristic	
Total Patients, N	1045
Age yrs, mean (SD)	48.1 (9.6)
Race, N (%)	
White	807 (77)
Black	191 (18)
Asian	9 (0.9)
Female:Male ratio	3:1
DMT, N (%)	
Higher Efficacy	468 (44.8)
Moderate Efficacy	445 (42.6)
Patients w/ New Lesions, N (%)	95 (9)
Time since relapse yrs, mean (SD)	8.7 (9.5)

Table 2: Algorithm bins the DMTs into the above traditional (moderate efficacy) and aggressive (high efficacy) groups.

Traditional (moderate efficacy)	**Aggressive (high efficacy)**
• Glatiramer acetate (Copaxone, Glatopa) • Interferons (Avonex, Betaseron, Rebif, Plegridy) • Teriflunomide • Fumarates (Tecfidera, Vumerity, Bafiertam) • Sphingo-1 phosphate modulators (fingolimod, siponimod, ozanimod, ponesimod)	• Alemtuzumab (Lemtrada) • Natalizumab (Tysabri) • Ofatumumab (Kesimpta) • Ocrelizumab (Ocrevus) • Rituximab • Cladribine (Mavenclad)

Algorithm Results

The Area Under the Curve (AUC) for our model is 0.8. With the 8% clinically reviewed threshold chosen, we obtained 67 true positives: new lesions predicted on next MRI, correctly chosen; 252 false positives, corresponding to new lesions being predicted incorrectly; 28 false negatives, corresponding to incorrectly missed lesions; and 698 true negatives, indicating an

medRxiv preprint doi: https://doi.org/10.1101/2024.06.20.24309267; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

algorithm decision that correctly predicts no new lesion. This corresponds to a sensitivity (i.e., recall) of 0.71. The algorithm specificity is 0.73. This operating point was chosen to minimize missed lesions as this could pose clinical risk. However, different thresholds can be chosen depending on the needs of the clinical environment (Figure 1).

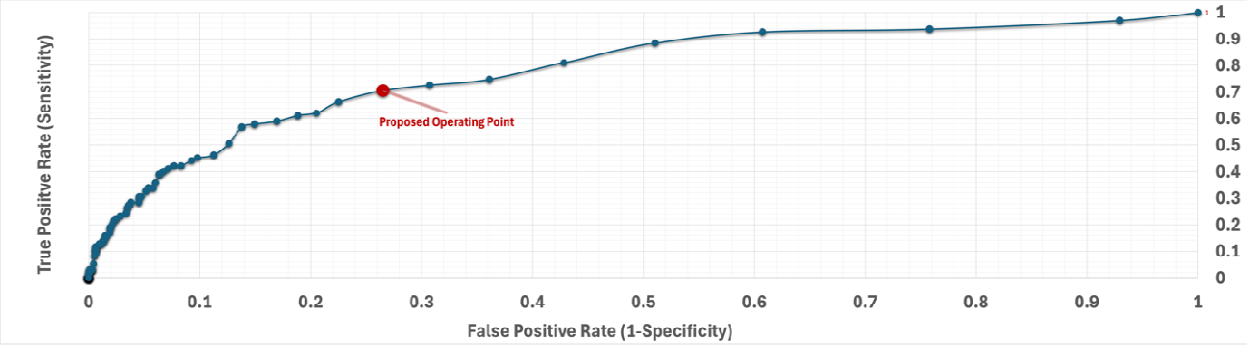
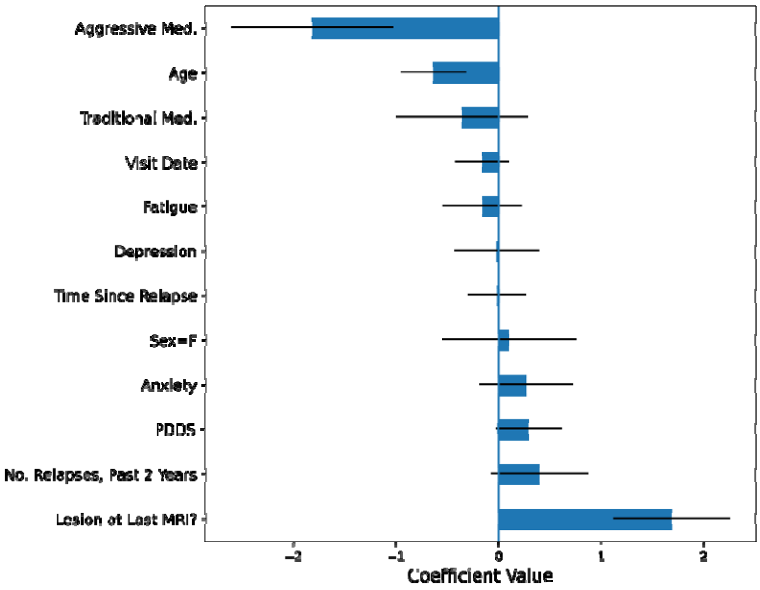


Figure 1. Classifier results showing true positive rate vs. false positive rate. The clinician-reviewed operating point is shown in red. AUC = 0.8.

Because features are normalized, variables with higher-magnitude coefficients can be interpreted as more strongly driving prediction (Figure 2). We identify several statistically significant contributors, including higher-efficacy MS DMTs, lower age, and lack of new lesion on prior MRI.



medRxiv preprint doi: https://doi.org/10.1101/2024.06.20.24309267; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 2: Logistic regression coefficients: We can assess the importance of each feature in predicting the presence of a new lesion.

Discussion

Algorithm

Results from our predictive model and prototype visualization show promising results in identifying patients that may be able to choose a longer interval between MRI scans. Because of the interpretability of the classifier and the relatively low proportion of missed lesions, this can be straightforwardly integrated into a clinical support tool. We have created an initial mock-up demonstrating how this might be integrated into an electronic health record system, recognizing that significant future development will likely be needed to aid a clinician-patient decision making conversation (Figure 3).

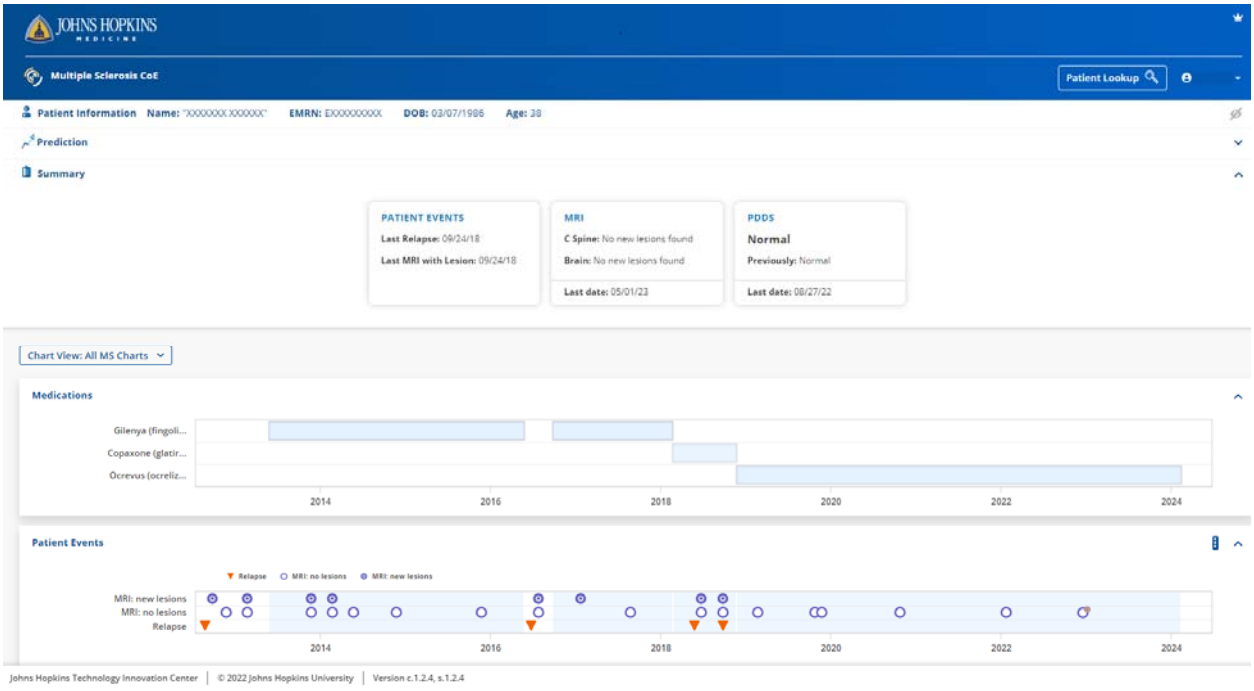


Figure 3: Visualization tool prototype for an example patient, to illustrate the environment for our algorithm, as envisioned as a clinical decision support tool supporting clinician-patient conversations.

In this initial prototype, we identify patients that should obtain follow up scans as they have a high risk of demonstrating new lesions on the next MRI. But, also as important, we identified a

medRxiv preprint doi: https://doi.org/10.1101/2024.06.20.24309267; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

large number of patients that could skip their next MRI with a low risk of missing a new lesion. If the clinician has already been considering extending the interval between MRI scans, this result may provide additional reassurance or a "second opinion" to support the clinician's assessment.

While not every patient will decide to skip the MRI, the patient burden and health system cost may be substantially reduced at an individual and health system level (e.g., a typical brain MRI might cost as much as $3500 per scan). In addition to the cost savings, this can target resources for those who need more frequent monitoring. Access to MRI scans can be limited, especially in certain geographic areas, both in the United States and in other countries; furthermore, even when MRIs are available, patients often must wait months to obtain an appointment. It is also important to emphasize the benefit to the patients who can forego frequent scans, especially those with claustrophobia, limited mobility, or chronic back pain, for whom an MRI scan can be a difficult ordeal, or for whom copay burden posits substantial resource strain.

Developing this model allowed us to study strong predictors of future MRI lesions. Our significant features were congruent with findings from previous studies - patients using higher efficacy DMTs were less likely to have new disease activity on follow up MRI as were patients with stable disease (no recent relapses or lesions) and older age. Other features included in the model did not show significance and may warrant further study, especially as sample sizes increase. More complex models that consider these predictions could be used in addition to or instead of simple logistic regression. During our design process, we explored the use of more complex traditional models (e.g., Random Forest, support vector machines), as well as deep learning and emerging methods but found that they were either inappropriate (e.g., due to volume of available data), or achieved similar performance with a loss of interpretability, which can be a major concern for gaining clinician and patient trust. Although we prioritize care for patients from minoritized backgrounds, we chose to not explicitly encode race as a classifier variable to enhance robustness. We plan to carefully monitor the effect of the model on relevant subgroups throughout the deployment process. Furthermore, we hope that future versions will also make use of other data types, including biomarkers (e.g. serum neurofilament light chain[8]), to improve the accuracy of model predictions. We validated the stability of our approach by adding an updated tranche of patient data (approximately a year of visits), which resulted in similar output performance of the model after retraining. Further validation may be performed by replicating our approach in another clinical population such as MS PATHS[5], DISCO[9], VIDAMS[10], and CombiRx[11].

medRxiv preprint doi: https://doi.org/10.1101/2024.06.20.24309267; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Implementation Considerations

Our implementation design needs to consider real-life implementation into the EHR and the values that are reported to the clinician. User-center design concepts such as transparency of information need to be considered. For example, a clinician may need access not only to the recommendation but also the reasons behind the recommendation from the model perspective. This would allow the provider to assess if those aspects are consistent with the patient history and evidence.

Our initial design needs further validation with larger data sets and prospective studies. The threshold that was chosen was based on data from this study and could be different once it is studied prospectively. One component of our approach is to consider both algorithmic and implementation bias as tool implementation is studied. Sociodemographic information is available in the EHR to support these analyses, including patient sex at birth and gender, race, insurance data, and geocode-based determinants of health. A limitation worth considering is that the system relies on data availability from sources that need to be manually entered by a clinician. If the MS Smartform is incomplete, algorithm training is affected and results may not be as meaningful, or a result may not be available for a particular patient.

In conclusion, our approach produced a pilot algorithm to help clinicians triage patients that would benefit from a subsequent surveillance MRI versus those could be good candidates for decreased frequency of imaging. Once refined into a clinical decision-support tool, this approach may help support clinician-patient conversations regarding the value of image monitoring based on their individual patient characteristics.

medRxiv preprint doi: https://doi.org/10.1101/2024.06.20.24309267; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

References

1. Zeydan B, Kantarci O. Progressive forms of multiple sclerosis. Distinct entity or age-dependent phenomenon. Neurol Clin 2018;36:163-171.

2. Sormani MP, Gasperini C, Romeo M, Rio J, Calabrese M, Cocco E, Enzingher C, Fazekas F, Filippi M, Gallo A, Kappos L, Marrosu MG, Martinelli V, Prosperini L, Rocca MA, Rovira A, Sprenger T, Stromillo ML, Tedeschi G, Tintorè M, Tortorella C, Trojano M, Montalban X, Pozzilli C, Comi G, De Stefano N; MAGNIMS study group. Assessing response to interferon-β in a multicenter dataset of patients with MS. Neurology 2016;87(2):134-40.

3. Wattjes MP, Ciccarelli O, Reich DS, Banwell B, de Stefano N, Enzinger C, Fazekas F, Filippi M, Frederiksen J, Gasperini C, Hacohen Y, Kappos L, Li DKB, Mankad K, Montalban X, Newsome SD, Oh J, Palace J, Rocca MA, Sastre-Garriga J, Tintoré M, Traboulsee A, Vrenken H, Yousry T, Barkhof F, Rovira À; Magnetic Resonance Imaging in Multiple Sclerosis study group; Consortium of Multiple Sclerosis Centres; North American Imaging in Multiple Sclerosis Cooperative MRI guidelines working group. 2021 MAGNIMS-CMSC-NAIMS consensus recommendations on the use of MRI in patients with multiple sclerosis. Lancet Neurol 2021;20:653-670.

4. Saslow L, Li DKB, Halper J, Banwell B, Barkhof F, Barlow L, Costello K, Damiri P, Dunn J, Giri S, Maes M, Morrow SA, Newsome SD, Oh J, Paul F, Quarterman P, Reich DS, Shewchuk JR, Shinohara RT, Van Hecke W, van de Ven K, Wallin MT, Wolinsky JS, Traboulsee A. An international standardized magnetic resonance imaging protocol for diagnosis and follow-up of patients with multiple sclerosis: advocacy, dissemination, and implementation strategies. Int J MS Care 2020;22:226-232.

5. Mowry EM, Bermel RA, Williams JR, Benzinger TLS, de Moor C, Fisher E, Hersh CM, Hyland MH, Izbudak I, Jones SE, Kieseier BC, Kitzler HH, Krupp L, Lui YW, Montalban X, Naismith RT, Nicholas JA, Pellegrini F, Rovira A, Schulze M, Tackenberg B, Tintore M, Tivarus ME, Ziemssen T, Rudick RA. Harnessing Real-World Data to Inform Decision-Making: Multiple Sclerosis Partners Advancing Technology and Health Solutions (MS PATHS). Front Neurol. 2020 Aug 7;11:632.

6. Roy S, Grader-Beck T, Mann H, Kantawala M, Fitzgerald KC, Scott TF, Mowry EM. Estimated time saved in MS clinic using MS Smartforms to retrieve clinical data. In: Consortium of MS Centers Annual Meeting, June 2, 2022, Washington, DC.

7. Medina LD, Torres S, Alvarez E, Valdez B, Nair KV. Patient-reported outcomes in multiple sclerosis: Validation of the Quality of Life in Neurological Disorders (Neuro-QoL™) short forms. Mult Scler J Exp Transl Clin. 2019 Nov 26;5(4):2055217319885986.

8. Sotirchos ES, Fitzgerald KC, Singh CM, Smith MD, Reyes-Mantilla M, Hersh CM, Hyland MH, Canissario R, Simmons SB, Arrambide G, Montalban X, Comabella M, Naismith RT, Qiao M, Krupp LB, Nicholas JA, Akgün K, Ziemssen T, Rudick R, Fisher E, Bermel RA, Mowry EM, Calabresi PA. Associations of sNfL with clinico-radiological measures in a large MS population. Ann Clin Transl Neurol. 2023 Jan;10(1):84-97.

medRxiv preprint doi: https://doi.org/10.1101/2024.06.20.24309267; this version posted July 5, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

9. Corboy JR, Fox RJ, Kister I, Cutter GR, Morgan CJ, Seale R, Engebretson E, Gustafson T, Miller AE; DISCOMS investigators. Risk of new disease activity in patients with multiple sclerosis who continue or discontinue disease-modifying therapies (DISCOMS): a multicentre, randomised, single-blind, phase 4, non-inferiority trial. Lancet Neurol. 2023 Jul;22(7):568-577.

10. Cassard SD, Fitzgerald KC, Qian P, Emrich SA, Azevedo CJ, Goodman AD, Sugar EA, Pelletier D, Waubant E, Mowry EM. High-dose vitamin D_3 supplementation in relapsing-remitting multiple sclerosis: a randomised clinical trial. EClinicalMedicine. 2023 Apr 13;59:101957. doi: 10.1016/j.eclinm.2023.101957. PMID: 37125397; PMCID: PMC10130605.

11. Koch MW, Moral E, Brieva L, Mostert J, Strijbis EM, Comtois J, Repovic P, Bowen JD, Wolinsky JS, Lublin FD, Cutter G. Relapse recovery in relapsing-remitting multiple sclerosis: An analysis of the CombiRx dataset. Mult Scler. 2023 Dec;29(14):1776-1785.