SUPPLEMENTARY METHODS

Study design and patient population

For each patient, we recorded demographic characteristics, past medical history and information on the clinical presentation including the temporal profile, initial symptoms and the neurological exam. CSF features recorded included pleocytosis, protein levels, oligoclonal bands (OCB), and IgG index. Collected MRI characteristics included presence of gadolinium enhancement (Gad+), lesion localization on T2-weighted sagittal images (reference to the corresponding vertebral level) and lesion topography on axial images (anterior, central, lateral or posterior spinal cord). All patients underwent extensive work up, including spine and brain MRI and if clinically indicated, lumbar puncture, NMO, rheumatological and paraneoplastic panels, chest CT, serology and PCRs for infections known to cause myelopathy, and spinal angiogram when a vascular etiology was suspected.

Statistical Analysis

Our analysis had two aims: 1) to assess descriptively how specific characteristics are associated with myelopathies of different etiologies and 2) derive a subset of predictors that improve the prediction accuracy of identifying etiologic origin of a given myelopathy. For the first stage, we grouped characteristics into five sets of potentially relevant predictors:

1) demographic/past medical history (age, sex, ethnicity, smoking status, obesity, hypertension, diabetes, dyslipidemia, autoimmune disease, infection in the last 30 days, and vaccination in the last 90 days),
2) clinical presentation (temporal profile, presence of motor, sensory or bladder/bowel symptoms, new onset back pain and worsening by exercise),

3) neurological exam (motor exam [weakness, defined as MRC score ≤ 4 in at least one major muscle group considered in the ASIA score, and muscle tone classified as normal, flaccid or spastic], sensory abnormality [vibration and proprioception as indicator of posterior column dysfunction, light touch, pain/temperature, sensory level], presence of urinary retention or abnormal rectal tone, reflexes),
4) MRI (presence of longitudinally extensive lesions, Gad+, multifocality, axial lesion topography, sagittal lesion location) and
5) CSF (pleocytosis [≤5 cells/uL, 6-19 cells/uL, ≥20 cells/uL], protein [≤45 mg/dL, 46-75 mg/dL, >75 mg/dL], IgG index [≤0.7, >0.7] and OCB). For each group of predictors, we fit a multinomial regression model where we considered each myelopathy type as an outcome. Models for MRI and CSF features were adjusted for time to MRI or time to lumbar puncture.

Our second objective was to evaluate, after accounting for Gad+ and pleocytosis, whether a subset of predictors would improve accuracy in discriminating the different myelopathy categories. We focused on estimation of the multinomial-generalized integrated discrimination increment (IDI) and the net reclassification improvement (NRI) as they are measures that estimate prediction increment of new variables (1). The IDI is a measure of separation of the predicted probabilities for each type of event, while the NRI is roughly the proportion of participants which are correctly *versus* incorrectly classified between two models. For this analysis, we divided the data into a training (n=290) and testing set (n=151) and fit a series of multinomial models in the testing set, following a forward selection algorithm. We excluded patients who didn't receive contrast on initial MRI (n=16). For individual characteristics, we excluded patients with missing information, but performed subsequent sensitivity analyses using

missing indicator variables. We started with a model only adjusting for Gad+ and pleocytosis, and fit individual multinomial models for each of the other possible characteristics (37 models). We then calculated the IDI and correct classification rate (CCR) for each model. The characteristic which maximizes the IDI (IDI_{max}) was added to the model. We fit a second series of multinomial models (including Gad+, pleocytosis, and IDI_{max}) and calculated the IDI, NRI, and CCR for this set of models. The algorithm continued to add predictors until IDI_{max} did not significantly improve or improvements of CCR were less than 1%. Once the algorithm terminated and a final model was derived, we calculated the NRI, IDI, CCR, and multinomial-generalized AUC (2) of this model fit in the testing set. In sensitivity analyses, we restricted the set of predictors related to demographic information, clinical presentation and neurologic exam findings to mitigate possible incorporation bias (potentially introduced by the inclusion of predictors related to MRI findings). Similar to the strategy described above, we started with a model only adjusting for Gad+ and pleocytosis, and fit individual multinomial models for each of the other possible characteristics (24 models) and followed an identical algorithm to arrive at a final model, which we then refit in the testing set.

REFERENCES

1. Li J, Jiang B, Fine JP. Multicategory reclassification statistics for assessing improvements in diagnostic accuracy. Biostatistics. 2013 Apr; 14(2):382-394
2. Li J, Fine JP. ROC analysis with multiple classess and multiple tests: methodology and its application in microarray studies. Biostatistics. 2008 Jul;9(3):566-76