

RESEARCH ARTICLE

Systematic review of prediction models in relapsing remitting multiple sclerosis

Fraser S. Brown^{1*}, Stella A. Glasmacher¹, Patrick K. A. Kearns¹, Niall MacDougall², David Hunt^{1,3,4}, Peter Connick^{1,4}, Siddharthan Chandran^{1,4,5}

1 Anne Rowling Regenerative Neurology Clinic, University of Edinburgh, Edinburgh, United Kingdom, **2** Institute of Neurological Sciences, Glasgow, United Kingdom, **3** MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom, **4** Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, United Kingdom, **5** UK Dementia Research Institute, University of Edinburgh, Edinburgh, United Kingdom

* fbrown34@exseed.ed.ac.uk



OPEN ACCESS

Citation: Brown FS, Glasmacher SA, Kearns PKA, MacDougall N, Hunt D, Connick P, et al. (2020) Systematic review of prediction models in relapsing remitting multiple sclerosis. PLoS ONE 15(5): e0233575. <https://doi.org/10.1371/journal.pone.0233575>

Editor: Sreeram V. Ramagopalan, University of Oxford, UNITED KINGDOM

Received: February 10, 2020

Accepted: May 7, 2020

Published: May 26, 2020

Copyright: © 2020 Brown et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: PKAK is supported by the Wellcome Trust. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: NM has received hospitality for educational events from Biogen,

Abstract

The natural history of relapsing remitting multiple sclerosis (RRMS) is variable and prediction of individual prognosis challenging. The inability to reliably predict prognosis at diagnosis has important implications for informed decision making especially in relation to disease modifying therapies. We conducted a systematic review in order to collate, describe and assess the methodological quality of published prediction models in RRMS. We searched Medline, Embase and Web of Science. Two reviewers independently screened abstracts and full text for eligibility and assessed risk of bias. Studies reporting development or validation of prediction models for RRMS in adults were included. Data collection was guided by the checklist for critical appraisal and data extraction for systematic reviews (CHARMS) and applicability and methodological quality assessment by the prediction model risk of bias assessment tool (PROBAST). 30 studies were included in the review. Applicability was assessed as high risk of concern in 27 studies. Risk of bias was assessed as high for all studies. The single most frequently included predictor was baseline EDSS ($n = 11$). T2 Lesion volume or number and brain atrophy were each retained in seven studies. Five studies included external validation and none included impact analysis. Although a number of prediction models for RRMS have been reported, most are at high risk of bias and lack external validation and impact analysis, restricting their application to routine clinical practice.

Introduction

The natural history of relapsing remitting multiple sclerosis (RRMS) is variable and prediction of individual prognosis is challenging [1]. The inability to reliably prognosticate at diagnosis has important implications for informed decision making especially in relation to disease modifying therapy (DMT). Risk stratification at diagnosis into disease severity categories (mild, moderate or severe) could better allow treating physicians and people with RRMS to make treatment decisions, but this is difficult early in the disease process.

Novartis, Genzyme, Merck and Roche over the past 5 years. NM has received honoraria for talks or advisory boards from Biogen, Novartis, Genzyme, Merck and Roche over the same time period. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

As a consequence, broadly speaking, there are two treatment strategies in early RRMS: induction and escalation [1]. An induction strategy involves initiation of potent DMTs early in disease course [1]. An escalation strategy, on the other hand, involves initiating therapy with less potent agents with lower risk of serious adverse reactions, and then subsequently offering escalation to more potent DMTs if necessary. The induction strategy offers early control of disease but may cause harm from overtreatment. The escalation strategy risks harm from undertreatment and preventable neuroinflammation. As RRMS disproportionately affects individuals of working age, including females of childbearing potential, often pragmatic decisions need to be made that fall between these two strategies. For many reasons, therefore, there is a need for predictive tools that can be used by individual patients to inform treatment and life choices [2,3].

Multiple individual clinical and paraclinical factors have been studied for their ability to discriminate between patients with differing short and long-term prognoses. Poor prognosis has been associated with male sex and older age at disease onset [1,2]. However, a systematic review identified that evidence supporting the former is poor while predictive effect of older age is dependent on its definition [2]. Early clinical features such as sphincter involvement, higher baseline disability [2,4–7] and certain magnetic resonance imaging (MRI) measures—brain atrophy rate and T2-weighted lesion number and volume [8–12]—appear to be the most robust predictors of poor prognosis but these rely on established damage and so are not ideal prognostic measures. In contrast, biomarkers such as vitamin D level may confer prognostic effect at an earlier time point: an inverse relationship between serum vitamin D levels and hazard of relapse at six months has been reported [13]. The presence of cerebrospinal fluid (CSF) immunoglobulin M oligoclonal bands (IgMOB) is a putative biomarker for future relapse and conversion to secondary progression in RRMS, but requires further validation [14,15]. Lifestyle factors have attracted attention as they are potentially modifiable: Smoking has been shown to shorten time to onset of secondary progression [16,17]. However, whilst obesity appears to increase the chances of developing multiple sclerosis, its role in determining prognosis remains to be determined [18].

In a previous systematic review Langer-Gould et al focused on individual clinical and demographic factors in RRMS rather than composite models and did not include imaging variables [2]. Havas et al reviewed prediction models in RRMS focusing on predicting treatment response [19]. Predictive modelling, using patient-specific data points to predict outcome, is an unmet need in RRMS. Using published guidance for reporting and risk of bias assessment, our systematic review aims to add to this literature by describing and evaluating the methodological quality of studies that develop and validate predictive models in RRMS.

Methods

Review aim, scope, target population, outcomes and intended moment of model use were defined as guided by CHARMS [20 and S1 File]. Study details and pre-specified search strategy were registered through PROSPERO, reference CRD42019149140 (<https://www.crd.york.ac.uk/prospERO/>).

This review reports on studies that identify predictors of target outcomes, assign weights (eg. using regression coefficients) to each predictor using multivariable analysis, and develop a prediction model for adult patients with RRMS. Herein, a prediction model is taken to mean a model which uses multiple predictors in combination to determine probability of an outcome [23]. Intended moment of model use is at diagnosis. External validation studies were also included. We excluded studies predominantly selecting children (<18 years old), predicting response to disease modifying therapy, predicting conversion of clinically isolated syndrome (CIS) to MS, studies exclusively including patients with CIS, primary progressive multiple

sclerosis (PPMS), secondary progressive multiple sclerosis (SPMS) or studies investigating a single predictor, test or marker as they do not meet the definition of prediction model as above. Outcomes of interest included inflammatory disease activity (clinical relapse rate, T2 lesion load change), rate of neurodegeneration (brain atrophy, clinical progression of fixed disability), progression to SPMS and degree of disability.

A search of OVID MEDLINE, Embase and ISI Web of Science was conducted using a pre-specified search strategy [S2 File]. Records not meeting inclusion criteria or clearly not prediction modelling studies were excluded by one reviewer. The remaining records were screened by two medically qualified reviewers [FSB and SAG] independently and full articles were reviewed if eligible. Disagreements were resolved by consensus. There were no limitations with regard to study language or publication date.

Data extraction (following CHARMS [20]) was performed by one reviewer and quality assessment by two reviewers. The categories for data extraction are detailed in full in the PROSPERO record but include source of data, participants, outcome candidate predictors, model development and model evaluation. Quality assessment of studies was carried out following Prediction model Risk Of Bias ASsessment Tool [PROBAST] which rates study methodology and applicability to review question as at “high”, “low” or “unclear” risk of bias based on a predetermined set of questions and scoring guide [21]. Inter-rater agreement in these domains was measured by Cohen’s kappa statistic.

Results

Database searches from inception to August week three 2019 identified 5193 studies of which 30 studies met the pre-defined inclusion criteria (Fig 1) [22–51]. 23 studies were model development only [22–24,26,29–34,37,39–46,48–51], five were model development and external validation in the same study [25,35,36,38,47] and two were external validation studies of the same model [27,28]. Studies used Poser (n = 16), McDonald 2001 (n = 4), McDonald 2005 (n = 6) and McDonald 2010 (n = 2) diagnostic criteria (S3 File). Diagnostic criteria were not specified in five studies [23,32,35,36,51]. Four studies used more than one set of diagnostic criteria [22,31,37,38]. Two studies used data from multiple clinical trials likely with heterogeneous diagnostic criteria [24,33]. A summary of study attributes is included in S4 File and risk of bias assessment in Table 1. Agreement in PROBAST assessment between reviewers was 96.5% and 79.3% for overall concern for risk of bias and applicability to our research question, respectively.

Source of data and participants

27 studies [22,23,25–32,34–46,48–51] used cohort design, which is recognised as an optimal strategy for prediction model development. Three studies used data from clinical trials [24,33,47]. 21 studies were single centre and nine multicentre. All 30 studies reported inclusion and exclusion criteria. 11 studies featured populations consisting only of patients with RRMS [27,28,35,36,38,39,43,46–49]. Percentage of patients treated with DMTs in studied cohorts varied from 0–100% (S5 File). Only four studies were judged not to be at high risk of selection bias [28,39,42,47].

Candidate predictors

Demographic, clinical, MRI, CSF and electrophysiology variables were retained as predictors in final models (Table 2). The single most common clinical predictor was baseline EDSS (n = 11). Age (n = 6), age at onset (n = 6) and gender (n = 5) were also commonly retained. T2 lesion volume or number and brain atrophy were each retained in seven studies. In ten studies, predictor measurement timing matched our review question’s target timing (that is, the



PRISMA 2009 Flow Diagram

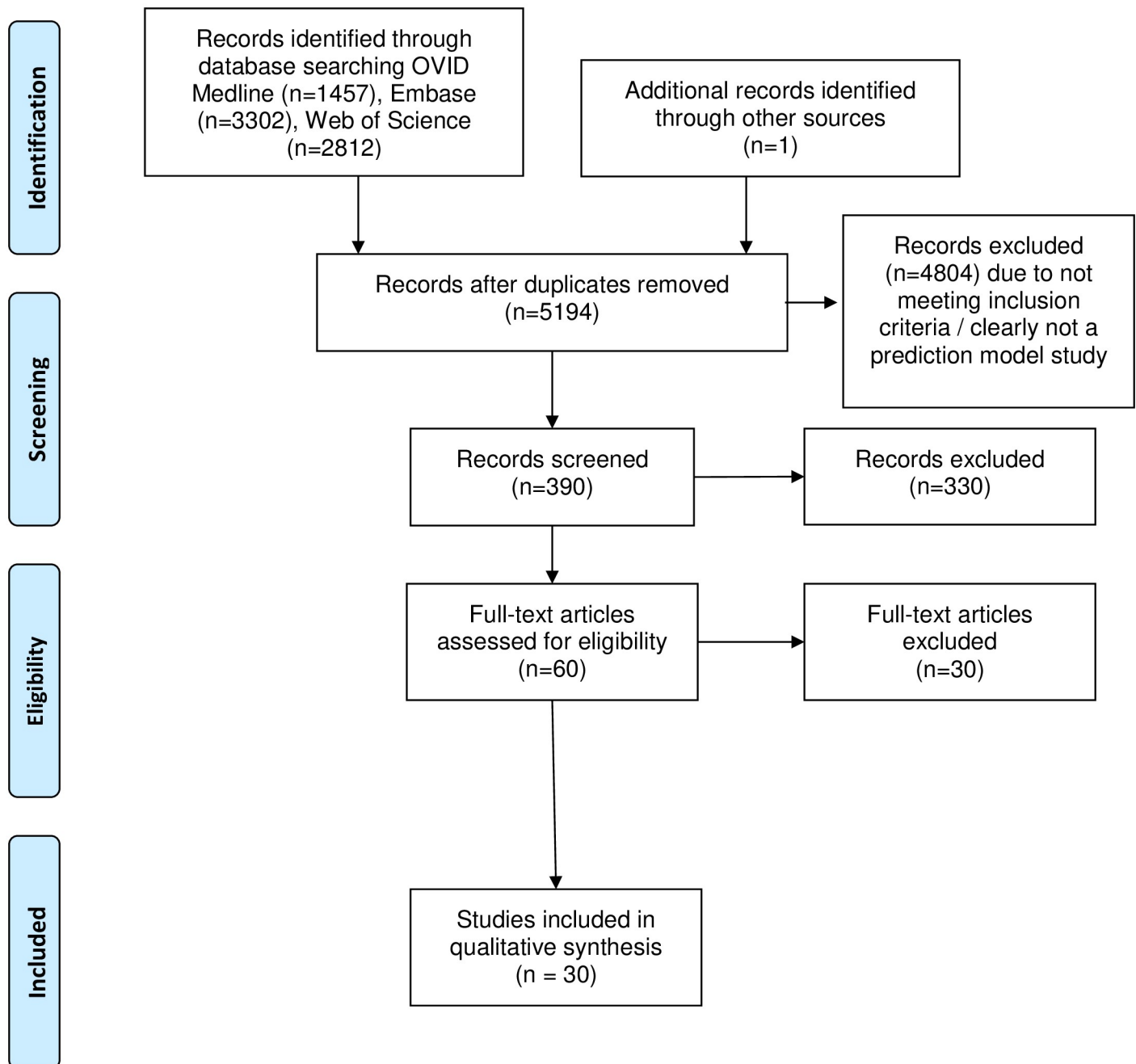


Fig 1. PRISMA flow diagram.

<https://doi.org/10.1371/journal.pone.0233575.g001>

Table 1. PROBAST: Assessment of risk of bias and applicability of a) development and b) external validation papers.

| a | | | | | | | |
|-----------------------|--------------|------------|---------|----------|---------------|------------|---------|
| Study | ROB | | | | Applicability | | |
| | Participants | Predictors | Outcome | Analysis | Participants | Predictors | Outcome |
| Agosta 2006 | ✘ | ✓ | ✓ | ✘ | ✘ | ✘ | ✓ |
| Bakshi 2008 | ✘ | ✓ | ✓ | ✘ | ✘ | ✘ | ✓ |
| Barkhof 2005 | ✘ | ✘ | ✘ | ✘ | ✘ | ✓ | ✓ |
| Bejarno 2011 | ✘ | ✓ | ✓ | ✘ | ✘ | ✘ | ✓ |
| Bergamaschi 2001 | ✘ | ✓ | ✘ | ✘ | ✘ | ✓ | ✓ |
| De Groot 2009 | ✘ | ✘ | ✘ | ✘ | ✘ | ✓ | ✓ |
| Dekker 2019 | ✘ | ✓ | ✘ | ✘ | ✘ | ✘ | ✓ |
| Filippi 2012 | ✘ | ✓ | ✘ | ✘ | ✘ | ✘ | ✓ |
| Gauthier 2007 | ✓ | ✓ | ✘ | ✘ | ✘ | ✘ | ✓ |
| Held 2005 | ✘ | ✘ | ✘ | ✘ | ✘ | ✘ | ✓ |
| Liguori 2011 | ✘ | ✘ | ✘ | ✘ | ✘ | ✘ | ✓ |
| Mandrioli 2008 | ✘ | ✘ | ✘ | ✘ | ✘ | ✓ | ✓ |
| Manouchehrinia 2019 | ✘ | ✓ | ✘ | ✘ | ✓ | ✓ | ✓ |
| Margaritella 2012 (A) | ✘ | ✘ | ✘ | ✘ | ✘ | ✘ | ✓ |
| Margaritella 2012 (B) | ✘ | ✓ | ✘ | ✘ | ✘ | ✘ | ✓ |
| Mesaros 2008 | ✘ | ? | ✘ | ✘ | ✓ | ✘ | ✓ |
| Minneboo 2008 | ✘ | ✓ | ✘ | ✘ | ✘ | ✓ | ✓ |
| Popescu 2013 | ✘ | ✓ | ✘ | ✘ | ✘ | ✘ | ✓ |
| Ramsaransing 2007 | ✘ | ✘ | ✘ | ✘ | ✘ | ✘ | ✓ |
| Runmarker 1994 | ✘ | ✘ | ✘ | ✘ | ✓ | ✓ | ✓ |
| Schlaeger 2012 | ✘ | ✓ | ✓ | ✘ | ✘ | ✘ | ✓ |
| Schlaeger 2014 | ✘ | ✓ | ✓ | ✘ | ✘ | ✘ | ✓ |
| Skoog 2014 | ✘ | ✘ | ✘ | ✘ | ✓ | ✘ | ✓ |
| Sormani 2007 | ✘ | ✓ | ✘ | ✘ | ✘ | ✘ | ✓ |
| Uher 2017 | ✘ | ✓ | ✘ | ✘ | ✘ | ✘ | ✓ |
| Von Gumberz 2016 | ✘ | ✓ | ✘ | ✘ | ✓ | ✘ | ✓ |
| Weideman 2017 | ✘ | ✓ | ✘ | ✘ | ? | ✘ | ✘ |
| Weinshenker 1991 | ✘ | ✘ | ✘ | ✘ | ✘ | ✓ | ✓ |

| b | | | | | | | |
|------------------|--------------|------------|---------|----------|---------------|------------|---------|
| Study | ROB | | | | Applicability | | |
| | Participants | Predictors | Outcome | Analysis | Participants | Predictors | Outcome |
| Bergamaschi 2007 | ✘ | ✘ | ✘ | ✘ | ✘ | ✓ | ✓ |
| Bergamaschi 2015 | ✘ | ✘ | ✘ | ✘ | ✓ | ✓ | ✓ |

PROBAST assessment performed by two independent reviewers with Kappa value 0.458. Final agreed assessment presented. Where more than one model was developed in a study, PROBAST scoring is reported only once. ✓ = low risk of bias, ✘ = high risk of bias, ? = unclear risk of bias.

<https://doi.org/10.1371/journal.pone.0233575.t001>

authors studied variables present at time of diagnosis with RRMS) [24,26–29,35,36,40,50,51]. In nine studies, subjective predictor definitions or variable determination methods were used [24,29,33–35,37,43,46,51]. In nine studies, continuous predictors were categorised, another potential source of bias [26–28,32–34,36,44,48]

Model outcomes

Three studies had outcomes which were not objectively defined [34,43,46]. Within studies, the same outcome assessment method was generally applied to all patients. In four studies using

Table 2. Frequency of variables included in prediction models by development study.

| Study | Variable | | | | | | | | |
|-----------------------|----------|-----------|--------|----------|-----|----|-----|------|-----|
| | Age | Onset age | Gender | Clinical | MRI | EP | CSF | FMHx | DMT |
| Agosta 2006 | | | | | ✓ | | | | |
| Bakshi 2008 | | | | ✓ | ✓ | | | | |
| Barkhof 2005 | | ✓ | | ✓ | ✓ | | | | |
| Bejarno 2011 | | | | ✓ | | ✓ | | | |
| Bergamaschi 2001 | | ✓ | ✓ | ✓ | | | | | |
| De Groot 2009 | | | | ✓ | ✓ | | | | |
| Dekker 2019 | | | | ✓ | ✓ | | | | |
| Filippi 2012 | | | | | ✓ | | | | |
| Gauthier 2007 | ✓ | | | ✓ | ✓ | | | | |
| Held 2005 | | | | ✓ | | | | | |
| Liguori 2011 | | | | | ✓ | | | | |
| Mandrioli 2008 | | | | ✓ | | | ✓ | | |
| Manouchehrinia 2019 | ✓ | ✓ | ✓ | ✓ | | | | | |
| Margaritella 2012 (A) | | ✓ | ✓ | ✓ | | ✓ | | | |
| Margaritella 2012 (B) | | | | ✓ | | ✓ | | | |
| Mesaros 2008 | | | | ✓ | | | | | |
| Minneboo 2008 | ✓ | | | ✓ | ✓ | | | | |
| Popescu 2013 | | | | | ✓ | | | | |
| Ramsaransing 2007 | | | | ✓ | | | | | |
| Runmarker 1994 | | ✓ | ✓ | ✓ | | | | | |
| Schlaeger 2012 | | | | | ✓ | | | | |
| Schlaeger 2014 | | | | | | ✓ | | | ✓ |
| Skoog 2014 | ✓ | | | ✓ | | | | | |
| Sormani 2007 | | | | ✓ | ✓ | | | | |
| Uher 2017 | | | | ✓ | ✓ | | | | |
| Von Gumberz 2016 | | | | ✓ | ✓ | | | | ✓ |
| Weideman 2017 | ✓ | | | ✓ | | | | ✓ | ✓ |
| Weinshenker 1991 | ✓ | ✓ | ✓ | | | | | | |

MRI: magnetic resonance imaging. EP: electrophysiology. CSF: cerebrospinal fluid. FMHx: Family history. DMT: disease modifying therapy.

<https://doi.org/10.1371/journal.pone.0233575.t002>

EDSS as an outcome measure different assessment methods (telephone EDSS as opposed to full examinations) were used in some patients [31,41,44,45]. External validation in one study used a different definition of secondary progression to the development cohort [36]. No studies reported blinding of outcome assessors to all predictor information.

Model development and evaluation

Regression analysis was the most common modelling technique (n = 24). Neural networks, gradient boosting machine, Bayesian and Markov modelling techniques were each used in one model development study. Ten studies used univariate or bivariate analyses to filter potential predictors. Outcome events per predictor ratio (EPV) of less than 10 is a widely recognised criterion for identifying models at risk of overfitting [21]. This calculation is not applicable to models with continuous outcomes. There was insufficient information to calculate EPV in four studies [32,37,43,47]. 12 of the 16 studies in which EPV was applicable and could be calculated had scores of <10 (S6 File). 18 studies used complete case analysis. Four studies reported

imputation: three used last observation carried forward and one used multiple imputation. In eight studies there was insufficient information to determine missing data handling.

Internal validation was present in nine studies: four used cross-validation, three used split-sample and two used bootstrap. Only one study reported applying shrinkage methods [29]. Discrimination and calibration are common prediction model performance measures [21]. Discrimination is commonly assessed by area under the receiver operator curve (AUC) while it is recommended calibration be presented as a plot of observed versus predicted outcomes [21]. AUC was reported in ten studies and ranged from 0.64 to 0.89. Calibration was reported graphically in four studies. Goodness of fit performance statistics R^2 or nagelkerke R^2 were reported in seven studies.

Five studies included external validation in their model development. In two instances, this was restricted to temporal external validation [35,38]. Only one study reported performance measures in the external validation cohort where AUC ranged between 0.77–0.87 [35]. The Bayesian Risk Estimate for MS score [26] was externally validated in two subsequent studies with the second removing predictors [29,30]. AUC was not reported in either of the BREMS validation cohorts [29,30]. All external validation was performed by the authors of the respective development models.

Presentation of model and utility

Five studies presented the outcome of model development as a risk score [26–29, 48]. Two presented a web-based application [46,50]. One presented a nomogram [36]. For example, Skoog et al produced an online prediction score calculator on a freely available website which allows input of the current age of the patient, time since most recent attack, the main symptom type and whether there has been complete remission of most recent attack [46]. The output of this score is a percentage annual risk of conversion to secondary progression [46]. None of the studies carried out an impact assessment.

Discussion

Here, we present a systematic review of studies investigating prognostic models for use in people with RRMS. In the models studied, the single most common clinical predictor was baseline EDSS ($n = 11$). Demographic variables, including age and sex, and MRI markers, including T2 lesion volume or number and brain atrophy, were also often retained in the models studied here. Only one study included a CSF marker- IgMOB- in its final model. Vitamin D levels and smoking status, which have some published support for their prognostic relevance, did not feature in any models.

Our results demonstrate that there is agreement between a limited number of studies showing the prognostic effects of demographic and radiological parameters [8–12, 19]. We identified no studies that incorporated demographic, radiological, and biomarker data. Other work, including reviews by Langer-Gould et al (focusing on individual predictors) and Havas et al (focusing on predicting treatment response) also identified early disease course as a predictor of outcome [2,4–6,19].

Applicability

Most of the models studied were not developed and / or validated for use at time of diagnosis of RRMS. In addition, this cross section of newly diagnosed patients has changed over time with the evolution of diagnostic criteria [52]. None of the models were developed or validated in cohorts whose diagnosis was made using the 2017 McDonald criteria.

Many predictors required information unavailable at the time of diagnosis such as longitudinal disease course features. Optimally, a prediction tool would be applicable at time of diagnosis to facilitate initial treatment decisions and would incorporate predictively relevant information from all domains that are available at that point. That is, to tailor a treatment strategy for a patient that falls between the escalation and initiation strategies based on a best estimation of that individual's risks. Furthermore, the majority of studies included patients taking DMTs. Variable DMT usage introduces heterogeneity between studies and between participants within studies, and therefore hampers interpretation and comparisons. An improved understanding of the impact of DMTs on long-term outcomes will be needed in order to fully inform model-guided treatment strategies. As such, there is still a major unmet need with regard to developing prediction models applicable to patients with newly diagnosed RRMS.

Risk of bias

All included studies were at an overall risk of bias. Selection bias was a concern in the majority of studies. Often this was due to exclusion or inappropriate imputation of participants with missing data. The majority of studies used complete case analysis which can introduce bias given potential non-random distribution of missing data [21]. In addition, the last visit carried forward approach will flatter participants who are lost to follow up. Further selection bias was judged likely due to inclusion being limited to non-representative subgroups of the RRMS population. Predictor determination and definitions were subject to variability in some studies meaning associations with outcome may not be generalizable. Blinding of outcome assessors to predictor information was poorly reported. Blinding outcome assessors to predictor information is especially for preventing bias when assessments are subjective and require interpretation, as is the case with many of the clinical outcomes employed here [21, 53].

Small sample sizes were common which limited the power of many models to examine multiple parameters or interactions between parameters [54]. Univariable analysis was often used to select predictors for model inclusion, which risks omitting predictors with important relationships with the outcome present only after adjustment for confounding covariates, and risks inclusion of covariates that hold no independent predictive power when other covariates are included [21, 55]. Model calibration was poorly reported. AUC- a measure of discrimination- was reported only in ten studies. Reported AUC values in ten model development studies ranged between 0.64–0.89 (0.7–0.8 is regarded as acceptable and 0.8–0.9 excellent [56]). Without reporting of calibration and discrimination it is challenging to quantify model accuracy [21].

The majority of studies did not perform internal validation. Models without internal validation may be at risk of misspecification (e.g. overfitting to development data sets) [57]. External validation was only reported in three studies [35,36,38] and performance statistics were only presented in one of these [36]. Lack of reporting of external validation and model performance therein undermines use of model in different patients [58]. No studies performed impact analysis, an essential step which quantifies changes in clinician behaviour, outcomes and cost-effectiveness of implementing models and provides an evidence base for clinical practice [59]. None of the studies incorporated clinical, radiological, demographic, lifestyle and biomarker predictors though independently, each of these has been demonstrated to show predictive power. As such, PROBAST assessment has identified areas for improvement in order to limit risk of bias in future studies.

In summary, issues of applicability and methodological quality limit the application of the studied models.

Future perspectives

The present study does not investigate fatigue and cognitive impairment as outcomes. These symptoms are increasingly recognised as contributors to morbidity in MS [60]. Inclusion of these factors was beyond the scope of this review but they should be researched further and may be worthy of inclusion in future attempts to construct predictive models. An improved understanding of the underlying pathobiological and molecular mechanism(s) of MS is likely to lead to a range of biomarkers that may feature in future predictive models. Differential gene transcription levels have been shown to predict interferon beta responsiveness in RRMS [61]. RNA profiling can identify patients with different levels of disease activity [62]. Single Molecule Array (SIMOA) technology offers increasingly accurate quantification of biomarkers such as neurofilament light chain [63, 64]. Imaging measures also show promise. Atrophied T2 lesion volume, a result of both inflammatory and degenerative processes, has been identified as a predictor of future disease activity in RRMS [65]. Ultra-high field (7 tesla) MRI shows promise in longitudinal investigation of multiple sclerosis lesions [66].

For individuals with newly diagnosed RRMS, reliable prognostic models are urgently needed. However, with a growing number of promising biomarkers, improvements in capabilities in novel imaging techniques, and increased understanding of the demographic, clinical, and immunological basis of MS heterogeneity, large well-powered cohorts will be necessary in order to have sufficient power to combine these predictive modalities into clinically useful tools. Persons newly diagnosed with RRMS face uncertainty regarding future disease course and the effect of treatment. Methodologically sound models developed in appropriate patient populations are vital to improve prognostication and inform therapeutic decision-making.

Supporting information

S1 Checklist.

(DOC)

S1 File. CHARMS protocol. Design of the systematic review based on Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist.

(DOCX)

S2 File. Search strategy.

(DOCX)

S3 File. Diagnostic criteria.

(DOCX)

S4 File. CHARMS data extraction.

(DOCX)

S5 File. Percentage of study participants on DMTs.

(DOCX)

S6 File. Events per variable per study and calculation method.

(DOCX)

Acknowledgments

We thank Marshall Dozier, University of Edinburgh Academic Support Librarian, for her invaluable advice and guidance on search strategy.

Author Contributions

Conceptualization: Fraser S. Brown, Patrick K. A. Kearns, Peter Connick, Siddharthan Chandran.

Data curation: Fraser S. Brown.

Formal analysis: Fraser S. Brown, Stella A. Glasmacher.

Investigation: Fraser S. Brown, Stella A. Glasmacher, Patrick K. A. Kearns, Siddharthan Chandran.

Methodology: Fraser S. Brown, Stella A. Glasmacher, Patrick K. A. Kearns, Peter Connick, Siddharthan Chandran.

Project administration: Fraser S. Brown, Stella A. Glasmacher.

Resources: Siddharthan Chandran.

Supervision: Patrick K. A. Kearns, Peter Connick, Siddharthan Chandran.

Writing – original draft: Fraser S. Brown.

Writing – review & editing: Fraser S. Brown, Stella A. Glasmacher, Patrick K. A. Kearns, Niall MacDougall, David Hunt, Peter Connick, Siddharthan Chandran.

References

1. Comi G, Radaelli M, Soelberg Sørensen P. Evolving concepts in the treatment of relapsing multiple sclerosis. *Lancet*. 2017;1; 389: 1347–1356. [https://doi.org/10.1016/S0140-6736\(16\)32388-1](https://doi.org/10.1016/S0140-6736(16)32388-1) PMID: 27889192
2. Langer-Gould A, Popat RA, Huang SM, Cobb K, Fontoura P, Gould MK et al. Clinical and demographic predictors of long-term disability in patients with relapsing-remitting multiple sclerosis: a systematic review. *Arch Neurol*. 2006; 63(12): 1686–91. <https://doi.org/10.1001/archneur.63.12.1686> PMID: 17172607
3. Koch-Henriksen N, Sørensen PS. The changing demographic pattern of multiple sclerosis epidemiology. *Lancet Neurol*. 2010 9(5):520–32. [https://doi.org/10.1016/S1474-4422\(10\)70064-8](https://doi.org/10.1016/S1474-4422(10)70064-8) PMID: 20398859
4. Weinshenker BG, Bass B, Rice GP, Noseworthy J, Carriere W, Baskerville J et al. The natural history of multiple sclerosis: a geographically based study. 2. Predictive value of the early clinical course. *Brain*. 1989; 112(6): 1419–28.
5. Amato MP, Ponziani G, Bartolozzi ML, Siracusa G. A prospective study on the natural history of multiple sclerosis: clues to the conduct and interpretation of clinical trials. *J Neurol Sci*. 1999;15; 168(2): 96–106. [https://doi.org/10.1016/s0022-510x\(99\)00143-4](https://doi.org/10.1016/s0022-510x(99)00143-4) PMID: 10526190
6. Trojano M, Avolio C, Manzari C, Calò A, De Robertis F, Serio G, et al. Multivariate analysis of predictive factors of multiple sclerosis course with a validated method to assess clinical events. *J Neurol Neurosurg Psychiatry*. 1995; 58(3): 300–6. <https://doi.org/10.1136/jnnp.58.3.300> PMID: 7897410
7. Kantarci O, Siva A, Eraksoy M, Karabudak R, Süttaş N, Ağaoğlu J et al. Survival and predictors of disability in Turkish MS patients. Turkish Multiple Sclerosis Study Group (TUMSSG). *Neurology* 1998; 51(3): 765–72. <https://doi.org/10.1212/wnl.51.3.765> PMID: 9748024
8. Lukas C, Minneboo A, de Groot V, Moraal B, Knol DL, Polman Chet et al. Early central atrophy rate predicts 5 year clinical outcome in multiple sclerosis. *J Neurol Neurosurg Psychiatry*. 2010; 81(12): 1351–6. <https://doi.org/10.1136/jnnp.2009.199968> PMID: 20826873
9. Kappos L, Moeri D, Radue EW, Schoetzau A, Schweikert K, Barkhof F et al. Predictive value of gadolinium-enhanced magnetic resonance imaging for relapse rate and changes in disability or impairment in multiple sclerosis: a meta-analysis. Gadolinium MRI Meta-analysis Group. *Lancet*. 1999;20; 353(9157): 964–9. [https://doi.org/10.1016/s0140-6736\(98\)03053-0](https://doi.org/10.1016/s0140-6736(98)03053-0) PMID: 10459905
10. Enzinger C, Fuchs S, Pichler A, Wallner-Blazek M, Khalil M, Langkammer C et al. Predicting the severity of relapsing-remitting MS: the contribution of cross-sectional and short-term follow-up MRI data. *Mult Scler*. 2011; 17(6): 695–701. <https://doi.org/10.1177/1352458510394454> PMID: 21228028

11. Horakova D, Dwyer MG, Havrdova E, Cox JL, Dolezal O, Bergsland Net al. Gray matter atrophy and disability progression in patients with early relapsing–remitting multiple sclerosis: A 5-year longitudinal study. *J Neurol Sci* 2009; 282: 112–119. <https://doi.org/10.1016/j.jns.2008.12.005> PMID: 19168190
12. Koudriatseva T, Thompson AJ, Fiorelli M. Gadolinium-enhanced MRI predicts clinical and MRI disease activity in relapsing–remitting multiple sclerosis. *J Neurol Neurosurg Psychiatry* 1997; 67: 285–287.
13. Simpson S Jr, Taylor B, Blizzard L, Ponsonby AL, Pittas F, Tremlett Het al. Higher 25-hydroxyvitamin D is associated with lower relapse risk in multiple sclerosis. *Ann Neurol*. 2010; 68(2): 193–203. <https://doi.org/10.1002/ana.22043> PMID: 20695012
14. Perini P, Ranzato F, Calabrese M, Battistin L, Gallo P. Intrathecal IgM production at clinical onset correlates with a more severe disease course in multiple sclerosis. *J Neurol Neurosurg Psychiatry*. 2006; 77(8): 953–5. <https://doi.org/10.1136/jnnp.2005.086116> PMID: 16574727
15. Villar LM, Masjuan J, González-Porqué P, Plaza J, Sádaba MC, Roldán Eet al. Intrathecal IgM synthesis is a prognostic factor in multiple sclerosis. *Ann Neurol*. 2003; 53(2): 222–6. <https://doi.org/10.1002/ana.10441> PMID: 12557289
16. Sundström P, Nyström L. Smoking worsens the prognosis in multiple sclerosis. *Mult Scler*. 2008; 14(8): 1031–5. <https://doi.org/10.1177/1352458508093615> PMID: 18632778
17. Ramanujam R, Hedström AK, Manouchehrinia A, Alfredsson L, Olsson T, Bottai M, et al. Effect of smoking cessation on multiple sclerosis prognosis. *JAMA Neurol*. 2015; 72(10): 1117–23. <https://doi.org/10.1001/jamaneurol.2015.1788> PMID: 26348720
18. Manouchehrinia A, Hedström AK, Alfredsson L, Olsson T, Hillert J, Ramanujam R. Association of Pre-Disease Body Mass Index With Multiple Sclerosis Prognosis. *Front Neurol*. 2018 11; 9:232. <https://doi.org/10.3389/fneur.2018.00232> PMID: 29867705
19. Havas J, Leray E, Rollot F, et al. Predictive medicine in multiple sclerosis: A systematic review. *Mult Scler Relat Disord*. 2020; 40:101928. <https://doi.org/10.1016/j.msard.2020.101928> PMID: 32004856
20. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DGet al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med*. 2014;14; 11(10): e1001744. <https://doi.org/10.1371/journal.pmed.1001744> PMID: 25314315
21. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med*. 2019;1; 170(1): 51–58. <https://doi.org/10.7326/M18-1376> PMID: 30596875
22. Agosta F, Rovaris M, Pagani E, Sormani MP, Comi G, Filippi M. Magnetization transfer MRI metrics predict the accumulation of disability 8 years later in patients with multiple sclerosis. *Brain* 2006; 129: 2620–27. <https://doi.org/10.1093/brain/awl208> PMID: 16951409
23. Bakshi R, Neema M, Healy BC, Liptak Z, Betensky RA, Buckle GJet al. Predicting clinical progression in multiple sclerosis with the magnetic resonance disease severity scale. *Archives of Neurology* 2008; 65(11): 1449–53. <https://doi.org/10.1001/archneur.65.11.1449> PMID: 19001162
24. Barkhof F, Held U, Simon JH, Daumer M, Fazekas F, Filippi M et al. Predicting gadolinium enhancement status in MS patients eligible for randomized clinical trials. *Neurology* 2005; 65(9): 1447–54. <https://doi.org/10.1212/01.wnl.0000183149.87975.32> PMID: 16275834
25. Bejarano B, Bianco M, Gonzalez-Moron D, Sepulcre J, Goñi J, Arcocha J et al. Computational classifiers for predicting the short-term course of Multiple sclerosis. *BMC Neurology* 2011;7; 11: 67. <https://doi.org/10.1186/1471-2377-11-67> PMID: 21649880
26. Bergamaschi R, Berzuini C, Romani A, et al. Predicting secondary progression in relapsing-remitting multiple sclerosis: a Bayesian analysis. *J Neurol Sci*. 2001; 189(1–2): 13–21. [https://doi.org/10.1016/s0022-510x\(01\)00572-x](https://doi.org/10.1016/s0022-510x(01)00572-x) PMID: 11535229
27. Bergamaschi R, Quaglini S, Trojano M, Amato MP, Tavazzi E, Paolicelli Det al. Early prediction of the long term evolution of multiple sclerosis: The Bayesian Risk Estimate for Multiple Sclerosis (BREMS) score. *J Neurol Neurosurg Psychiatry*. 2007; 78(7): 757–59. <https://doi.org/10.1136/jnnp.2006.107052> PMID: 17220286
28. Bergamaschi R, Montomoli C, Mallucci G, Lugaesi A, Izquierdo G, Grand'Maison F et al. BREMSO: a simple score to predict early the natural course of multiple sclerosis. *Eur J Neurol* 2015; 22(6): 981–9. <https://doi.org/10.1111/ene.12696> PMID: 25808578
29. de Groot V, Beckerman H, Uitdehaag BM, Hintzen RQ, Minneboo A, Heymans MWet al. Physical and Cognitive Functioning After 3 Years Can Be Predicted Using Information From the Diagnostic Process in Recently Diagnosed Multiple Sclerosis. *Arch Phys Med Rehabil*. 2009; 90(9): 1478–88. <https://doi.org/10.1016/j.apmr.2009.03.018> PMID: 19735774

30. Dekker I, Eijlers AJC, Popescu V, Balk LJ, Vrenken H, Wattjes MP et al. Predicting clinical progression in multiple sclerosis after 6 and 12 years. *Eur J Neurol* 2019; 26(6): 893–902. <https://doi.org/10.1111/ene.13904> PMID: 30629788
31. Filippi M, Preziosa P, Copetti M, Riccitelli G, Horsfield MA, Martinelli Vet al. Grey matter damage predicts the accumulation of disability 13 years later in MS. *Neurology* 2012; 18(4): 34–35.
32. Gauthier SA, Mandel M, Guttmann CR, Glanz BI, Khoury SJ, Betensky RA et al. Predicting short-term disability in multiple sclerosis. *Neurology* 2007; 68(24): 2059–65. <https://doi.org/10.1212/01.wnl.0000264890.97479.b1> PMID: 17562826
33. Held U, Heigenhauser L, Shang C, Kappos L, Polman C; Sylvia Lawry Centre for MS Research. Predictors of relapse rate in MS clinical trials. *Neurology* 2005; 65(11): 1769–73. <https://doi.org/10.1212/01.wnl.0000187122.71735.1f> PMID: 16344520
34. Liguori M, Meier DS, Hildenbrand P, Healy BC, Chitnis T, Baruch NF et al. One year activity on subtraction MRI predicts subsequent 4 year activity and progression in multiple sclerosis. *J Neurol Neurosurg Psychiatry*. 2011; 82(10): 1125–31. <https://doi.org/10.1136/jnnp.2011.242115> PMID: 21429902
35. Mandrioli J, Sola P, Bedin R, Gambini M, Merelli EA multifactorial prognostic index in multiple sclerosis—Cerebrospinal fluid IgM oligoclonal bands and clinical features to predict the evolution of the disease. *J Neurol* 2008; 255(7): 1023–31. <https://doi.org/10.1007/s00415-008-0827-5> PMID: 18535872
36. Manouchehrinia A, Zhu F, Piani-Meier D, Lange M, Silva DG, Carruthers Ret al. Predicting risk of secondary progression in multiple sclerosis: A nomogram. *Mult Scler* 2019; 25(8): 1102–12. <https://doi.org/10.1177/1352458518783667> PMID: 29911467
37. Margaritella N, Mendozzi L, Garegnani M, Colicino E, Gilardi E, Deleonardis Let al. Sensory evoked potentials to predict short-term progression of disability in multiple sclerosis. *Neurol Sci*. 2012; 33(4): 887–92. <https://doi.org/10.1007/s10072-011-0862-3> PMID: 22120189
38. Margaritella N, Mendozzi L, Garegnani M, Nemni R, Colicino E, Gilardi E et al. Exploring the predictive value of the evoked potentials score in MS within an appropriate patient population: A hint for an early identification of benign MS? *BMC Neurol* 2012; 22; 12: 80. <https://doi.org/10.1186/1471-2377-12-80> PMID: 22913733
39. Mesaros S, Rocca MA, Sormani MP, Charil A, Comi G, Filippi M. Clinical and conventional MRI predictors of disability and brain atrophy accumulation in RRMS. A large scale, short-term follow-up study. *J Neurol* 2008; 255(9): 1378–83. <https://doi.org/10.1007/s00415-008-0924-5> PMID: 18584233
40. Minneboo A, Jasperse B, Barkhof F, Uitdehaag BM, Knol DL, de Groot V et al. Predicting short-term disability progression in early multiple sclerosis: added value of MRI parameters. *J Neurol Neurosurg Psychiatry*. 2008; 79(8): 917–23. <https://doi.org/10.1136/jnnp.2007.124123> PMID: 18077480
41. Popescu V, Agosta F, Hulst HE, Sluimer IC, Knol DL, Sormani MP et al. Brain atrophy and lesion load predict long term disability in multiple sclerosis. *J Neurol Neurosurg Psychiatry*. 2013; 84(10): 1082–91. <https://doi.org/10.1136/jnnp-2012-304094> PMID: 23524331
42. Ramsaransing GSM, De Keyser J. Predictive value of clinical characteristics for 'benign' multiple sclerosis. *Eur J Neurol*. 2007; 14(8): 885–9. PMID: 17662009
43. Runmarker B, Andersson C, Odén A, Andersen O. Prediction of outcome in multiple sclerosis based on multivariate models. *J Neurol* 1994; 241(10): 597–604. <https://doi.org/10.1007/BF00920623> PMID: 7836963
44. Schlaeger R, D'Souza M, Schindler C, Grize L, Dellas S, Radue EW et al. Prediction of long-term disability in multiple sclerosis. *Mult Scler* 2012; 18(1): 31–8. <https://doi.org/10.1177/1352458511416836> PMID: 21868486
45. Schlaeger R, Schindler C, Grize L, Dellas S, Radue EW, Kappos Let al. Combined visual and motor evoked potentials predict multiple sclerosis disability after 20 years. *Mult Scler* 2013; 19(11): 196–97.
46. Skoog B, Tedeholm H, Runmarker B, et al. Continuous prediction of secondary progression in the individual course of multiple sclerosis. *Mult Scler Relat Disord*. 2014; 3(5): 584–92. <https://doi.org/10.1016/j.msard.2014.04.004> PMID: 26265270
47. Sormani MP, Rovaris M, Comi G, Filippi M. A composite score to predict short-term disease activity in patients with relapsing-remitting MS. *Neurology* 2007; 69(12): 1230–35. <https://doi.org/10.1212/01.wnl.0000276940.90309.15> PMID: 17875911
48. Uher T, Vaneckova M, Sobisek L, Tyblova M, Seidl Z, Krasensky J et al. Combining clinical and magnetic resonance imaging markers enhances prediction of 12-year disability in multiple sclerosis. *Mult Scler* 2017; 23(1): 51–61. <https://doi.org/10.1177/1352458516642314> PMID: 27053635
49. von Gumberg J, Mahmoudi M, Young K, Schippling S, Martin R, Heesen Cet al. Short-term MRI measurements as predictors of EDSS progression in relapsing-remitting multiple sclerosis: Grey matter atrophy but not lesions are predictive in a real-life setting. *PeerJ* 2016; 2016(9)

50. Weideman AM, Barbour C, Tapia-Maltos MA, Tran T, Jackson K, Kosa Pet al. New multiple sclerosis disease severity scale predicts future accumulation of disability. *Front Neurol* 2017;10; 8: 598. <https://doi.org/10.3389/fneur.2017.00598> PMID: 29176958
51. Weinshenker BG, Rice GP, Noseworthy JH, Carriere W, Baskerville J, Ebers GC. The natural history of multiple sclerosis: a geographically based study. 3. Multivariate analysis of predictive factors and models of outcome. *Brain* 1991; 114(Pt 2): 1045–56.
52. Brownlee WJ, Swanton JK, Altmann DR, Ciccarella O, Miller DH. Earlier and more frequent diagnosis of multiple sclerosis using the McDonald criteria. *J Neurol Neurosurg Psychiatry*. 2015 86(5): 584–5. <https://doi.org/10.1136/jnnp-2014-308675> PMID: 25412872
53. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA*. 1997. 12; 277(6): 488–94. PMID: 9020274
54. Steyerberg EW, Eijkemans MJ, Harrell FE Jr, Habbema JD. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. *Med Decis Making*. 2001 21(1): 45–56. <https://doi.org/10.1177/0272989X0102100106> PMID: 11206946
55. Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J Clin Epidemiol*. 1996; 49(8): 907–16. [https://doi.org/10.1016/0895-4356\(96\)00025-x](https://doi.org/10.1016/0895-4356(96)00025-x) PMID: 8699212
56. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol*. 2010 Sep; 5(9): 1315–6. <https://doi.org/10.1097/JTO.0b013e3181ec173d> PMID: 20736804
57. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ*. 2009;31; 338: b604. <https://doi.org/10.1136/bmj.b604> PMID: 19336487
58. Altman DG, Vergouwe Y, Royston P, Moons KGM. Validating a prognostic model. *BMJ*. 2009; 338: b605. <https://doi.org/10.1136/bmj.b605> PMID: 19477892
59. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009 4; 338: b606. <https://doi.org/10.1136/bmj.b606> PMID: 19502216
60. Penner IK. Evaluation of cognition and fatigue in multiple sclerosis: daily practice and future directions. *Acta Neurol Scand*. 2016;134; 200: 19–23.
61. Baranzini SE, Madireddy LR, Cromer A, D'Antonio M, Lehr L, Beelke M, et al. Prognostic biomarkers of IFN β therapy in multiple sclerosis patients. *Mult Scler*. 2015 Jun; 21(7):894–904. <https://doi.org/10.1177/1352458514555786> PMID: 25392319
62. Ottoboni L, Keenan BT, Tamayo P, Kuchroo M, Mesirov JP, Buckle GJ, et al. An RNA profile identifies two subsets of multiple sclerosis patients differing in disease activity. *Sci Transl Med*. 2012 Sep 26; 4(153):153ra131. <https://doi.org/10.1126/scitranslmed.3004186> PMID: 23019656
63. Siller N, Kuhle J, Muthuraman M, Barro C, Uphaus T, Groppa S et al. Serum neurofilament light chain is a biomarker of acute and chronic neuronal damage in early multiple sclerosis. *Mult Scler*. 2019 25(5): 678–686. <https://doi.org/10.1177/1352458518765666> PMID: 29542376
64. Ferraro D, Guicciardi C, De Biasi S, Pinti M, Bedin R, Camera V, et al. Plasma neurofilaments correlate with disability in progressive multiple sclerosis patients. *Acta Neurol Scand*. 2020 Jan; 141(1):16–21. <https://doi.org/10.1111/ane.13152> PMID: 31350854
65. Zivadinov R, Bergsland N, Dwyer MG. Atrophied brain lesion volume, a magnetic resonance imaging biomarker for monitoring neurodegenerative changes in multiple sclerosis. *Quant Imaging Med Surg*. 2018; 8(10):979–983. <https://doi.org/10.21037/qims.2018.11.01> PMID: 30598875
66. Chawla S, Kister I, Sinnecker T, Wuerfel J, Brisset JC, Paul F, et al. Longitudinal study of multiple sclerosis lesions using ultra-high field (7T) multiparametric MR imaging. *PLoS One*. 2018 Sep 13; 13(9): e0202918. <https://doi.org/10.1371/journal.pone.0202918> PMID: 30212476