Prognostic models for predicting clinical disease progression, worsening and activity in people with multiple sclerosis (Protocol)


www.cochranelibrary.com
ABSTRACT

Objectives

This is a protocol for a Cochrane Review (prognosis). The objectives are as follows:

Primary objectives

To identify and summarize multivariable prognostic models for quantifying the risk of clinical disease progression, worsening, and activity in MS.

To this end we will:

- describe the characteristics of the identified multivariable prognostic models, including prognostic factors considered and evaluation measures used;
- describe changes in outcome definitions, time frames, prognostic factors and statistical methods over time;
- qualitatively summarize the validation performance of the models;
- summarize model performance across external validation studies via meta-analysis, where possible;
- assess the risk of bias in the models;
- evaluate moderating effects on model performance by meta-regression, where possible; and
- make recommendations for future MS prognostic research.

Investigation of sources of heterogeneity between studies

We expect to find substantial heterogeneity, as diagnostic criteria for MS subtypes and available treatment options, as well as the technology used to assess disease activity, have evolved over time. Heterogeneity is also typically high in prognostic studies. We expect heterogeneity both between development studies and their corresponding validation studies for specific models and also between different development models for the same outcome. Potential sources of heterogeneity related to either or both of these include:

- case mix (e.g. age, gender, disease duration, treatment status);
- study design (e.g. follow-up time, source of data, outcome and prognostic factor definitions); and
• statistical analysis methods and reporting (e.g. number of prognostic factors included, traditional statistics versus machine learning, risk of bias, validation methods).

We will extract relevant information and will include a narrative summary of these potential sources of heterogeneity. We will further investigate heterogeneity using meta-regression, as detailed in the protocol.
BACKGROUND

Description of the health condition and context

Multiple sclerosis (MS) is a chronic inflammatory disease of the central nervous system (CNS) that usually begins in young adulthood and affects more than two million people worldwide (Adelman 2013). The course of MS varies greatly and is characterized by clinical, radiological, genetic and pathological heterogeneity. The exact aetiology of MS is still unclear, even though there are convincing arguments for an (auto-)immunopathogenesis. These include the neuropathological findings, the various analogies to the autoimmune animal models and, above all, the response of MS to various immunosuppressive therapies (Thompson 2018b). Genetic research also supports the (auto-)immunopathogenesis (Sawcer 2011). In addition, environmental factors have been shown to have a major influence on the development and course of the disease. Not only new imaging techniques, such as diffusion tensor imaging, but also neuropathological investigations have shown that, in addition to demyelination in MS, significant damage occurs to axons (Thompson 2018a).

The current diagnosis of MS and classification into subtypes are based on the modified 'McDonald criteria' (Thompson 2018b). In 80% of patients, MS initially presents as clinically isolated syndrome (CIS), which may convert to the relapsing-remitting multiple sclerosis (RRMS) subtype. Usually, RRMS later changes into a more severe course, secondary progressive MS (SPMS). About 15% of people with MS, however, have progressive disease from the start, the primary progressive subtype (PPMS) (Reich 2018). This classification was made at a time when few biomarkers were available and is still used in clinical practice, especially for communication with patients and the definition of study cohorts. Inflammation, indicated by relapses, as well as neurodegeneration can, however, be observed throughout the CNS in all subtypes of MS. From a clinical point of view, this means that even people with relapsing MS may have a gradual and relapse-independent worsening of disability in addition to accumulation of residual disability from relapses. Similarly, people with progressive MS may continue experiencing relapses. In order to address varying disease courses, Lublin 2014 differentiates between phenotypes within diagnostic subtypes.

Although MS is still incurable, pharmacological treatment for MS, particularly RRMS, has developed with increasing speed since the introduction of the first interferon-beta preparation more than 20 years ago. The arsenal of MS therapeutics includes various substances with different mechanisms of action. The main therapeutic goals of treatment are reduction in relapse rate; and delaying onset, slowing or stopping confirmed disability progression (Wingerchuk 2015). A network meta-analysis of randomized controlled trials compared the efficacy and safety of 15 treatments for the RRMS indication and reached the conclusion that there was insufficient evidence on treatments’ effect on disability worsening. Average treatment effects on relapse rates up to 24 months could be compared, however, and alemtuzumab, natalizumab, and fingolimod were found to be the best options (Tramacere 2015). Over- or undertreatment must be avoided and refraining from treatment is also an option. Filippini 2017, another review collating evidence on people with an initial clinical attack suggestive of MS, found that compared to placebo, early treatment had a small but uncertain benefit on disability worsening, relapses and conversion to clinically definite multiple sclerosis (CDMS). However the findings were based on low-quality evidence.

The current guidelines usually classify the available therapies as first- or second-line according to their efficacy and safety profiles and recommend selection of a therapy based on the patient’s disease activity and preferences, reserving efficacious but high-risk second-line medications for highly active disease (Montalban 2018; Rae-Grant 2018). The definition of highly active disease varies across the literature however (Freedman 2016; Diaz 2019), and how to define benign multiple sclerosis is unclear (Correale 2012). With its broad spectrum of clinical manifestations and an armamentarium of therapeutic approaches with different risk profiles, MS is a prime example of a disease which requires precision medicine. Prognostication of the future disease course in persons with MS could allow individualized decisions of treatment and disease management.

Description of the prognostic models

Many potential prognostic factors for predicting disease progression, worsening and activity have been identified. These include but are not limited to age, gender, body mass index, smoking history, and disease duration (Briggs 2019). Various biomarkers for MS have also been proposed, with magnetic resonance imaging (MRI) being the most commonly investigated (Rotstein 2019). Compared to using single prognostic factors, combining them may result in higher predictive power in this multifactorial disease. A prognostic score for clinically relevant outcomes could be developed based on a combination of clinical, paraclinical, environmental and demographic data (Kalinic 2017).

Many researchers have turned to multivariable models for prognostic prediction in people with MS. These models vary in many ways, including functional form, prognostic factors included, and the populations and time points to which they should be applied. For instance, the Bayesian Risk Estimate for Multiple Sclerosis score, predicting time to secondary progression in people diagnosed with RRMS, was developed by Bayesian Markov chain Monte Carlo simulation based on demographic and clinical factors recorded during the first year from disease onset (Bergamaschi 2001), whereas WottscheI and colleagues used support vector machines to predict conversion of CIS to CDMS in 1- and 3-year periods using MRI and clinical factors (WottscheI 2015).

To our knowledge, no widely accepted prognostic model for MS is being used in clinical practice yet. In order to be useful in patient care, these models need to be valid, clinically relevant and, preferably, integrated early in the diagnostic process (Steyerberg 2013). This would enable the optimisation of treatment decisions for individuals with MS. Regardless of the initial diagnosis, in order to balance the benefits and risks of therapy at the time point of treatment decisions, clinicians and patients should take into account the risk of disease activity and progression based on prognostic factors.

Health outcomes

Clinically relevant outcomes in MS are related to disability and relapses. Disease progression is characterized by a creeping and unstoppable accumulation of neurological deficits and usually manifests as a decrease in walking ability which occurs over varying time spans (Warnke 2019). Related to disability accrual, progression
is most commonly measured by the Expanded Disability Status Scale (EDSS). It has also been operationalized by the Multiple Sclerosis Functional Composite (MSFC). It has been suggested that the term ‘progression’ should only be used for the progressive subtypes of the disease and that relapse-related increases in disability be referred to as disease ‘worsening’ (Lublin 2014). The terminology used in the literature may not, however, exactly match these definitions. For the purposes of this review, increase in disability, either dependent on or independent of relapses, is relevant. Disability progression is ranked as the highest priority outcome by people with MS (Day 2018).

Relapses, another high-priority clinical outcome indicative of disease activity, manifest themselves as acute and transient episodes of neurological symptoms. Subacute episodes can lead to different neurological symptoms, which may remit completely in the course of the disease, but may also be accompanied by residual disability. Despite the fact that the primary outcome in most of the confirmatory clinical trials leading to market approval of RRMS therapies is relapse rate, it is not yet clear whether reduction in relapse rate is associated with a better overall prognosis. For example, the strength of the effect of the reduction in relapse rate with regard to the prevention of long-term disability progression remains controversial (Cree 2019).

Diagnostic transition to a more advanced disease stage, indicative of worsening and active disease, is also of interest to prognostic research in this field. For example, people initially diagnosed with CIS can meet the criteria of CDMS by experiencing another relapse or patients initially diagnosed with RRMS can be considered to have converted to a progressive course, SPMS, by retrospective assessment of sustained relapse-independent disability progression over a period of time, for example one year (Thompson 2018b).

We expect the definition, time span of prognostication, and measurement methods of clinical disease progression, worsening, and activity to be highly heterogeneous in the literature.

Why it is important to do this review

While there are more than 50 published Cochrane Reviews on interventions for MS or associated symptoms and another nine are ongoing, there is still no published Cochrane title on prognosis studies in MS (Cochrane 2020). Independent of the Cochrane network, Hempel and colleagues reviewed 59 studies of single modifiable prognostic factors in MS progression, such as vitamin D levels and smoking status (Hempel 2017). Also, Rio and Ruiz-Peña reviewed 45 studies that predict long-term treatment response by short-term response criteria, including both single factors and multivariable algorithms (Rio 2016). Both reviews found a wide variety of methods, timing, and outcome and prognostic factor definitions. Simultaneously, another research group has started working on a systematic review of multivariable prognostic models in relapsing-remitting MS (Brown 2019). Recently, another prognostic review title on factors predicting differential response to therapy in people with CIS or RRMS has been registered in Cochrane (Rahn 2019).

We aim to conduct a systematic Cochrane Review of multivariable prognostic models for predicting future clinical outcomes indicative of disease progression, worsening, and activity in people with multiple sclerosis at any time point following diagnosis. The results from this review will provide a long-sought systematic summary and assessment of the evidence base, which has thus far only been described by many non-systematic reviews (Miller 2008; Derfuss 2012; Gafson 2017; Rotstein 2019). Identified models could potentially provide MS patients and physicians with informative and clinically relevant models for making decisions on disease management.

Based on previous reviews, we expect the moment of prognostication and the time period for prediction to vary across the literature. Our review will include a discussion of changes in prognostic factors and methods over time. This study could identify changes in prognostic factors used to make predictions. It would also provide a comprehensive systematic assessment of the risk of bias in the models in the literature. Where possible, we will perform meta-analysis to summarize model performance. This review will also form a solid basis from which to make recommendations for future prognosis research in MS.

OBJECTIVES

Primary objectives

To identify and summarize multivariable prognostic models for quantifying the risk of clinical disease progression, worsening, and activity in MS.

To this end we will:

- describe the characteristics of the identified multivariable prognostic models, including prognostic factors considered and evaluation measures used;
- describe changes in outcome definitions, time frames, prognostic factors and statistical methods over time;
- qualitatively summarize the validation performance of the models;
- summarize model performance across external validation studies via meta-analysis, where possible;
- assess the risk of bias in the models;
- evaluate moderating effects on model performance by meta-regression, where possible; and
- make recommendations for future MS prognostic research.

Investigation of sources of heterogeneity between studies

We expect to find substantial heterogeneity, as diagnostic criteria for MS subtypes and available treatment options, as well as the technology used to assess disease activity, have evolved over time. Heterogeneity is also typically high in prognostic studies. We expect heterogeneity both between development studies and their corresponding validation studies for specific models and also between different development models for the same outcome. Potential sources of heterogeneity related to either or both of these include:

- case mix (e.g. age, gender, disease duration, treatment status);
- study design (e.g. follow-up time, source of data, outcome and prognostic factor definitions); and
- statistical analysis methods and reporting (e.g. number of prognostic factors included, traditional statistics versus machine learning, risk of bias, validation methods).
We will extract relevant information and will include a narrative summary of these potential sources of heterogeneity. We will further investigate heterogeneity using meta-regression, as detailed in the protocol.

**METHODS**

**Criteria for considering studies for this review**

Aiming to refine the eligibility criteria and ensure a common understanding among the review authors, we conducted a pilot title and abstract screening with a random subset of results from the draft search strategy. This was followed by full-text screening of titles marked for inclusion. These criteria are described by the PICOTS table below and in the following sections.

<table>
<thead>
<tr>
<th>Population</th>
<th>Adults with multiple sclerosis, including all subtypes (CIS, RRMS, SPMS, PPMS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
<td>All multivariable prognostic models and their validation studies</td>
</tr>
<tr>
<td>Comparator</td>
<td>There are no comparators in this review</td>
</tr>
<tr>
<td>Outcome</td>
<td>Clinical disease progression, worsening and activity which are measured based on disability, relapses, or conversion to a more advanced disease subtype</td>
</tr>
<tr>
<td>Timing</td>
<td>The models are to be used anytime following diagnosis for predicting future disease course</td>
</tr>
<tr>
<td>Setting</td>
<td>Any clinical setting where people with multiple sclerosis receive medical care</td>
</tr>
</tbody>
</table>

**Types of studies**

We will include studies that report development or validation of multivariable prognostic models of future disease progression in people with MS. We will include longitudinal study types, including cohort studies — either prospective or retrospective in nature — case-control studies, and randomized controlled trials. We will include studies based on both primary use and secondary use of data. We will only consider a study eligible if it aims to develop or validate a multivariable prognostic model.

We will exclude studies investigating only the clinical impact of prognostic models because the effect measures they report and their design, usually comparative, are different from prognostic model development and validation studies (Riley 2019).

**Targeted population**

We will include studies in adults with diagnosis of multiple sclerosis, irrespective of the subtype or treatment status. We will include clinically isolated syndrome, relapsing, and progressive subtypes.

**Types of prognostic models**

We will include studies that develop, validate, extend or update multivariable prognostic models. We will exclude studies which do not include prediction of future outcomes in individuals. We will exclude studies predicting outcomes only based on single prognostic factors. We will also exclude models that focus on predicting treatment response, either beneficial or harmful.

**Types of outcomes to be predicted**

We will include clinical outcomes indicating disease progression, worsening and activity. We will accept author definitions based on any of the following or a composite outcome containing at least one of them.

- Disability progression/worsening
- Relapse/attack
- Conversion to a more advanced disease subtype (e.g. CIS to RRMS, RRMS to SPMS)

We will include studies with any of the above outcomes, including models validated for a different outcome than originally developed. We will exclude models that predict only paraclinical outcomes, like laboratory measurements or images, because their translation to patient-relevant outcomes at the individual level are unclear and they are not prioritized by people with MS (Day 2018). We will also exclude studies predicting only quality of life outcomes, due to the difficulty of interpreting their clinical meaning.

We will not exclude any studies based on time point of prognostication or the time horizon for which the prognostic models apply because our preliminary review of the prognostic literature in MS indicates very liberally defined (in years) and heterogeneous time points of prognostication, both in relation to diagnosis and start of treatment. Defining a time horizon without a specific starting point and with multiple outcomes and disease subtypes is too restrictive to include different predictive models. For clinically meaningful outcomes, however, we expect disability progression/worsening and conversion from RRMS to SPMS to be measured in years. Relapses and conversion from CIS to RRMS are expected to be measured in months to a couple of years.

**Search methods for identification of studies**

**Electronic searches**

We will primarily identify eligible studies by searching the following databases and sources from 1996 to present.
We will restrict the search to studies published since 1996, the year of publication of an important tutorial on multivariable prognostic models in *Statistics in Medicine* (Harrell 1996). Before this time, methods were rapidly being developed but at the same time concerns over the misuse of statistical modelling for prediction of health outcomes were being raised (Diamond 1989; Concato 1993; Chatfield 1995). We consider Harrell 1996 to be a turning point, after which many papers (Altman 2000), textbooks (Harrell 2001; Steyerberg 2009), and guidelines (EQUATOR network and TRIPOD) addressing proper analysis and reporting became readily available (Simera 2008; Collins 2015). We will not impose language restrictions on the search.

We will use a search strategy for systematic reviews of prognostic models based on that of Geersing 2012 and further refined for this review. This modified filter is split into three subsearches: search terms specific for prediction or prognostic models (2a); terms for general models (2b); and the statistical search terms (2c).

The search is constructed in a dual manner with either two or three main concepts (detailed strategy for MEDLINE and Embase are in Appendix 1 and Appendix 2).

- (1) Multiple sclerosis and (2a) specific prognostic models or
- (1) Multiple sclerosis and (2b or 2c) general models or general statistical terms and (3) clinical outcomes

The search strategy uses combinations of thesaurus search terms (MeSH or Emtree) and free text search including synonyms in the title and abstract. Animal studies and studies in children are excluded.

### Searching other resources

We will handsearch the references of all included studies, as well as the references of all multiple sclerosis prognosis reviews we identify during screening. We will also track citations of the included studies. We will contact authors of all included studies for further information on unpublished or ongoing studies.

### Data collection

#### Selection of studies

We will carry out and document selection of eligible studies from the search results using the criteria outlined in ‘Criteria for considering studies for this review’ (see above) and the Rayyan web application (Ouzzani 2016). The criteria were piloted by several authors to ensure that the final criteria are applied similarly across the team.

Two review authors will perform title and abstract screening independently and we will include all titles marked for inclusion by at least one author in the full-text screening. We will perform, independently and in duplicate, assessment of full texts for their inclusion in the review or reasons for exclusion. We will resolve disagreements at the full-text screening stage through discussion and, if necessary, by involving a third review author.

We will summarize the study selection process with the flowchart from the PRISMA statement, showing the number of articles we identify; the number of articles we exclude, with reasons; and the total number of articles we include (Moher 2009).

#### Data extraction and management

Review authors will extract data from the included studies independently and in duplicate. We will collect data into a predefined, piloted electronic spreadsheet (see Appendix 3) based on the CHARMS checklist (Moons 2014) and TRIPOD guidelines (Collins 2015). We will resolve disagreements at the data extraction stage through discussion and, if necessary, by involving a third review author.

We will extract the following data.

- Article information (title, author, year)
- Data sources (e.g. prospective/retrospective, randomized trial participants, registry data, case-control)
- Participants (e.g. inclusion/exclusion criteria, recruitment method, country, number of centres, setting, participant description, treatments received, MS subtype)
- Outcomes (e.g. definitions and methods of measurement, duration of follow-up or time of outcome assessment, binding)
- Candidate predictors (e.g. demographics, MRI measures, relapses, EDSS, MSFC or other scores, duration of disease, predictor definitions and method/timing of measurement, handling/transformations)
- Sample size (e.g. number of participants, number of events, number of events per predictor)
- Missing data (e.g. number of participants with missing predictor or outcome data, handling of missing data)
- Model development (e.g. type of model, method for predictor consideration, model/predictor selection method, predictor selection criteria, shrinkage)
- Model performance and evaluation (e.g. discrimination, calibration, and classification measures with standard errors or confidence intervals, internal or external validation)
- Model presentation and interpretation (e.g. final model and secondary models including all predictors and weights, alternative presentations, exploratory research or for use in practice, comparison with other studies, generalisability, strengths and limitations)
- Factors related to readiness of model in practice (sufficient explanation to allow for further use, availability of predictors, external validation, prediction intervals, timing)

#### Assessment of risk of bias in included studies

Two review authors will perform ‘Risk of bias’ assessments independently and in duplicate using the prediction model ‘Risk of bias’ assessment tool (PROBAST) (Wolff 2019). In order to develop a common understanding of the form, two review authors (KR
and BIOS) have piloted the tool, discussed discrepancies in use, and agreed on rules for further use. When multiple models are developed in a single study or development and external validation of a model are included in the same study, we will assess the quality of each model or external validation separately. The tool consists of signalling questions in four domains, covering sources of possible bias due to data sources, definition or measurements of predictors and outcomes, sample size and analysis sets, model development, and model performance and evaluation. Each domain is graded as having low, high, or unclear risk of bias, which forms the basis for the overall risk of bias assessment for each identified prognostic model and validation (as described in Moons 2019). We will resolve disagreements at the domain or model/validation level through discussion and, if necessary, with the help of a third review author. We will present the assessments in a tabular form and summarize them graphically.

**Measures of association or predictive performance measures to be extracted**

An objective of this systematic review of multivariable prognostic models is to describe the changes in prognostic factors included in such models over time. To this end, we will extract the adjusted effect measures (e.g. risk ratio, odds ratio, hazard ratio, or beta coefficient) and their measure of uncertainty (e.g. standard error, variance, or confidence interval) for each prognostic factor included in models for each outcome. We will standardize direction of comparison across studies, if necessary.

We will extract the performance measures for discrimination – the model’s ability to distinguish between participants developing/not developing the outcome (e.g. C-statistic, area under curve [AUC], Harrel’s C index, Gonen and Heller’s concordance index, Royston-Sauerbrei D statistic); and calibration – the extent to which the expected outcomes and observed outcomes agree (e.g. calibration slope, calibration-in-the-large, and observed-to-expected (O:E) ratio).

We expect the C-statistic to be the most frequently reported measure of discrimination. It gives the proportion of randomly chosen pairs from the sample (one participant with the outcome and one without) in which the participant with the outcome has the higher predicted score/risk. A C-statistic of 0.5 means that the model's discriminative performance is no better than chance, while a value of 1.0 is considered perfect discrimination. We expect calibration to be reported infrequently, but the O:E ratio can be computed with the total observed number of events and the expected number of events, which are often reported in prognostic model studies. This measure is strongly related to calibration-in-the-large and measures the average O:E ratio across the range of predicted risks (Debray 2017). Values close to 1 indicate a well-calibrated model overall, but this does not rule out poor calibration in some subgroups.

**Dealing with missing data**

We will contact the corresponding authors via e-mail to request missing information required for quantitative data synthesis or risk of bias assessment. We will apply methods to derive missing performance measures (the C-statistic for discrimination and the O:E ratio for calibration) and their precision from the reported information, where possible (Parmar 1998; Tierney 2007; Debray 2017; Debray 2019). When otherwise similar studies report on different outcome horizons, we will apply extrapolation methods (Debray 2019).

**Assessment of heterogeneity**

We will measure the magnitude of heterogeneity in all meta-analyses across studies by the I² statistic (and its 95% confidence interval), which quantifies the amount of variability in the estimate due to heterogeneity rather than chance (Higgins 2002). We will interpret estimates below 40% as unimportant, estimates above 75% as considerable heterogeneity, and estimates in between as moderate to substantial heterogeneity (Deeks 2019). We will also assess heterogeneity using Tau² – the estimate of between-study variance in a random-effects meta-analysis. We will compute approximate 95% prediction intervals, providing ranges for potential model performance in new studies.

Further, we will use meta-regression to explore heterogeneity between:

- development studies and their corresponding validation studies for specific models;
- different development models for the same outcome.

We will perform these analyses only when the number of studies with reported or derivable performance measures is sufficient (at least 10 studies).

**Assessment of reporting deficiencies**

We will describe major deficiencies in reporting, as assessed by the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines (Collins 2015).

**Data synthesis**

This broad review intends to identify all prognostic models of clinical disease progression, worsening, and activity across all types of MS. We expect to identify numerous model validation studies, but only few external validation studies overall. We will (1) qualitatively summarize all identified multivariable models and prognostic factors included in these models for disease progression in MS; and, when possible, (2) meta-analyse prognostic model performance statistics for single models validated in several samples; and (3) perform meta-regression to explore possible sources of heterogeneity.

**Data synthesis and meta-analysis approaches**

We will summarize each study in tabular form, organized by MS subtype and model outcome. We will summarize models by: source, statistical method, follow-up time, derivation/validation, number of predictors, number of events relative to sample size, list of final predictors, internal validation method, external validation indicator, measure of association, and discrimination/calibration/accuracy measures.

To our knowledge, no single prognostic model is dominant in either the literature or in clinical practice. If, however, we find validation studies for a specific model, then we will additionally tabulate reported information on the participants, number of events, sample size, and performance measures. We will also perform a meta-analysis to pool the predictive performance measures of the specific model.
We will separately conduct meta-analyses with prognostic model performance measures as outcomes: the C-statistic for discrimination and the O:E ratio for calibration. We will employ random-effects meta-analysis to allow for differences in these models across studies due to both sampling variability and heterogeneity (Debray 2017). Hence, we will not assume that a single true prognostic model performance measure is shared across studies but rather we will interpret the summary measure of the random-effects meta-analysis as the mean of true performance which differs across studies. We will use a logit transformation for the C-statistics and a natural logarithm transformation for the O:E ratio, as suggested in Snell 2018. We will conduct meta-analysis if there are at least three studies evaluating a certain model. We will report all estimates with 95% confidence intervals as well as 95% prediction intervals to aid in interpretation of results in the context of a new study.

Subgroup analysis and investigation of heterogeneity

We will investigate potential sources of heterogeneity between a development study and its corresponding external validations if at least 10 studies for the specific model are available. Using random-effects meta-regression (Berkey 1995), we will assess the effect of the following factors on each of the transformed performance measures (logit C-statistic or log O:E ratio), separately.

- Diagnostic criteria (McDonald versus other)
- Percentage treated with disease-modifying therapies
- Publication year
- Prospective versus retrospective design

Additionally, we will investigate potential sources of heterogeneity between different models developed for the same outcome (for example, between all models for prediction of annualized relapse rate) using the same procedure as above, and using the following factors.

- Diagnostic criteria (McDonald versus other)
- Percentage treated with disease-modifying therapies
- Publication year
- Prospective versus retrospective design
- Traditional statistics versus machine-learning modelling approaches
- Whether imaging prognostic factors were used

Sensitivity analysis

We will also perform all the analyses excluding studies at high risk of bias (overall), as determined using PROBAST. If too few studies are available after exclusion of high-risk studies, we will not perform the sensitivity analyses.

Conclusions and summary of findings

The summary of findings will highlight the best performing models and models of practical use according to MS subtype and model outcomes. We will also highlight areas in need of improved reporting in the literature and make recommendations for future research. Currently, the GRADE framework has not yet been adapted for prognostic model research (Guyatt 2011). This is an ongoing project for the GRADE working group, and if it is finalized during our review, we will adhere to the adapted GRADE in our conclusions.

Acknowledgements

We would like to thank Graziella Filippini (Co-ordinating Editor) and Liliana Coco (former Managing Editor) of the Cochrane Multiple Sclerosis and Rare Diseases of the CNS Group for their ongoing support. We would also like to thank our peer reviewers for their helpful comments during the preparation of this protocol, including but not limited to, Nina Kreuzberger.
Additional references

Adelman 2013

Altman 2000

Bergamaschi 2001

Berkey 1995

Briggs 2019

Brown 2019

Chatfield 1995

Cochrane 2020
Cochrane Multiple Sclerosis and Rare Disease of the CNS. Our reviews. msrdcns.cochrane.org/our-review (accessed 23 February 2020).

Collins 2015

Concato 1993

Correale 2012

Cree 2019

Day 2018

Debray 2017

Debray 2019

Deeks 2019

Derfuss 2012

Diamond 1989

Diaz 2019
Prognostic models for predicting clinical disease progression, worsening and activity in people with multiple sclerosis [Protocol]

Cochrane Database of Systematic Reviews

Filippini 2017

Freedman 2016

Gafson 2017

Guyatt 2011

Harrell 1996

Harrell 2001

Hempel 2017

Higgins 2002

Kalincik 2017

Lublin 2014

Miller 2008

Moher 2009

Montalban 2018

Moons 2014

Moons 2019

Ouzzani 2016

Parmar 1998

Rae-Grant 2018

Rahn 2019
Rahn A. Factors predicting differential treatment response in people with clinically isolated syndrome or relapsing-remitting multiple sclerosis receiving immunotherapy (PROMISE) [title]. www.cochrane.org/title/factors-predicting-differential-

Reich 2018

Riley 2019

Rotstein 2019

Río 2016

Sawcer 2011
Sawcer S. The major cause of multiple sclerosis is environmental: genetics has a minor role--no. Multiple Sclerosis 2011; 17(10):1174-5. [DOI: 10.1177/1352458511421106]

Simera 2008

Snell 2018

Steyerberg 2009

Steyerberg 2013

Thomson 2018a

Thomson 2018b

Tierney 2007

Tramacere 2015

Warnke 2019

Wingerchuk 2016

Wolff 2019

Wottschel 2015

A P P E N D I C E S

Appendix 1. MEDLINE search strategy
Ovid MEDLINE® and Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Daily
Search Strategy:
# Concept Searches

1. **Concept 1**  
   **MS**  
   (exp Multiple Sclerosis/ OR ((multipl* OR disseminated OR insular) ADJ1 sclerosis).ti,ab.) NOT (animals NOT humans).sh. NOT (child NOT adult).sh.

2. **2a prognostic/prediction**  
   (exp Prognosis/ AND (exp disease progression/ OR exp Remission, Spontaneous/ OR exp Recurrence/) OR (predict OR prognos*).ti. OR ((predict* OR prognos*) ADJ3 (recurr- ence OR progression OR relaps* OR remission OR remitting OR 'multiple sclerosis' OR ms)).ti,ab. OR ((predict* OR prognos*) ADJ3 treat* ADJ3 response).ti,ab. OR ((predict* OR prognos*) ADJ3 disease ADJ3 activity).ti,ab.

3. **2b general models**  
   ([(model* OR decision* OR identif*) ADJ3 (history OR variable* OR multicomponent* OR multivariable* OR multinomial* OR covariate* OR criteria OR criterion OR scor* OR character- istic* OR finding* OR factor* OR rule*)].ti,ab. OR (decision* ADJ6  model*)).ti,ab.

4. **2c statistical terms**  
   ([(logistic OR statistic*) ADJ3 model*]).ti,ab. OR (decision*.ti,ab. AND exp models, statisti- cal/) OR (Strategy OR Discrimination OR Discriminate OR "c-statistic" OR "c statistic" OR "Area under the curve" OR AUC OR Calibration OR Indices OR index OR Algorithm OR Multivariable* OR multivariate* OR covariate* OR valid*).ti. OR (Strategy OR Discrimination OR Discriminate OR "c-statistic" OR "c statistic" OR "Area under the curve" OR AUC OR Calibration OR Indices OR index OR Algorithm OR Multivariable* OR multivariate* OR covariate* OR valid*).ab. AND (predict* OR prognos*).ti,ab.

5. **3 Outcomes**  
   ([(disease OR disability OR invalid* OR function* OR outcome OR impairment OR com- posite OR activity OR severity OR cognitive OR edss OR treatement OR ms OR 'multiple sclerosis' OR brems) ADJ6 (score* OR scal* OR status OR assess* OR index OR classification)].ti,ab). OR (clinical ADJ3 (assess* OR activity)).ti,ab. OR ((disease OR disable* OR risk OR calculate*) ADJ3 (course OR progression)).ti,ab. OR (relaps* ADJ3 (rate OR frequen* OR time OR prognos* OR predict*)).ti,ab. OR (clinical* ADJ3 decision*).ti,ab. OR ((ms OR cdm* OR 'multiple sclerosis') ADJ3 (develop* OR course OR progress* OR relaps* OR clinical*)).ti,ab.

1 AND 2 OR (1 AND (3 OR 4) AND 5)

### Appendix 2. Embase search strategy

**Ovid Embase**

Search Stratgey:

<table>
<thead>
<tr>
<th>#</th>
<th>Concept</th>
<th>Searches</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Concept 1</td>
<td>('multiple sclerosis'/exp/mj OR (ms/ti AND 'multiple sclerosis'/exp) OR (((multipl* OR disseminated OR insular) NEAR/1 sclerosis):ti,ab)) NOT [(animals)/lim NOT (humans)/lim] NOT (nonhuman'/exp NOT 'human'/exp) NOT 'animal model'/exp NOT ('child'/exp NOT 'adult'/exp) NOT [conference abstract]/lim</td>
</tr>
<tr>
<td>2</td>
<td>2a prognostic/prediction</td>
<td>('predictive value'/exp AND 'model'/exp) OR ('prognosis'/exp AND ('disease exacerbation'/exp OR 'recurrent disease'/exp OR 'recurrence risk'/exp OR 'relapse'/exp OR 'remission'/exp)) OR predict*:ti OR prognos*:ti OR ((predict* OR prognos*) NEAR/3 (recur* OR progress* OR relaps* OR remission OR remitting OR 'multiple sclerosis' OR ms):ti,ab OR ((predict* OR prognos*) NEAR/3 treat* NEAR/3 response):ti,ab OR ((predict* OR prognos*) NEAR/3 disease NEAR/3 activity):ti,ab)</td>
</tr>
</tbody>
</table>
Appendix 3. Data extraction form
Adapted from CHARMS checklist of Moons 2014.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Key items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study information</td>
<td>• Study identifier (last name of first author and publication year), citation</td>
</tr>
<tr>
<td></td>
<td>• Development with/without internal validation and/or external validation</td>
</tr>
<tr>
<td>Source of data</td>
<td>• Cohort, case-control, randomised trial, registry, routine care data</td>
</tr>
<tr>
<td></td>
<td>• Retrospective/prospective data collection</td>
</tr>
<tr>
<td>Participants</td>
<td>• Inclusion and exclusion criteria</td>
</tr>
<tr>
<td></td>
<td>• Recruitment method and details (location, number of centres, setting)</td>
</tr>
<tr>
<td></td>
<td>• Participant description (including disease duration, type of MS at prognostication, diagnostic criteria used)</td>
</tr>
<tr>
<td></td>
<td>• Details of treatments received</td>
</tr>
<tr>
<td></td>
<td>• Study dates</td>
</tr>
<tr>
<td>Outcomes to be predicted</td>
<td>• Definition and method of measurement of outcome</td>
</tr>
<tr>
<td></td>
<td>• Category of outcome measure (conversion to definite MS, conversion to progressive MS, relapses, progression score, composite)</td>
</tr>
<tr>
<td></td>
<td>• Was the same outcome definition and method of measurement used on all participants?</td>
</tr>
<tr>
<td></td>
<td>• Was the outcome assessed without knowledge of candidate predictors (i.e. blinded)?</td>
</tr>
<tr>
<td></td>
<td>• Were candidate predictors part of the outcome?</td>
</tr>
<tr>
<td></td>
<td>• Time of outcome occurrence or summary of duration of follow-up</td>
</tr>
<tr>
<td>Candidate predictors</td>
<td>• Number and type of predictors (e.g. demographics, symptoms, scores, CSF, imaging, electrophysiological, -omics, disease type)</td>
</tr>
<tr>
<td></td>
<td>• Definition and method for measurement of candidate predictors</td>
</tr>
</tbody>
</table>
(Continued)

- Timing of predictor measurement (e.g. at patient presentation, as diagnosis, at predefined intervals)
- Were predictors assessed blinded for outcome (if relevant)?
- Handling of predictors in the modelling (e.g. continuous, linear, (fractional) polynomial/spline/non-linear transformations, categorizations)

### Sample size
- Number of participants and number of events
- Number of events in relation to number to candidate predictors (EPV)
- Power of study assessed?

### Missing data
- Number of participants with any missing value
- Number of participants with missing values for each predictor
- Method for handling missing data (e.g. complete-case analysis, single imputation, multiple imputation)
- Loss to follow-up discussed?

### Model development
- Modelling method (e.g. logistic, survival, neural network, specific machine learning technique)
- Modelling assumptions satisfied?
- Was it a longitudinal model?
- Method for selection of predictors for inclusion in multivariable modelling (e.g. all candidate predictors, pre-selection based on unadjusted association with outcome, etc)
- Method for selection of predictors during multivariable modelling (e.g. full model approach, backward selection, forward selection)
- Criteria used for selection of predictors during multivariable modelling (e.g. P value, AIC, BIC)
- Shrinkage of predictor weights/regression coefficients (e.g. no shrinkage, uniform shrinkage, penalized estimation)

### Model performance
- Measure and estimate of calibration with confidence intervals (calibration plot, calibration slope, Hosmer-Lemeshow test)
- Measure and estimate of discrimination with confidence intervals (C-statistic, D-statistic)
- Log-rank used for discrimination (yes, no, not applicable)
- Measure and estimate of classification with confidence intervals (sens, spec, ppv, npv, net reclassification, accuracy rate (TP+TN/N), error rate (1-acc))
- Were a-priori cut points used for classification measures? (yes, no, NR, NA)
- Overall performance (R^2, Brier score, etc)

### Model evaluation
- If model development, model performance tested on development dataset only or on separate external validation
- If model development, method used for testing model performance on development dataset (random split of data, resampling methods e.g. bootstrap or crossvalidation, none)
- In case of poor validation, was model adjusted or updated (e.g. intercept recalibrated, predictor effects adjusted, new predictors added)?

### Results
- Multivariable model (e.g. basic, extended, simplified) presented, including predictor weights/regression coefficients, intercept, baseline survival; with standard errors or confidence intervals
- Any alternative presentation of the final prediction models (e.g. sum score, nomogram, score chart, predictions for specific risk subgroups with performance)
- Provide details on how risk groups were created, if done and the observed values at which they occur
- Comparison of the distribution of predictors (including missing data) for development and validation data sets
- If validation, is the same model used as presented in development (same intercept and weights, no dropping of variables, etc)?

### Interpretation and discussion
- Aim according to authors (abstract, discussion)
• Was the primary aim prediction of individual patient outcomes?
• Are the models interpreted as probably/confirmatory (model useful for practice) or probably/exploratory (more research is needed)?
• Comparisons made with other studies, discussion of generalisability, strengths, and limitations
• Suggested improvements for the future

Model readiness for practical use

• Sufficient explanation to allow for further use, availability of predictors, external validation, prediction intervals, timing

HISTORY

Protocol first published: Issue 5, 2020

CONTRIBUTIONS OF AUTHORS

<table>
<thead>
<tr>
<th>Task</th>
<th>Authors responsible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Draft the protocol</td>
<td>IOS, KAR, JH, MG, JB, UH, UM, AL, SS</td>
</tr>
<tr>
<td>Develop and run the search strategy</td>
<td>MG, KAR, IOS</td>
</tr>
<tr>
<td>Obtain copies of studies</td>
<td>IOS, KAR, MG</td>
</tr>
<tr>
<td>Select which studies to include</td>
<td>IOS, KAR, UH, JH, JB, UM, AL, SS</td>
</tr>
<tr>
<td>Extract data from the studies</td>
<td>IOS, KAR, UH, JH, JB, UM, AL, SS</td>
</tr>
<tr>
<td>Enter data into RevMan 5</td>
<td>IOS, KAR</td>
</tr>
<tr>
<td>Carry out the analysis</td>
<td>KAR, IOS</td>
</tr>
<tr>
<td>Interpret the analysis</td>
<td>KAR, IOS, JH, JB, UH, UM, AL, SS</td>
</tr>
<tr>
<td>Draft the final review</td>
<td>KAR, IOS, JH, JB, UH, UM, AL, SS</td>
</tr>
<tr>
<td>Update the review</td>
<td>UM, UH</td>
</tr>
</tbody>
</table>

DECLARATIONS OF INTEREST

• JH reports a grant for OCT research from the Friedrich-Baur-Stiftung and Merck, personal fees and non-financial support from Merck, Alexion, Celgene, Novartis, Roche, Santhera, Biogen, Heidelberg Engineering, Sanofi Genzyme and non-financial support of the Guthy-Jackson Charitable Foundation, all outside the submitted work.
• AL received financial compensation and/or travel support for lectures and advice from Biogen, Merck, Novartis, Teva, Genzyme, Bayer, Celgene and he is a co-founder of Cellerys and co-inventor on a patent held by the University of Zurich on the use of peptide-coupled cells for treatment of MS.
• SS has received non-personal compensation from Novartis, Merck and Roche for board memberships and invited lectures. He has received non-personal consultancy fees by Novartis. He is currently an employee of Roche, while holding his University position.
• UM consults for the BfArM (German Institute of Medicine and Medical products), receives payments from the French National Agency of Research (ANR) for reviewer activities, and has received several grants by the BMBF (Federal Ministry of Science and Research) and the DFG (German Research Organisation) for biometrical support in investigator-initiated clinical trials.
• IOS, KAR, MG, JB, UH: none known
SOURCES OF SUPPORT

Internal sources

- DIFUTURE Project at Ludwig-Maximilians-Universität München, Germany
  DIFUTURE is funded by the German Federal Ministry of Education and Research under 01ZZ1603[A-D] and 01ZZ1804[A-I].
- Clinical Research Priority Program (CRPP), University of Zurich, Switzerland
  PrecisionMS: Implementing Precision Medicine in Multiple Sclerosis

External sources

- No sources of support supplied