OPEN ACCESS

Original research

Predictors of early disability accumulation in newly diagnosed multiple sclerosis: clinical, imaging and cerebrospinal fluid measures

Markus Lauerer ,[1,2] Tun Wiltgen,[1,2] Carolin Brückner,[1] Christina Engl,[1] Katrin Giglhuber,[1] Sebastian Lambrecht,[1] Viola Pongratz,[1,2] Achim Berthele,[1] Christiane Gasperi,[1] Jan S Kirschke ,[3] Claus Zimmer,[3] Bernhard Hemmer ,[1,4] Mark Mühlau [1,2]

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/jnnp-2024-335037).

[1]Department of Neurology, School of Medicine and Health, Technical University of Munich, Munich, Germany
[2]TUM-Neuroimaging Center, School of Medicine and Health, Technical University of Munich, Munich, Germany
[3]Department of Neuroradiology, School of Medicine and Health, Technical University of Munich, Munich, Germany
[4]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany

Correspondence to
Dr Mark Mühlau; mark.muehlau@tum.de

Received 8 September 2024
Accepted 22 January 2025
Published Online First 17 February 2025

To cite: Lauerer M, Wiltgen T, Brückner C, *et al*. *J Neurol Neurosurg Psychiatry* 2025;**96**:900–907.

ABSTRACT

Background A growing arsenal of treatment options for relapsing multiple sclerosis (RMS) emphasises the need for early prognostic biomarkers. While evidence for individual markers exists, comprehensive analyses at the time of diagnosis are sparse.

Methods Brain and spinal cord lesion numbers, cerebrospinal fluid parameters, initial symptoms, and Expanded Disability Status Scale (EDSS) score were determined at the time of diagnosis. Confirmed disability accumulation (CDA), defined as a sustained EDSS increase over 6 months, was determined during a 5-year follow-up. All-subsets multivariable logistic regression was performed to identify predictors of CDA. Model performance was assessed via receiver operating characteristic analysis, and individual risks were calculated. Analyses were repeated with progression independent of relapse activity (PIRA) as an outcome.

Results 113/417 (27.1%) people with RMS experienced CDA on follow-up. Intrathecal IgG synthesis, a higher number of spinal cord lesions, age and polysymptomatic manifestation were identified as independent predictors of CDA. The resulting prediction model yielded an area under the curve (AUC) of 0.75 with a 95% CI of 0.70 to 0.80. Individuals exceeding the optimal thresholds for the three most significant predictors had a 61.8% likelihood of experiencing CDA, whereas those below all three thresholds had a CDA rate of 4.5%. The only significant baseline predictor differentiating PIRA from relapse-associated worsening was a higher number of spinal cord lesions (AUC=0.64, 95% CI 0.54 to 0.74).

Conclusions Intrathecal IgG synthesis, spinal cord lesion number, age and polysymptomatic manifestation are independent predictors of early CDA in newly diagnosed RMS.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Different clinical, imaging and cerebrospinal fluid biomarkers are known to be associated with confirmed disability accumulation (CDA) in relapsing multiple sclerosis (RMS). However, little is known about their comprehensive prognostic value at the time of diagnosis.

WHAT THIS STUDY ADDS

⇒ In a cohort of 417 people with RMS, intrathecal IgG synthesis, spinal cord lesion number, age and polysymptomatic manifestation were identified as independent predictors of CDA over a 5-year follow-up.

⇒ The individual risk of CDA was higher with each additional threshold exceeded for the different predictors.

⇒ Spinal cord lesion number was the only baseline predictor differentiating progression independent of relapse activity from relapse-associated worsening on follow-up.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ By relying on a number of routine diagnostic parameters, our findings may help in estimating the likelihood of disability accumulation in individuals newly diagnosed with RMS, thereby guiding early treatment decisions.

INTRODUCTION

Treatment options for multiple sclerosis (MS), a chronic inflammatory disorder of the central nervous system, have seen a dramatic expansion over the last three decades. The current landscape of disease-modifying therapies (DMTs) is characterised by a multitude of drugs with different modes of action, potencies and side effects. This abundance of choice confronts physicians with the conundrum of finding the best DMT for the individual patient. Since treatment initiation often takes place at the time of first diagnosis, reliable prognoses of disease courses at this early stage are crucial for an informed decision. Initial attempts at predicting the accumulation of disability in MS were made even before any form of DMT was available.[1] Since then, a considerable number of prognostic models have been suggested, employing a wide range of potential predictors, outcome measures and methodologies.[2]

Focusing on parameters available in clinical routine, associations with future disability accumulation have been reported for the number of both brain lesions[3–5] and spinal cord lesions[6–9] as well as cerebrospinal fluid (CSF) pleocytosis[10] and intrathecal Ig synthesis—assessed qualitatively through

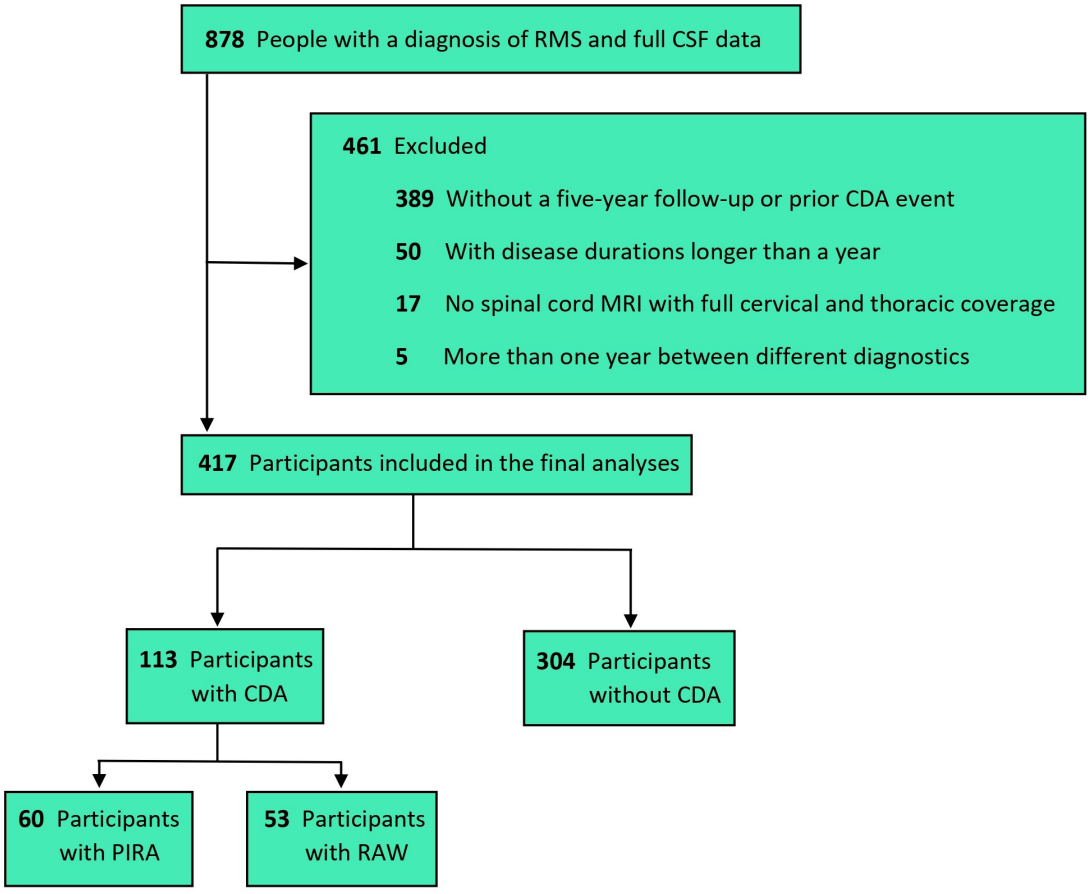

Figure 1 Participant flow chart. CDA, confirmed disability accumulation; CSF, cerebrospinal fluid; PIRA, progression independent of relapse activity; RAW, relapse-associated worsening; RMS, relapsing multiple sclerosis.

the detection of oligoclonal bands (OCB)[11–14] or quantitatively via Ig indices.[15–17] Furthermore, an association between disability accumulation and the anatomical localisation(s) of the demyelinating event leading to diagnosis has been observed.[1 18–21] As of yet, however, the combined prognostic value of these routine parameters at the time of diagnosis has not been comprehensively evaluated. Moreover, it has recently been shown that progression independent of relapse activity (PIRA) is responsible for a relevant part of disability accumulation, even in the early course of the disease.[22 23] Whether the combination of clinical parameters with routine diagnostic MRI and CSF measures can help in identifying individuals at risk of early PIRA is another question that currently remains unanswered.

In this retrospective study of 417 people recently diagnosed with relapsing MS (RMS), we investigated the prognostic value of several (para)clinical measures at baseline regarding confirmed disability accumulation (CDA) and PIRA over a 5-year period. To ensure the potential for clinical translation, we restricted our analyses to predictors available in diagnostic routine.

METHODS

Participants

Inclusion criteria of this retrospective, single-centre study were (1) diagnosis of RMS or clinically isolated syndrome with conversion to RMS on follow-up; (2) age between 18 and 60 years; (3) disease duration (defined as the interval between the first demyelinating event and completion of initial diagnostics) of less than 1 year; (4) brain MRI, spinal cord MRI (with full cervical and thoracic coverage), CSF data (including white cell count (WCC) and Ig synthesis/indices) and non-relapse Expanded Disability Status Scale (EDSS) score, all within a year of one another, and (5) clinical follow-up of at least 5 years for individuals who did not experience CDA earlier. External spinal cord imaging (ie, MRI not performed at our clinic) was considered for inclusion if it fulfilled our predefined criteria and was of sufficient quality. Data were assessed for eligibility from 2009 to 2023 (for the latest possible completion of follow-up).

Definition of disability accumulation and subtypes

CDA was defined as an EDSS increase of 1.5 for baseline EDSS scores of 0, an increase of 1 for baseline EDSS scores between 1.0 and 5.0 and an increase of 0.5 for baseline EDSS scores of 5.5 or higher; the increase had to be confirmed on clinical follow-up over at least 6 months. In order for a CDA event to be classified as PIRA, both the initial EDSS increase and the respective confirmation visit could not take place within 30 days before and 90 days after the onset of a relapse event. Otherwise, CDA was classified as relapse-associated worsening (RAW).[24]

MRI acquisition and processing

Spinal cord MRI was performed on three different 3 Tesla scanners (Philips Achieva dStream, Philips Ingenia, Siemens Magnetom Verio). All scans included 2D T2-weighted turbo spin echo sequences in sagittal and axial orientation with full cervical and thoracic coverage. The number of spinal cord lesions was taken from the respective neuroradiology report and visually checked for every scan by an experienced neurologist (ML). In

Table 1 Characteristics of study participants

	Total (n=417)	No CDA (n=304)	CDA (n=113)	P value
Age (years), median (IQR)	34.0 (26.6–41.2)	32.7 (26.6–39.4)	37.0 (26.6–45.1)	**0.018**
Sex (female), n (%)	270 (64.7)	188 (61.8)	82 (72.6)	**0.050**
EDSS, median (IQR)	1.0 (0–1.5)	1.0 (0–1.5)	1.0 (1.0–2.0)	0.073
Disease duration (days), median (IQR)	52.0 (29.0–80.0)	52.5 (30.0–76.3)	51.0 (25.0–87.0)	0.897
Imaging parameters				
Number of spinal lesions, median (IQR)	1.0 (0–3.0)	1.0 (0–2.0)	2.0 (1.0–5.0)	**<0.001**
Cervical lesion number, median (IQR)	1.0 (0–2.0)	0 (0–1.0)	1.0 (0–3.0)	**<0.001**
Thoracic lesion number, median (IQR)	0 (0–1.0)	0 (0–1.0)	1.0 (0–3.0)	**<0.001**
Number of brain lesions, median (IQR)	20.0 (10.0–44.0)	18.0 (9.0–37.0)	28.0 (14.0–61.0)	**<0.001**
CSF parameters				
WCC (x10^6/L), median (IQR)	6.0 (3.0–12.0)	5.0 (3.0–11.0)	7.0 (4.0–16.0)	**0.004**
Intrathecal IgG synthesis, n (%)	213 (51.1)	133 (43.8)	80 (70.8)	**<0.001**
Intrathecal IgM synthesis, n (%)	75 (18.0)	52 (17.1)	23 (20.4)	0.474
Intrathecal IgA synthesis, n (%)	28 (6.7)	21 (6.9)	7 (6.2)	1.000
IgG index, median (IQR)	0.72 (0.55–1.0)	0.65 (0.53–0.92)	0.87 (0.66–1.4)	**<0.001**
IgM index, median (IQR)	0.074 (0.049–0.15)	0.071 (0.047–0.14)	0.083 (0.058–0.17)	**0.015**
IgA index, median (IQR)	0.27 (0.24–0.32)	0.27 (0.23–0.32)	0.28 (0.24–0.32)	0.397
OCB, n (%)	371 (89.0)	261 (85.9)	110 (97.3)	**<0.001**
Functional systems affected by first demyelinating event				
Pyramidal, n (%)	82 (19.7)	47 (15.5)	35 (31.0)	**<0.001**
Cerebellar, n (%)	55 (13.2)	34 (11.2)	21 (18.6)	0.052
Brainstem, n (%)	99 (23.7)	69 (22.7)	30 (26.5)	0.438
Sensory, n (%)	198 (47.5)	135 (44.4)	63 (55.8)	**0.047**
Bowel/bladder, n (%)	18 (4.3)	10 (3.3)	8 (7.1)	0.105
Visual, n (%)	148 (35.5)	110 (36.2)	38 (33.6)	0.647
Polysymptomatic, n (%)	149 (35.7)	92 (30.3)	57 (50.4)	**<0.001**
Longitudinal parameters				
EDSS visits per year, median (IQR)	1.6 (1.2–2.4)	1.6 (1.2–2.0)	2.4 (1.7–4.4)	**<0.001**
Time on DMT (% of follow-up), median (IQR)	92.9 (64.3–98.4)	94.5 (66.7–98.5)	91.5 (58.3–97.9)	0.264
HET during follow-up, n (%)	104 (24.9)	71 (23.7)	32 (28.3)	0.373
No DMT during follow-up, n (%)	53 (12.7)	40 (13.2)	13 (11.5)	0.742
Time to CDA (years), median (IQR)	n.a.	n.a.	1.9 (1.1–3.3)	n.a.

Significant (p<0.05) differences between groups are given in bold.
CDA, confirmed disability accumulation; CSF, cerebrospinal fluid; DMT, disease-modifying therapy; EDSS, Expanded Disability Status Scale; HET, high-efficacy therapy; n.a., not applicable; OCB, oligoclonal bands; WCC, white cell count.

case of conflicting results, manual recounts were given precedence. 34 participants who had external spinal cord imaging of sufficient quality were included in the analyses.

Brain MRI was performed on two different 3 Tesla scanners (Philips Achieva dStream, Philips Ingenia). Standardised brain MRI comprised a 3D spoiled gradient echo T1-weighted sequence and a turbo-spin echo T2-weighted FLAIR sequence. Brain lesions were segmented from FLAIR and T1-weighted images using lesion segmentation toolbox (LST)-AI, the deep learning-based extension of the LST.[25]

CSF parameters

CSF and serum concentrations for albumin as well as IgG, IgM and IgA were measured in parallel by standard nephelometric assays, and the respective CSF-to-serum quotients (Q_{IgG}, Q_{IgM}, Q_{IgA} and Q_{alb}) were calculated. Ig indices were defined as Q_{IgX}/Q_{alb}. Following the formulae proposed by Reiber,[26] intrathecal Ig synthesis was considered present in samples with Q_{IgX} greater than Q_{Lim}(IgX), the latter defined as:

$$Q_{Lim}\,(IgG) = 0.93 \times \sqrt{Q_{alb}^2 + 6 \times 10^{-6}} - 1.7 \times 10^{-3}$$

$$Q_{Lim}\,(IgM) = 0.67 \times \sqrt{Q_{alb}^2 + 120 \times 10^{-6}} - 7.1 \times 10^{-3}$$

$$Q_{Lim}\,(IgA) = 0.77 \times \sqrt{Q_{alb}^2 + 23 \times 10^{-6}} - 3.1 \times 10^{-3}$$

Detection of OCB was performed using isoelectric focusing followed by immunoblotting or immunofixation. OCB were considered positive if patterns 2 or 3, according to the 2005 consensus statement, were present.[27]

Model predictors

Predictor variables were chosen with three criteria in mind: (1) plausibility of prognostic value based on a survey of the literature (as mentioned in the introduction); (2) potential inclusion in a diagnostic workup of MS and (3) avoidance of problematic collinearity. The following were included: patient characteristics—sex, age at symptom onset; clinical parameters—EDSS score, monosymptomatic versus polysymptomatic manifestation of the demyelinating event leading to diagnosis; paraclinical measures—number of T2 lesions on brain (including infratentorial lesions) and spinal cord MRI, CSF WCC, intrathecal synthesis of IgG, IgM and IgA (defined dichotomously). The presence of OCB was not added to the models to avoid potential collinearity with Ig syntheses. Finally, two variables

Table 2 Best subset for each set of predictors of size n identified by all-subsets multivariable logistic regression with confirmed disability accumulation as an outcome

Set size n	Best subset of predictor variables	BIC
1	Number of spinal cord lesions	489.85
2	… + Intrathecal IgG synthesis	480.30
3	… + Age	477.88
4	**Number of spinal cord lesions (β*=0.47)+intrathecal IgG synthesis (β*=0.52)+age (β*=0.39)+polysymptomatic demyelinating event (β*=0.34)**	**475.79**
5	… + Sex	476.09
6	… + Number of brain lesions	481.44
7	… + Intrathecal IgM synthesis	487.16
8	… + Proportion of follow-up on HET+proportion of follow-up on other DMT–intrathecal IgM synthesis	492.52
9	… + Intrathecal IgM synthesis	498.20
10	… + Intrathecal IgA synthesis	503.75
11	… + CSFWCC	509.56
12	… + EDSS at baseline	515.45

The best subset for each n was determined by the lowest residual sum of squares (RSS). + indicates the addition of a predictor to the previous model while – denotes the removal of a previously included predictor. The respective values of the BIC are given, and the best overall model (lowest RSS and BIC) is emphasised in bold. Standardised (β*) coefficients are given for this model.
BIC, Bayesian information criterion; CSF, cerebrospinal fluid; DMT, disease-modifying therapy; EDSS, Expanded Disability Status Scale; HET, high-efficacy therapy; WCC, white cell count.

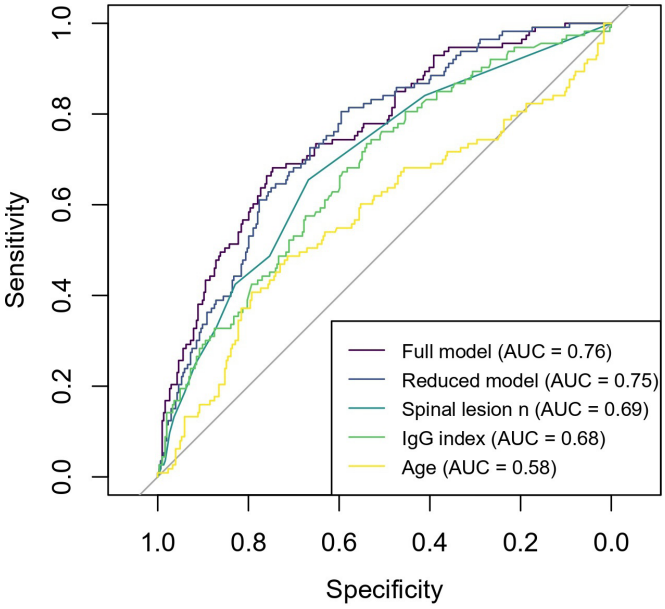

Figure 2 Receiver operating characteristic curves for different prediction models with confirmed disability accumulation as an outcome. Full model: including all 12 potential predictors; reduced model: including only significant predictors identified by all-subsets regression (number of spinal cord lesions, intrathecal IgG synthesis, age, polysymptomatic manifestation). A curve for IgG index as a surrogate for intrathecal IgG synthesis was added since the latter's dichotomous nature prevents it from being displayed this way. AUC, area under the curve; n, number.

concerning DMT were added. They were longitudinal in nature and would not initially be available to the diagnosing clinician, so their inclusion was made primarily with model adjustment in mind. The following were chosen for this purpose: proportion of follow-up on high-efficacy therapy (HET) (alemtuzumab, cladribine, fingolimod, natalizumab, ocrelizumab, ofatumumab, ozanimod, rituximab and siponimod) and proportion of follow-up on other DMT (dimethyl fumarate, glatiramer acetate, interferon and teriflunomide).

Statistical analyses

Since none of the variables of interest were normally distributed (as determined by the Shapiro-Wilk test), non-parametric tests were chosen. Values are given as median with IQR, accordingly. For group comparisons of dichotomous variables, Fisher's exact test was used. For group comparisons of continuous variables, two-sided, independent Wilcoxon-Mann-Whitney tests were performed. Spearman's rank correlation coefficient (r_s) was used for simple correlations. Individual variance inflation factors (VIFs) were calculated to assess potential collinearity. Predictors with a VIF<2 were considered unproblematic for model inclusion to ensure a conservative estimate.[28]

Our selection of N=12 variables of interest resulted in a theoretical total of 2^{12}=4096 possible statistical models. To mitigate the pitfalls of model selection bias inherent in the significance-driven inclusion or exclusion of parameters,[29] we applied all-subsets multivariable logistic regression. This method determines the best model following two steps: (1) For each set of predictors of size n (where n=1…, N), find the subset with the smallest residual sum of squares (RSS). (2) Identify the model with the best fit among these subsets according to some predefined criterion. Two common metrics for the latter are the Akaike information criterion and the Bayesian information criterion (BIC). The conceptual differences between the two are beyond the scope of this work and have been extensively discussed elsewhere.[30] We chose the BIC for our analyses as, generally speaking, it tends to favour smaller/simpler models by more severely penalising the addition of further predictors: $BIC = k \ln(n) - 2 \ln \hat{L}$, where k is the number of estimated parameters in the model, n is the sample

Table 3 Performance parameters for different prediction models with confirmed disability accumulation as an outcome

Predictor	AUC (95% CI)	Threshold	Accuracy	Sensitivity	Specificity	PPV	NPV
Full model	0.76 (0.71 to 0.81)	n.a.	0.73	0.68	0.75	0.50	0.86
Reduced model	0.75 (0.70 to 0.80)	n.a.	0.65	0.81	0.60	0.43	0.89
Spinal cord lesions	0.69 (0.64 to 0.75)	1	0.66	0.65	0.67	0.42	0.84
Intrathecal IgG synthesis	0.64 (0.58 to 0.69)	n.a.	0.60	0.71	0.56	0.37	0.84
Age	0.58 (0.51 to 0.64)	38	0.65	0.49	0.71	0.39	0.79
Polysymptomatic	0.60 (0.55 to 0.65)	n.a.	0.48	0.50	0.70	0.30	0.80

Full model: including all 12 potential predictors; reduced model: including only significant baseline predictors identified by all-subsets regression (number of spinal cord lesions, intrathecal IgG synthesis, age, polysymptomatic manifestation). Thresholds were calculated using Youden's method.[33]
AUC, area under the curve; n.a., not applicable; NPV, negative predictive value; PPV, positive predictive value.

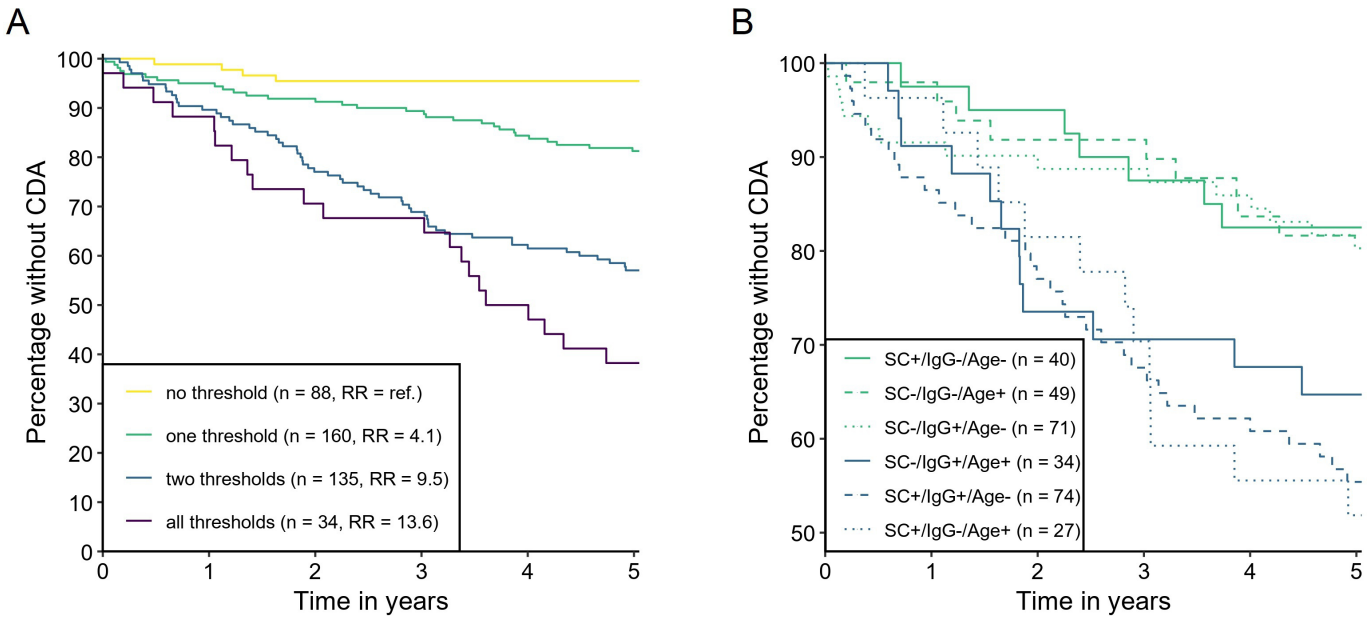

Figure 3 Survival curves based on optimal thresholds for the three main predictors of CDA. (A) Survival curves of four subgroups based on the number of threshold values exceeded. RRs were calculated in reference to the group without any predictor above threshold. (B) Survival curves of the one and two thresholds groups further stratified by the specific predictors exceeding their respective threshold. + and − denote whether individuals in a group exceed the respective threshold for a variable or not (or are positive for an attribute or not in the case of intrathecal IgG synthesis). The y-axis has been cropped for better discrimination of the curves. Threshold values were calculated using Youden's method[33]: SC lesion number >1, age >38 years. CDA, confirmed disability accumulation; ref., reference; RR, relative risk; SC, spinal cord.

size and $\hat{L}$ is the maximum value of the likelihood function for the model.[31] Smaller BIC values are indicative of better fits. The intercept was retained in all models, and standardised beta coefficients (β*) were calculated for the best model variables.

A number of sensitivity analyses were conducted: one in a subgroup of participants with at least one spinal cord lesion to rule out biases through a non-negligible amount of zero values; another one in a subgroup of participants with positive OCB to investigate a potential influence of this variable otherwise not tested for model inclusion. To gain a relative estimate of model performance, three further analyses were implemented: (continuous) Ig indices were chosen as variables instead of (dichotomous) intrathecal synthesis parameters. The number of spinal cord lesions was substituted with either cervical cord lesion number or thoracic cord lesion number, and the resulting models were compared. Additionally, each of the predictors identified for model inclusion was substituted with a conceptually related parameter (one at a time) and the resulting models' performances were evaluated. Any of the DMT variables identified as significant by regression analysis were excluded when assessing model performance (except for the full variable model) as they would not be available at the time of diagnosis. Statistical comparison of the area under the curve (AUC) for two receiver operating characteristic (ROC) curves was performed according to DeLong's method for paired data.[32] Optimal thresholds in ROC analysis were identified using Youden's index.[33]

All statistical and graphical analyses were done in R (V.4.4.0) and its packages lmSubsets, pROC, survminer and tidyverse. P values <0.05 were considered statistically significant.

RESULTS

Sample characteristics

A flow chart for data selection and outcomes is given in figure 1. 417 persons were included in the final analyses, 113 (27.1%) of whom experienced CDA during the 5-year follow-up (60 PIRA, 53 RAW). The median time from the beginning of clinical follow-up to disability accumulation in the CDA group was 1.9 years (IQR: 1.1–3.3). As summarised in table 1, people who experienced CDA tended to be older and female.

They had significantly more T2 lesions on both brain and spinal cord MRI. Additionally, their CSF WCC as well as IgG and IgM indices were higher, and intrathecal IgG synthesis and OCB were more frequent. More demyelinating events leading to the first diagnosis were polysymptomatic in the CDA group, and the latter had more annual EDSS visits on average. There were no differences regarding the time spent on DMT and the proportion of participants on HET between the groups. However, a higher proportion of follow-up spent on HET was significantly associated with higher EDSS scores at baseline (r_s=0.21) as well as more brain (r_s=0.31) and spinal cord lesions (r_s=0.25, p<0.001 in each case). Correlations among (para)clinical parameters did not raise concerns regarding collinearity (each VIF<2).

Predictors of disability accumulation

Among all possible variable permutations, a four-predictor model yielded the best model fit. This model contained intrathecal IgG synthesis, the number of spinal cord lesions, age and polysymptomatic manifestation (table 2).

Intrathecal IgG synthesis (β*=0.52) and spinal cord lesion number (β*=0.47) were the most statistically relevant variables (both p<0.001), followed by age (β*=0.39, p<0.01) and polysymptomatic manifestation (β*=0.34, p<0.01).

Model performance and individual risk of disability accumulation

Including all 12 initial variables in the prediction model resulted in an AUC of 0.76 (95% CI 0.71 to 0.81) with a positive

predictive value of 0.50, and a negative predictive value of 0.86 at the optimal threshold (table 3).

When reducing the model parameters to only the four classified for inclusion by all-subsets regression, the AUC was calculated at 0.75 (95% CI 0.70 to 0.80). At the optimal threshold, this reduced model yielded a positive predictive value of 0.43 (against a CDA prevalence of 27%) and a negative predictive value of 0.89. No significant difference in performance was found between the full 12-variable model and the reduced model (p=0.35). ROC curves for the two models and each individual continuous predictor are shown in figure 2. Optimal threshold values were then calculated for the continuous predictors (number of spinal cord lesions >1, age >38 years) and individual performance was assessed based on these thresholds (table 3). A comparison between the best single predictor model (number of spinal cord lesions) and the reduced model showed a significant difference in performance, favouring the latter (p<0.05).

Based on the optimal thresholds previously calculated, four subgroups were formed according to how many threshold values an individual exceeded. Only the three most statistically relevant predictors (intrathecal IgG synthesis, spinal cord lesion number, age) were included in this analysis to avoid an excess of subdivisions. Absolute risks of experiencing CDA on follow-up ranged from 61.8% for participants exceeding the thresholds of all three predictors (n=34) to 4.5% for individuals that were below all three threshold values (n=88) (figure 3A). Absolute risks for the groups with one (n=160) and two (n=135) predictor values above threshold lay between those extremes (18.8% and 43.0%, respectively). Taking the lowest risk group as reference, relative risks were 4.1, 9.5 and 13.6 for the one, two and three thresholds groups, respectively. Further stratification of the groups according to which specific predictors exceeded the threshold values yielded largely similar trajectories (figure 3B; online supplemental figure 1 shows survival curves for each of the four main predictors of CDA).

Sensitivity analyses

As a first sensitivity analysis, all-subsets regression was performed in a subgroup of 274 participants with at least one spinal cord lesion to investigate a potential effect of excess zero values on model selection. The four predictors from the whole cohort were confirmed, although their relative statistical weights were slightly different (intrathecal IgG synthesis, age, polysymptomatic manifestation: p<0.01; spinal cord lesion number: p<0.05; online supplemental table 1). For a second sensitivity analysis, we repeated our calculations in a subgroup of 371 participants with positive OCB. Again, the same four predictors were identified for model inclusion in this subgroup (spinal cord lesion number: p<0.001; intrathecal IgG synthesis, age, polysymptomatic manifestation: p<0.01; online supplemental table 2). Substituting the three Ig indices for the respective synthesis variables did not have any appreciable effect on the overall performance of either the full 12-variable model (p=0.87) or the reduced model (p=0.76; online supplemental table 3). An ROC curve for IgG index as predictor is shown in figure 2. Similarly, substituting either the number of cervical cord lesions or the number of thoracic cord lesions for the number of total spinal cord lesions did not change the composition nor the performance of the model (p=0.42 and p=0.11, respectively; online supplemental table 3). To better assess the relative contribution of each predictor to the reduced model, each of the four main predictors was substituted with a conceptually related parameter. This led to a significant worsening in performance when substituting the number of brain lesions for spinal cord lesions (p<0.05) or CSF WCC for IgG synthesis (p<0.05; online supplemental table 3).

Progression independent of relapse activity

Participants experiencing PIRA on follow-up tended to be older and have more spinal cord lesions than those with RAW (online supplemental table 4). They also spent proportionally more time on DMT and had longer follow-up intervals until the CDA event. Participants with RAW were significantly more likely to not take any DMT during the entire follow-up than participants with PIRA. A three-variable model including the number of spinal cord lesions (β*=0.54, p<0.05) as well as the proportion of follow-up spent on both HET (β*=0.77, p<0.01) and other DMT (β*=0.90, p<0.01) best differentiated PIRA from RAW on follow-up (online supplemental table 5).

The full 12-variable model yielded an AUC of 0.81 (95% CI 0.72 to 0.89) for differentiating between PIRA and RAW among CDA events (online supplemental table 6). This model significantly outperformed a model containing the only baseline predictor suggested for inclusion by regression analysis, the number of spinal cord lesions (AUC=0.64, 95% CI 0.54 to 0.74; difference between model performance: p<0.01). ROC curves for the models (including the one with both DMT parameters) are shown in online supplemental figure 2.

DISCUSSION

In this retrospective study including 417 people with RMS, we identified predictors of early disability accumulation from a set of routine parameters accessible at the time of first diagnosis. Specifically, intrathecal IgG synthesis, spinal cord lesions, age and polysymptomatic manifestation of the initial demyelinating event were associated with a sustained EDSS increase within a follow-up period of 5 years. When investigating the same baseline parameters in the context of differentiating between PIRA and RAW, only a higher number of spinal cord lesions remained as a significant predictor of PIRA.

The median baseline age of our cohort was 34 years, which matches the average age at first diagnosis reported in a large Italian study.[34] The rate of CDA we observed (27.1%) was rather low compared with similarly designed studies (32.1% in Monreal *et al*[35]; 33% in Rocca *et al*[8]; 37.1% in Tur *et al*[23]; 45.4% in Portaccio *et al*[22]). This seems plausible, given the short disease durations in our data (52 days on average) as well as the longer follow-up intervals in some of these studies. Interestingly, as previously noted by others,[22] PIRA constituted the majority of disability accumulation events, even in our newly diagnosed cohort.

The predictors of CDA identified in our analyses are consistent with previous findings regarding associations between disability accumulation and the number of spinal cord lesions,[6–8] age,[1 8 12 18 19 21 22] polysymptomatic manifestation[1 18–21] and intrathecal IgG synthesis.[17] Furthermore, they confirm observations previously made by our own group regarding the prognostic relevance of spinal cord lesions[9] and intrathecal IgG synthesis,[16] although there is a partial overlap between our present cohort and the ones used in those studies (37.9% in Lauerer *et al*[9]; 29.3% in Gasperi *et al*[16]). In accordance with other groups, we also found significantly higher IgM indices[15] and CSF WCC[10] among individuals experiencing disability accumulation, although neither variable emerged as a significant predictor in our multivariable model. Similar observations were made for the number of brain lesions. There was a notable association between the presence of (IgG)OCB and CDA in our data. As expected from other

reports,[12 14] however, the overall prevalence of OCB was quite high (89.0% versus an event rate of 27.1%), so a potential prognostic application would go along with a low positive predictive value. The lack of difference between the CDA and non-CDA groups in respect to DMT might seem surprising at first glance. However, the apparent discrepancy is most likely explained by the fact that higher baseline EDSS scores and lesion loads often prompted early initiation of HET in our cohort, thereby diluting any quantifiable DMT effects.

Identifying age as a significant predictor of CDA raises some practical concerns: If one goal of early disability prognosis is the timely initiation of DMT, how should age factor into this decision? On the one hand, the increased risk of CDA with age could be seen as an argument for HET. On the other hand, the well-established decrease in relapse-activity,[36] the waning efficacy of DMT,[37] and the relative predominance of PIRA in older persons with MS[22 38] might all be seen as evidence to the contrary. Further complicating the picture is the fact that the time of MS diagnosis does not necessarily coincide with the biological onset of the disease.

A higher number of spinal cord lesions was the only significant baseline predictor differentiating PIRA from RAW in our analysis. This association has been observed by others[39] as well as in our previous study.[9] The relatively limited prognostic performance of this one-variable model (AUC=0.64) stems from the fact that the nature of a CDA event is closely associated with DMT dynamics on follow-up. On the whole, RAW took place earlier during follow-up than PIRA (1.1 years versus 2.8 years on average) and consequently was more likely to happen shortly after or even before initiation of DMT. Participants experiencing PIRA, on the other hand, were on medication during most of their follow-up and only rarely were without any DMT (5.0%). This supports the conception of PIRA as 'silent progression' that continues to affect people whose inflammatory activity may be well controlled under DMT.

By restricting our group of predictors to measures available to diagnosing clinicians early on, we sought to ensure the translatability of our findings into everyday practice. This aim was further emphasised by choosing strict inclusion criteria for our well-characterised and reasonably large cohort. We consider this a major asset of our analysis. Another strength of our study was the choice of all-subsets regression as a method for model selection. Finding the best prognostic model among a number of potential predictors poses the problem of which combinations of variables should be statistically tested in the first place. Established methods like hierarchical, forced entry or stepwise regression all carry the risk of introducing biases that can distort the outcome.[29] By evaluating all potential variable permutations and ranking them according to some predefined criteria (RSS and BIC, in our case), all-subsets regression seeks to minimise these pitfalls.

Our study has some limitations. The inclusion of different diagnostic modalities (brain MRI, spinal cord MRI and lumbar puncture) entails a number of time intervals between diagnostic procedures that vary from individual to individual. Although we tried to limit this heterogeneity by defining strict inclusion criteria, some influence on our results cannot be ruled out. Similarly, the choice of a minimum follow-up of 5 years with varying numbers of follow-up visits during this time may introduce a form of selection bias. As in a previous study,[9] we confined our outcome measure to the first CDA event on follow-up and did not account for further events. Consequently, individuals could either experience PIRA or RAW, not both. This decision was made to avoid a loss of statistical power by further fragmentation into clinical subgroups. Finally, although our sensitivity analyses indicated a robust data foundation, external validation of our findings would be desirable, preferably in a prospective setting.

In conclusion, our observations suggest that intrathecal IgG synthesis, a higher number of spinal cord lesions, age and polysymptomatic manifestation at the time of diagnosis independently increase the risk of experiencing early disability accumulation in RMS. This risk increases with each additional predictor exceeding its respective threshold. Among the same group of baseline predictors, the number of spinal cord lesions was the only one to differentiate PIRA from RAW on follow-up.

Acknowledgements We are grateful for the invaluable contributions of the study participants as well as the assistance of the support staff at our clinic.

Contributors ML, TW and MM contributed to the conception and design of the study as well as data analysis. CB, CE, KG, SL, VP, AB, CG, JSK, CZ and BH participated in the acquisition of data. ML and MM drafted the text and figures. As the guarantor of this study, ML accepts full responsibility for the work, had access to the data and controlled the decision to publish.

Funding ML received funding through the 'Kommission für klinische Forschung' (KKF) of the School of Medicine and Health, Technical University of Munich (TUM). MM and JSK were supported by the Bavarian State Ministry for Science and Art (Collaborative Bilateral Research Program Bavaria – Quebec: AI in medicine, grant F.4-V0134.K5.1/86/34). MM was supported by research grant 428223038 of the German Research Foundation, DFG Priority Programme 2177, Radiomics: Next Generation of Biomedical Imaging.

Competing interests None of the authors have reported conflicts of interest directly related to this study. KG has received reimbursement for traveling expenses from UCB and Viatris. AB has received consulting and/or speaker fees from Alexion, Biogen, Celgene, Horizon, Novartis, Roche and Sandoz/Hexal. His institution has received compensation for clinical trials from Alexion, Biogen, Merck, Novartis, Roche and Sanofi Genzyme. JSK has received speaker fees for Novartis. He is a shareholder of Bonescreen. BH has served on scientific advisory boards for Novartis; he has served as DMSC member for AllergyCare, Sandoz, Polpharma, Biocon and TG therapeutics; his institution received research grants from Roche for multiple sclerosis research. He has received honoraria for counselling (Gerson Lehrmann Group). He holds part of two patents; one for the detection of antibodies against KIR4.1 in a subpopulation of patients with multiple sclerosis and one for genetic determinants of neutralising antibodies to interferon.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by the internal review board of the Technical University of Munich (reference number: 5848/13). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. The anonymised data that supported the findings of this study as well as the code underlying the analyses are available on reasonable request to any qualified investigator.

ORCID iDs
Markus Lauerer http://orcid.org/0000-0002-5711-0054
Jan S Kirschke http://orcid.org/0000-0002-7557-0003
Bernhard Hemmer http://orcid.org/0000-0001-5985-6784
Mark Mühlau http://orcid.org/0000-0002-9545-2709

REFERENCES

1 Weinshenker BG, Rice GP, Noseworthy JH, *et al*. The natural history of multiple sclerosis: a geographically based study. 3. Multivariate analysis of predictive factors and models of outcome. *Brain (Bacau)* 1991;114 (Pt 2):1045–56.
2 Reeve K, On BI, Havla J, *et al*. Prognostic models for predicting clinical disease progression, worsening and activity in people with multiple sclerosis. *Cochrane Database Syst Rev* 2023;9:CD013606.
3 Uher T, Vaneckova M, Sobisek L, *et al*. Combining clinical and magnetic resonance imaging markers enhances prediction of 12-year disability in multiple sclerosis. *Mult Scler* 2017;23:51–61.
4 Chung KK, Altmann D, Barkhof F, *et al*. A 30-Year Clinical and Magnetic Resonance Imaging Observational Study of Multiple Sclerosis and Clinically Isolated Syndromes. *Ann Neurol* 2020;87:63–74.
5 Pisani AI, Scalfari A, Crescenzo F, *et al*. A novel prognostic score to assess the risk of progression in relapsing-remitting multiple sclerosis patients. *Eur J Neurol* 2021;28:2503–12.
6 Arrambide G, Rovira A, Sastre-Garriga J, *et al*. Spinal cord lesions: A modest contributor to diagnosis in clinically isolated syndromes but a relevant prognostic factor. *Mult Scler* 2018;24:301–12.
7 Brownlee WJ, Altmann DR, Prados F, *et al*. Early imaging predictors of long-term outcomes in relapse-onset multiple sclerosis. *Brain (Bacau)* 2019;142:2276–87.
8 Rocca MA, Valsasina P, Meani A, *et al*. Spinal cord lesions and brain grey matter atrophy independently predict clinical worsening in definite multiple sclerosis: a 5-year, multicentre study. *J Neurol Neurosurg Psychiatry* 2023;94:10–8.
9 Lauerer M, McGinnis J, Bussas M, *et al*. Prognostic value of spinal cord lesion measures in early relapsing-remitting multiple sclerosis. *J Neurol Neurosurg Psychiatry* 2023;95:37–43.
10 Lotan I, Benninger F, Mendel R, *et al*. Does CSF pleocytosis have a predictive value for disease course in MS? *Neurol Neuroimmunol Neuroinflamm* 2019;6:e584.
11 Villar LM, Masjuan J, González-Porqué P, *et al*. Intrathecal IgM synthesis is a prognostic factor in multiple sclerosis. *Ann Neurol* 2003;53:222–6.
12 Mandrioli J, Sola P, Bedin R, *et al*. A multifactorial prognostic index in multiple sclerosis. Cerebrospinal fluid IgM oligoclonal bands and clinical features to predict the evolution of the disease. *J Neurol* 2008;255:1023–31.
13 Capuano R, Zubizarreta I, Alba-Arbalat S, *et al*. Oligoclonal IgM bands in the cerebrospinal fluid of patients with relapsing MS to inform long-term MS disability. *Mult Scler* 2021;27:1706–16.
14 Monreal E, Sainz de la Maza S, Costa-Frossard L, *et al*. Predicting Aggressive Multiple Sclerosis With Intrathecal IgM Synthesis Among Patients With a Clinically Isolated Syndrome. *Neurol Neuroimmunol Neuroinflamm* 2021;8:e1047.
15 Perini P, Ranzato F, Calabrese M, *et al*. Intrathecal IgM production at clinical onset correlates with a more severe disease course in multiple sclerosis. *J Neurol Neurosurg Psychiatry* 2006;77:953–5.
16 Gasperi C, Salmen A, Antony G, *et al*. Association of Intrathecal Immunoglobulin G Synthesis With Disability Worsening in Multiple Sclerosis. *JAMA Neurol* 2019;76:841–9.
17 Akaishi T, Takahashi T, Fujihara K, *et al*. Impact of intrathecal IgG synthesis on neurological disability in patients with multiple sclerosis. *Mult Scler Relat Disord* 2020;45:102382.
18 Runmarker B, Andersen O. Prognostic factors in a multiple sclerosis incidence cohort with twenty-five years of follow-up. *Brain (Bacau)* 1993;116 (Pt 1):117–34.
19 Confavreux C, Vukusic S, Adeleine P. Early clinical predictors and progression of irreversible disability in multiple sclerosis: an amnesic process. *Brain (Bacau)* 2003;126:770–82.
20 Bergamaschi R, Quaglini S, Trojano M, *et 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;78:757–9.
21 Malpas CB, Manouchehrinia A, Sharmin S, *et al*. Early clinical markers of aggressive multiple sclerosis. *Brain (Bacau)* 2020;143:1400–13.
22 Portaccio E, Bellinvia A, Fonderico M, *et al*. Progression is independent of relapse activity in early multiple sclerosis: a real-life cohort study. *Brain (Bacau)* 2022;145:2796–805.
23 Tur C, Carbonell-Mirabent P, Cobo-Calvo Á, *et al*. Association of Early Progression Independent of Relapse Activity With Long-term Disability After a First Demyelinating Event in Multiple Sclerosis. *JAMA Neurol* 2023;80:151–60.
24 Müller J, Cagol A, Lorscheider J, *et al*. Harmonizing Definitions for Progression Independent of Relapse Activity in Multiple Sclerosis: A Systematic Review. *JAMA Neurol* 2023;80:1232–45.
25 Wiltgen T, McGinnis J, Schlaeger S, *et al*. LST-AI: A deep learning ensemble for accurate MS lesion segmentation. *Neuroimage Clin* 2024;42:103611.
26 Reiber H. Cerebrospinal fluid--physiology, analysis and interpretation of protein patterns for diagnosis of neurological diseases. *Mult Scler* 1998;4:99–107.
27 Freedman MS, Thompson EJ, Deisenhammer F, *et al*. Recommended standard of cerebrospinal fluid analysis in the diagnosis of multiple sclerosis: a consensus statement. *Arch Neurol* 2005;62:865–70.
28 O'brien RM. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Qual Quant* 2007;41:673–90.
29 Freedman DA Professor, Freedman DA Professor. A Note on Screening Regression Equations. *Am Stat* 1983;37:152–5.
30 Vrieze SI. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychol Methods* 2012;17:228–43.
31 Schwarz G. Estimating the Dimension of a Model. *Ann Statist* 1978;6:4.
32 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–45.
33 YOUDEN WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32–5.
34 Prosperini L, Lucchini M, Ruggieri S, *et al*. Shift of multiple sclerosis onset towards older age. *J Neurol Neurosurg Psychiatry* 2022;93:1137–9.
35 Monreal E, Fernández-Velasco JI, Álvarez-Lafuente R, *et al*. Serum biomarkers at disease onset for personalized therapy in multiple sclerosis. *Brain (Bacau)* 2024;147:4084–93.
36 Schwehr NA, Kuntz KM, Butler M, *et al*. Age-related decreases in relapses among adults with relapsing-onset multiple sclerosis. *Mult Scler* 2020;26:1510–8.
37 Weideman AM, Tapia-Maltos MA, Johnson K, *et al*. Meta-analysis of the Age-Dependent Efficacy of Multiple Sclerosis Treatments. *Front Neurol* 2017;8:577.
38 Cagol A, Schaedelin S, Barakovic M, *et al*. Association of Brain Atrophy With Disease Progression Independent of Relapse Activity in Patients With Relapsing Multiple Sclerosis. *JAMA Neurol* 2022;79:e221025:682–92:.
39 Prosperini L, Ruggieri S, Haggiag S, *et al*. Prognostic Accuracy of NEDA-3 in Long-term Outcomes of Multiple Sclerosis. *Neurol Neuroimmunol Neuroinflamm* 2021;8:e1059.