Received: 3 March 2021 | Accepted: 3 April 2021

DOI: 10.1111/ene.14859

european journal of neurology
the official journal of the european academy of neurology

ORIGINAL ARTICLE

A novel prognostic score to assess the risk of progression in relapsing–remitting multiple sclerosis patients

Anna Isabella Pisani[1] | **Antonio Scalfari**[2] | **Francesco Crescenzo**[1] | **Chiara Romualdi**[3] | **Massimiliano Calabrese**[1]

[1]Department of Neurological and Movement Sciences, University of Verona, Verona, Italy

[2]Brain Division, Imperial College London, London, UK

[3]Department of Biology, University of Padova, Padova, Italy

Correspondence
Massimiliano Calabrese, Department of Neurological and Movement Sciences, University of Verona, Verona, Italy.
Email: calabresem@hotmail.it

Abstract

Background: At the patient level, the prognostic value of several features that are known to be associated with an increased risk of converting from relapsing–remitting (RR) to secondary phase (SP) multiple sclerosis (MS) remains limited.

Methods: Among 262 RRMS patients followed up for 10 years, we assessed the probability of developing the SP course based on clinical and conventional and non-conventional magnetic resonance imaging (MRI) parameters at diagnosis and after 2 years. We used a machine learning method, the random survival forests, to identify, according to their minimal depth (MD), the most predictive factors associated with the risk of SP conversion, which were then combined to compute the secondary progressive risk score (SP-RiSc).

Results: During the observation period, 69 (26%) patients converted to SPMS. The number of cortical lesions (MD = 2.47) and age (MD = 3.30) at diagnosis, the global cortical thinning (MD = 1.65), the cerebellar cortical volume loss (MD = 2.15) and the cortical lesion load increase (MD = 3.15) over the first 2 years exerted the greatest predictive effect. Three patients' risk groups were identified; in the high-risk group, 85% (46/55) of patients entered the SP phase in 7 median years. The SP-RiSc optimal cut-off estimated was 17.7 showing specificity and sensitivity of 87% and 92%, respectively, and overall accuracy of 88%.

Conclusions: The SP-RiSc yielded a high performance in identifying MS patients with high probability to develop SPMS, which can help improve management strategies. These findings are the premise of further larger prospective studies to assess its use in clinical settings.

KEYWORDS
multiple sclerosis, demyelinating diseases, neurological disorders, risk factors

INTRODUCTION

Multiple sclerosis (MS) is an immune-mediated inflammatory and neurodegenerative disease affecting the central nervous system leading to chronic disability [1] Most cases initially present with a relapsing–remitting course (RR), characterized by acute relapses, which is almost inevitably followed by the secondary progressive phase (SP), leading to severe disability accumulation.

The conversion to SPMS is the key determinant of the long-term prognosis [2] but its prevention unfortunately remains an unmet

therapeutic need [3] At the individual level, the MS clinical course is extremely unpredictable. The probability of becoming progressive proportionally increases with the disease duration and with patients' age [4] mostly secondary to the failure of brain compensatory mechanisms [5] Several demographics, clinical, neuroradiological and environmental factors have been associated with a higher risk of converting to SPMS. Older age at disease onset [4,6,7] and a florid early clinical [2,8–10] and magnetic resonance imaging (MRI) multifocal white matter (WM) inflammatory activity [11,12] distinguish patients more likely to develop severe disability in the future. However, the severity of the grey matter (GM) damage (cortical lesions and atrophy) explains the disease progression better than the WM damage and it is a good predictor of the long-term outcome [13–15] since the early stage of the disease. Despite extensive efforts collecting and validating prognostic information, the accurate identification early in the disease course of patients destined to have poor outcome remains challenging, which hampers our ability to optimize the therapeutic approach before irreversible damage takes place.

By exploiting the flexibility of machine learning approaches, we developed the secondary progressive risk score (SP-RiSc), which integrates demographic, clinical and MRI data collected from a cohort of RRMS patients during the first 2 years after the disease diagnosis. The SP-RiSc may be a tool potentially applicable in specific clinical settings to objectively estimate the individual risk of transition from RRMS to the progressive phase.

METHODS

MS patient cohort and study design

The study was designed by retrospectively collecting demographic, clinical and MRI data (including measures of both focal and diffuse GM damage) at diagnosis and after the first 2 years of disease course, from 262 patients diagnosed with RRMS [16] who were recruited at the MS Specialist Centre of the University Hospital of Verona (Verona, Italy) between 2005 and 2010 and were followed up to 2018, for a mean of 9.55 (range 6.8–13.13) years. Each patient was treated initially with one of the licensed first-line disease-modifying therapies (DMTs) (interferon-β1a and glatiramer acetate) and clinically examined by MS neurologists every 6 months or when a relapse occurred. During the follow-up period, patients experiencing disease activity were switched to second-line therapies (fingolimod and natalizumab). Physical disability was evaluated with the Expanded Disability Status Scale (EDSS) [17] A relapse was defined as the acute or subacute development of new or recurrent symptoms, lasting >24 h and not preceded by fever.[18]

The progressive course was defined by the occurrence of continuous disability accumulation independently of relapses and was confirmed 12 months later. Although transitory plateaus in the progressive course were allowed, steady progression was the rule [18] All procedures in this study were performed in accordance with the ethical standards of the institutional research committee and the 2013 Helsinki Declaration.[19]

MRI acquisition protocol and analysis

MRI sequences were acquired by Philips Achieva 1.5T MRI scanner (Philips Medical Systems), with 33 mT/m power gradient and a 16-channel head coil. During the study period, the scanner underwent a specific functioning test every 2 months to guarantee parameter stability. 3D magnetization-prepared rapid gradient-echo (MP-RAGE) [TR/TE = 25/4.6 ms], 3D fluid attenuated inversion recovery (FLAIR) [TR/TE = 10,000/120 ms] and 3D double inversion recovery (DIR) [TR/TE = 6500/265 ms, TI1/TI2 = 500 ms/2800 ms] sequences have been acquired to brain scan. Spinal cord sequences included dual-echo proton density and T2-weighted fast spin-echo, and short-tau inversion recovery. Patients were carefully positioned according to published guidelines for serial MRI studies in MS patients [20] T0 and T2 refer to RRMS diagnosis time and 2 years later, respectively (see Methods S1 for details).

Statistical analysis

The Akaike Information Criterion for Cox regression (AIC-Cox) with backward phasing out and the Cox proportional hazard (PHM) model were performed to compare these “traditional” approaches with a “more innovative” machine learning-based one (random survival forest [RSF]). Moreover, before performing the Cox regression models the variables were standardized (variables will be shifted to be zero centred and scaled to have unique variance).

Random survival forest model

The design of the study is summarized in Figure 1. The whole cohort was randomly split into a training set (80%, n = 219), which was used to model both the RSF and the SP-RiSc, and a testing set (20%, n = 43), on which the score was validated (Figure 1a). The populations in both sets shared similar features and similar proportions of patients, who converted to SPMS during the observation period (Table S1). The RSF is a machine learning-ensemble method for the analysis of right-censored survival data (Figure 1b), which is based on the extension of Breiman's random forest [21] and provides flexibility in highly correlated complex data (Methods S2). The demographic, clinical and MRI data collected during the first 2 years of the disease have been included in the RSF model in order (i) to identify those variables with a strong impact on the risk of developing SPMS and (ii) to study the synergic cooperation of the selected risk factors. A total of 12 variables have been analyzed, including demographic (gender and age at T0) and clinical (EDSS score at T0 and T2; the number of relapses over the first 2 years of disease) information; MRI data (the changes over the first 2 years of both global cortical thickness (CTh) and cerebellar cortical volume (CCV), cortical lesions (CLs) and WM lesions number at T0 and number of new lesions at T2; the presence of spinal cord lesion at T0). The model's results were adjusted for the treatment categorical variable,

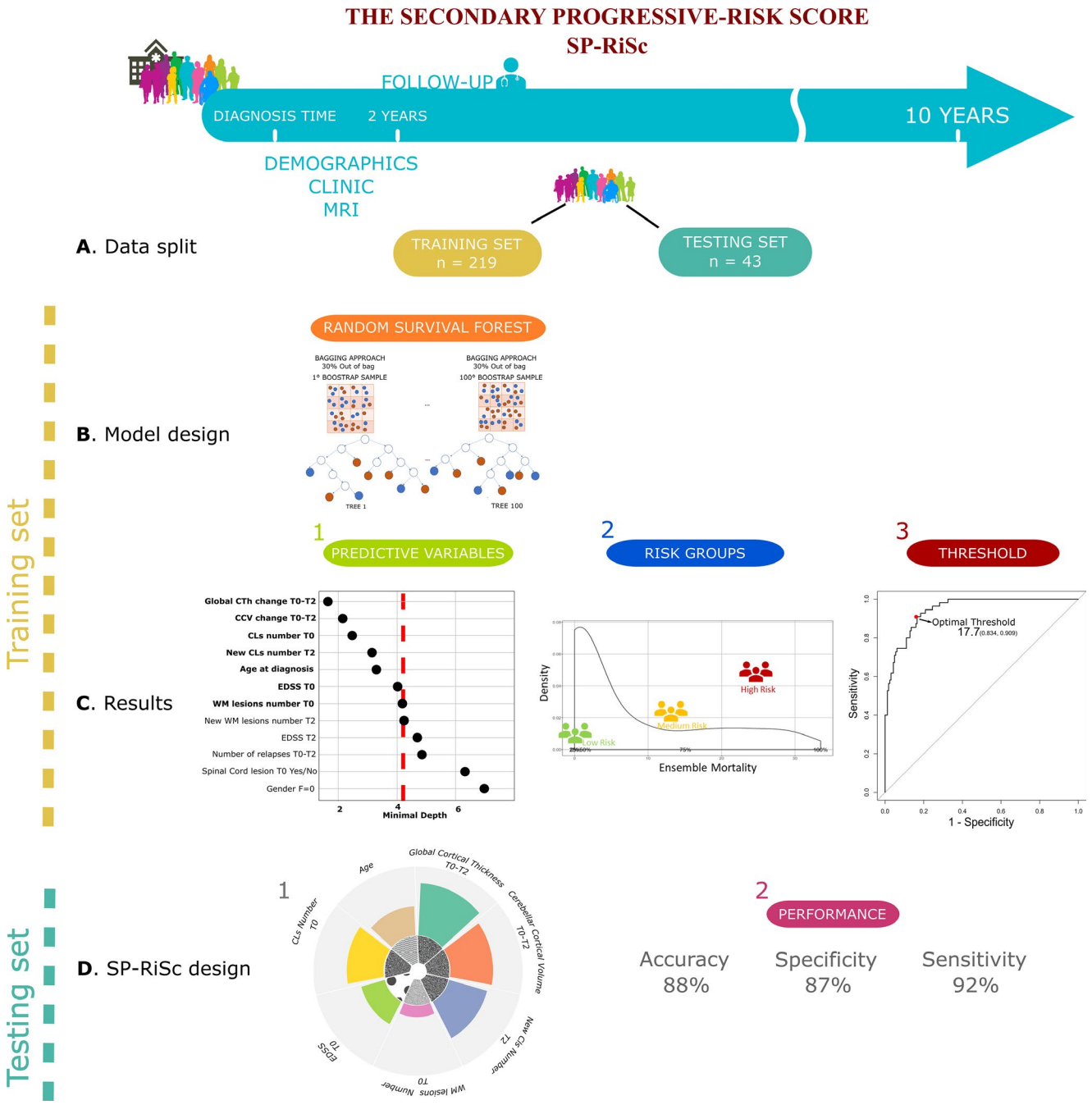

FIGURE 1 The overall study design. (a) *Data split*: The entire cohort was randomly split into training and testing set. (b) *Model design*: Random survival forest (RSF) modelling was performed on the training set. (c) *Results*: (1) The seven most predictive variables were selected, based on their minimal depth. (2) Risk groups were identified by ensemble mortality. (3) Receiver operating characteristic (ROC) analysis was used to identify the best score cut-off. (d) *Secondary progressive risk score (SP-RiSc) design*: (1) The Sp-RiSc tool was developed. (2) The Sp-RiSc performance specificity, sensitivity and overall accuracy were assessed on the testing set. [Colour figure can be viewed at wileyonlinelibrary.com]

indicating whether patients had been switched to second-line DMTs during the observation period.

The minimal depth (MD) is a dimensionless measure used to quantify the predictive effect of each variable included in the model. The mean of the MD model distribution (MD mean distribution = 4.21) was used as a reference threshold (Thr) to determine the size of the predictive effect of the variables included in the model [22] (Methods S2); variables with MD lower than Thr are those with the highest predictive power (Figure 1c1). The ensemble mortality (EM), which is the individual predicted outcome of RSF, was then calculated to identify three different SP risk groups (Methods S2). Finally, the EM and the MD parameters were combined to develop the SP-RiSc tool (Figure 1d1). Both the Brier score (BS) and the Harrell's Concordance Index (C-Index) were used to assess the prediction accuracy of the RSF.[23] The BS expresses the mean squared difference between the actual status and the predicted survival probability, while the C-Index

is a time-independent measure indicating how well the model discriminates between patients with and without the outcome. Higher C-index and lower BS indicate better prediction performance (Methods S2).

SP-RiSc implementation and evaluation

Based on EM measures, patients were stratified into three risk groups, using quartile distribution values: patients with EM lower than the 1st quartile and greater than the 3rd quartile are respectively at low and high risk of converting to SPMS, while the remaining patients belong to the medium-risk group (Figure 1c2).

The EM and the MD parameters were combined in order to design the SP-RiSc tool, which included only variables exerting the highest predictive effect (MD lower than 4.21) (Figure 2a). The score design procedure handles differently discrete (age, EDSS, CLs and WM lesion) and continuous (global CTh and CCV) variables. Continuous variables had to be transformed into discrete counterparts. The discretization was performed using the lower and the upper limits of the bootstrap confidence interval for each variable within each risk group. If we consider all the possible values for a variable, the three non-overlapping intervals generate seven different classes (considering values lower than the lower limit of low-risk group, higher than the higher limit of high-risk group, and values in the middle of the disjoint intervals). A numerical weight value ranging from 0.5 to 3.5 (in steps of 0.5) was assigned to each class: a higher numerical value indicates a higher magnitude of the variable considered. Then, the value of each category was divided by the variable MD. For the discrete variables, the value was divided by its own MD value (Figure 2b). The final SP-RiSc resulted from the sum of the risk factors weighted by the predictive ability. The formal definition is reported in Methods S2.

The receiver operating characteristic (ROC) analysis (Youden index method) was used to identify on the training set the SP-RiSc cut-off that maximizes specificity and sensitivity of identifying patients at risk of entering the SP phase (Figure 1c3). Finally, this threshold was tested on the validation set (Figure 1d2).

RSF model and score including only WM parameters

The RSF model was also performed on the training set using demographic, clinical and WM parameters collected during the two first years, therefore without considering information on GM damage. The most predictive variables were combined to develop

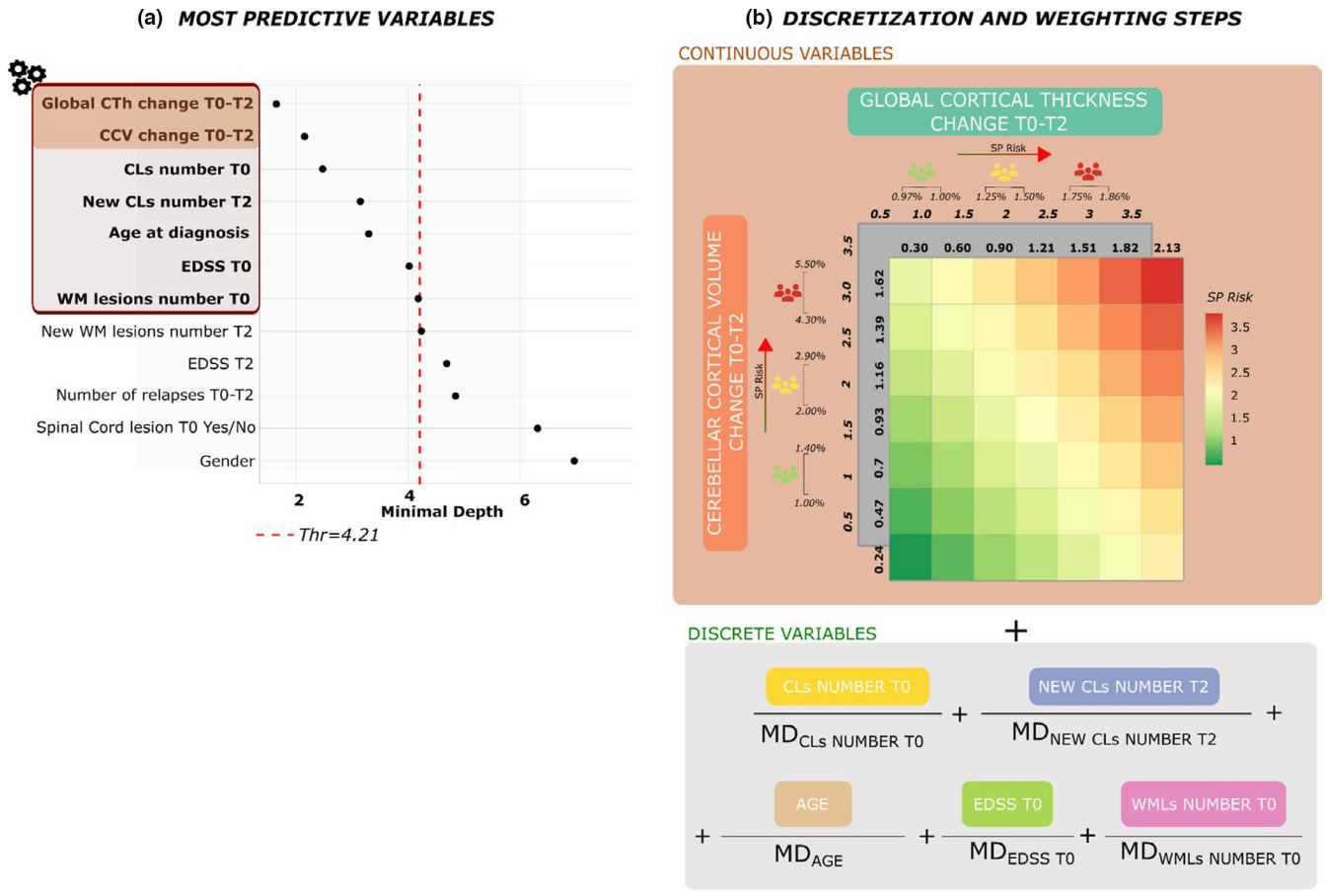

FIGURE 2 Secondary progressive risk score (SP-RiSc) design. (a) *Variables* are listed based on their minimal depth (MD) values; lower values indicate higher predictive accuracy. Predictive variables with MD lower than the estimated threshold (Thr) (4.21) are highlighted. (b) *Discretization steps:* Continuous and discrete variables were discretized and weighed for MD measure to be combined to build the SP-RiSc tool. [Colour figure can be viewed at wileyonlinelibrary.com]

a different version of the SP-RiSc which was validated on the testing set. The comparison of the ROC curves of the SP-RiSc and the alternative score with only WM parameters is reported in Appendix S1.

RESULTS

Demographic, clinical and MRI data for the 262 RRMS patients, at diagnosis and after 2 years, are shown in Table 1. During the 10 years follow-up period, 69 (26%) patients converted to SPMS; this subgroup, in comparison to those who remained in the RR phase, was distinguished at diagnosis by older age ($p < 0.001$) and higher number of CLs and WM lesions ($p < 0.001$) and global CTh at T0 ($p < 0.01$), and during the first 2 years by a larger number of relapses ($p < 0.001$), by a more significant accumulation of GM (new CLs, and CCV change; $p < 0.001$), and WM (new T2 lesions; $p < 0.01$) damage.

In the whole group, during the follow-up period 114 (43.5%) patients were switched to second-line DMTs (fingolimod and natalizumab), based on the occurrence of clinical and radiological disease breakthrough.

RSF model results

By applying the RSF, we have identified seven variables with MD lower than the estimated Thr (4.21) and therefore highly predictive of the risk of converting to SPMS. The strongest predictive effect was exerted by the cortical thinning (MD = 1.65) and by the cerebellar cortical volume loss (MD = 2.15) during the two first 2 years of disease, and by the CLs load at diagnosis (MD = 2.47) and its increase (MD = 3.15) after 2 years; age (MD = 3.30), EDSS (MD = 4.10) and the WM lesions total number at RRMS diagnosis (MD = 4.17) had a moderate impact on the probability of converting to SP (Figure 2a). The number of WM lesions at T2 (MD = 4.23), the EDSS at T2 (MD = 4.66), the number of relapses during the two first years of disease (MD = 4.84), the presence of the spinal cord at T0 (MD = 6.25) and gender (MD = 6.87) were found to have an MD higher than the Thr (=4.21) and were consequently excluded from the SP-RiSc tool design.

Both the BS and C-Index measures confirmed the goodness of fit and the predictive accuracy of the statistical model at different follow-up time quartiles: the BS parameter decreased over time reaching a value close to 0 and the C-Index increased close

TABLE 1 Demographical, clinical and radiological features collected during the two first years of disease

Feature	Whole group of RRMS at T0 *n* = 262	RRMS at the end of follow-up *n* = 193	SPMS at the end of follow-up *n* = 69	*p* value
Age at diagnosis, mean (SD), y	33.5 (9.5)	31.6 (9.6)	38.9 (6.6)	<0.001
Duration of follow-up, median (range)	9.55 (6.79–13.13)	9.26 (6.79–12.05)	10.16 (7.37–13.12)	<0.001
Gender, female, *n* (%)	153 (58.4)	73 (37.8)	33 (47.82)	0.05
Number of relapses T0–T2, median (range)	1 (1–3)	1 (1–3)	2 (1–3)	<0.001
EDSS T0, median (range)	1.5 (0–3.5)	1.5 (0–3)	1.5 (0–3.5)	<0.01
EDSS T2, median (range)	2 (0–2.5)	2 (0–2)	2 (0–2.5)	<0.05
CLs number T0, median (range)	2 (0–19)	1 (0–14)	6 (0–19)	<0.001
New CLs T2, median (range)	0 (0–8)	0 (0–3)	2 (0–8)	<0.001
WM lesion number T0, median (range)	8 (2–22)	7 (2–22)	10 (2–22)	<0.001
New WM lesion number T2, median (range)	1 (0–4)	1 (0–4)	1 (0–4)	<0.01
CCV T0, mean (SD), cm^3	106.3 (2.2)	103.1 (2.0)	97.9 (2.4)	<0.001
CCV change T0–T2, mean (SD), %	2.5 (1.99)	1.8 (1.3)	4.5 (2.07)	<0.001
Global CTh T0, mean (SD), mm	2.46 (0.21)	2.39 (0.14)	2.27 (0.25)	<0.01
Global CTh change T0–T2, mean (SD), %	1.3 (0.5)	1.16 (0.4)	1.8 (0.5)	<0.001
Spinal cord lesion T0, yes, *n* (%)	58 (22)	34 (17.60)	24 (34.78)	<0.01

Data were reported using mean and SD or median and range, based on their distribution. Dichotomous and categorical variables were described using proportions. Differences between groups were assessed using Mann–Whitney. Fisher's exact test was applied to test the contingency tables on categorical variables.

Statistical significance was considered at $p < 0.05$.

Abbreviations: CCV, cerebellar cortical volume; CL, cortical lesion; CTh, cortical thickness; EDSS, Expanded Disability Status Scale; RRMS, relapsing–remitting multiple sclerosis; SD, standard deviation; SPMS, secondary progressive multiple sclerosis; WM, white matter.

to 100%. At 7 years: BS = 7.7%, confidence interval (CI) 6.4–10 and C-Index = 92.0%; at 8.5 years BS = 8.9% CI 7.6–10.8 and C-Index = 91.0%; at 9.5 years BS = 8% CI 5.9–11 and C-Index = 91.4%; at 10.5 years BS = 4.6% CI 2.9–5.6 and C-Index = 90%; at 13.5 years BS = 2.0% CI 0.61–3.4 and C-Index = 90.0%. Therefore, the model resulted in being suitable both in terms of predictive accuracy and SP event discrimination, and its predictive accuracy increased proportionally with the disease duration. In contrast, when we applied the same predictive accuracy measures to the Cox PHM and AIC-Cox we found a poor performance in predicting the occurrence of SP, as the BS scores of both Cox models reached very high values close to 60% at all different follow-up time quartiles. In addition, in the Cox regressions models the C-index measures remained close to 80% and therefore lower than the RSF model (Table S4, S5 and S6).

SP-RiSc design and evaluation

According to the EM distribution, three groups of patients characterized by high (n = 55), medium (n = 54) and low (n = 110) risk to enter the SP phase were identified. Patients in the high-risk group had an EM higher than 16.35 (3rd quartile), while the low-risk group patients had an EM lower than 0.32 (1st quartile). In the high-risk group 46 (85.5%) patients entered the SP phase during the follow-up period (median [IQR] time to reach the SP phase = 7 [2.35] years), while in the medium group only 9 (17.7%) became SP (median [IQR] time to reach the SP phase = 8.5 [1.9] years). All the patients in the low-risk group remained in the RRMS phase during the entire study period. Clinical and MRI features of each group are reported in Table 2.

By combining the seven most predictive variables (Figure 2a) we developed the SP-RiSc tool, which takes into account the different size of the predictive effect exerted by each variable (Figure 3a). In the training set, the optimal SP-RiSc cut-off value, estimated by the ROC curve analysis, was 17.7 and had a sensitivity of 0.91 (95% CI 0.82–1) and specificity of 0.83 (95% CI 0.73–0.95) (Figure 3b). Therefore, patients with a SP- RiSc≥ of 17.7 have a 91% probability of converting to SPMS within 10 years from the disease diagnosis. In contrast, patients with SP-RiSc <17.7 had an 83% probability of remaining in the relapsing–remitting phase. The SP-RiSc predicted the individual risk of SPMS with an overall accuracy of 85% (95% CI 80%–92%); in the training set, the score was able to discriminate 50/56 SPMS patients (true positive, TP) and 136 of 163 patients as RRMS patients (true negative, TN), while 5 SPMS patients were incorrectly classified as RR (false negative, FN) and 27 RRMS patients were misclassified as SPMS (false positive, FP). Consequently, the positive predictive value (PPV), which indicates the probability of patients being classified as truly SP, was 65% (95% CI 55%–87%). On the contrary, the negative predicted value (NPV), which is the probability of patients being classified as truly RRMS, was 97% (95% CI 93%–100%). Finally, to evaluate the generalization property of the SP-RiSc, the threshold of 17.7 was applied on the testing set: the cut-off discriminated 12 of 13 as patients with SP condition (TP) and 26 of 30 as RR status (TN). The PPV was 75% (95% CI 48%–93%) and the NPV was 96% (95% CI 81%–100%). Therefore, results from the independent test analyses confirmed the great accuracy (88%, 95% CI 75%–96%), high sensitivity (92%, 95% CI 70%–100%) and specificity (87%, 95% CI 70%–96%) of the SP-RiSc performance in identifying patients at higher risk of conversion to SPMS.

Evaluation of WM parameters effect on the SP conversion.

We evaluated a different version of the SP-RiSc, without including GM damage variables, using only the age and the EDSS score at

TABLE 2 Descriptive statistics of the top seven predictive variables selected by random survival forest for each of the three risk groups

Significant predictive variable	High-risk group N = 55 46 (85.5%) SPMS patients	Medium-risk group N = 54 46 (17.7%) SPMS patients	Low-risk group N = 110 0 (0%) SPMS patients	p value
Global CTh change T0–T2, mean (SD), %	1.8 (0.3)	1.3 (0.4)	0.99 (0.1)	<0.001
CCV change T0–T2, mean (SD), %	4.9 (2.2)	2.5 (1.6)	1.3 (0.5)	<0.001
CLs number T0, median (range)	6 (1–18)	3 (1–8)	1 (0–5)	<0.001
New CLs T2, median (range)	2 (1–8)	0 (0–4)	0 (0–8)	<0.001
Age at diagnosis, mean (SD), y	37.7 (6.9)	33.5(11.1)	31. 3(10.2)	<0.001
EDSS T0, median (range)	1.5 (0–3)	1.5 (0–3.5)	1.5 (0–3.5)	<0.001
median (range)	10 (2–22)	7 (2–16)	7 (2–21)	<0.001

Note: Data were reported using mean and SD or median and range, based on their distribution. Dichotomous and categorical variables were described using proportions. Differences between groups were assessed using analysis of variance (ANOVA). Statistical significance was considered at $p < 0.05$.

Abbreviations: CCV, cerebellar cortical volume; CL, cortical lesion; CTh, cortical thickness; EDSS, Expanded Disability Status Scale; SD, standard deviation; SPMS, secondary progressive multiple sclerosis.

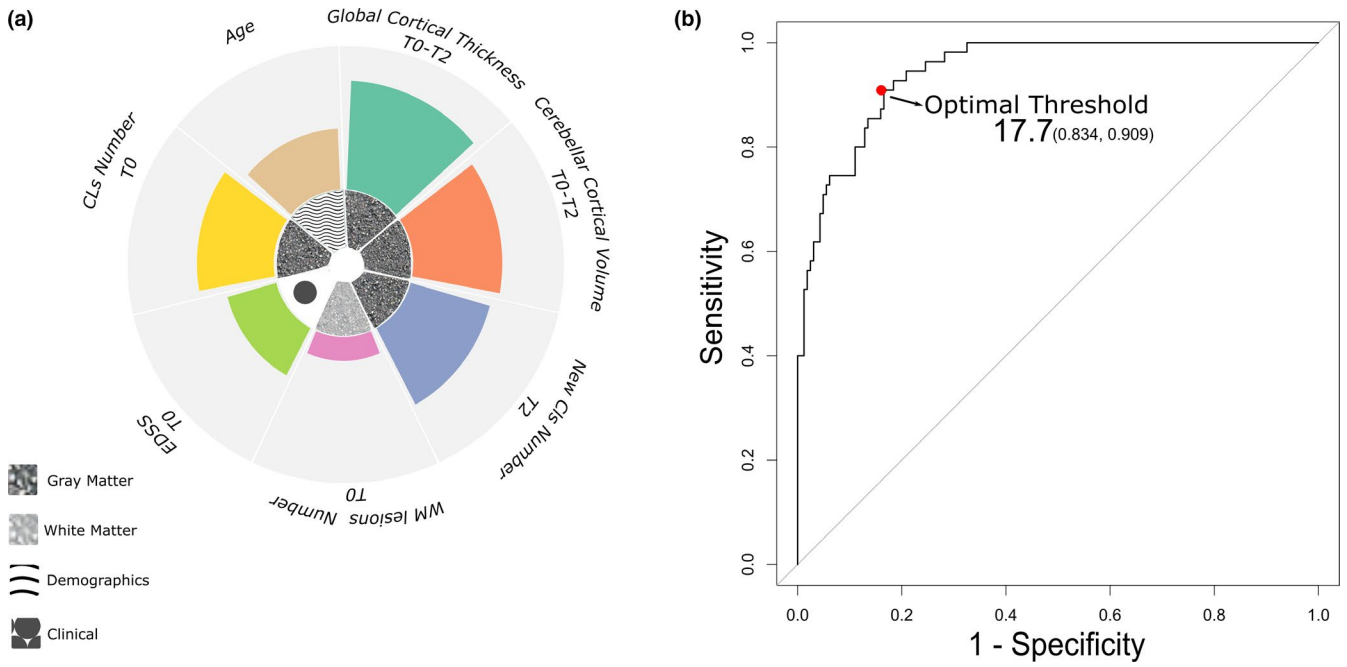

FIGURE 3 Secondary progressive risk score (SP-RiSc) visualization and receiver operating characteristic (ROC) analysis. (a) *SP-RiSc tool visualization*: The seven selected predictors are shown with different colours; the size of the predictive power on SP conversion is reflected by the size of the corresponding shape. Different white and black patterns for each significant variable were reported in the deepest circle. (b) *ROC curve analysis:* Detection of the optimal SP-RiSc cut-off on the training set. [Colour figure can be viewed at wileyonlinelibrary.com]

diagnosis and the number of T2 lesions and relapses during the first 2 years (Table S2); this was validated on the testing set and showed an accuracy of 74% (CI 95% 59%–87%), a specificity of 67% (CI 95% 47%–83%) and a sensitivity of 92% (CI 95% 64%–100%). The PPV and the NPV were 55% (CI 95% 32%–76%) and 75% (CI 95% 48%–93%), respectively. Therefore, the exclusion of GM damage parameters resulted in developing both the model and the score with lower performance, compared to the standard SP-RiSc, as shown by the C-Index and BS, and by the ROC curve analyses comparison (Table S3 and Figure S2).

DISCUSSION

Despite extensive efforts collecting information on the natural history of MS and identifying features associated with poor prognosis, the long-term outcome of the disease remains extremely unpredictable, especially at the individual level. The conversion to the SP phase is considered the key adverse event, leading to the accumulation of severe disability, but its prevention is still an unmet therapeutic target. The early and accurate identification of patients destined to experience a severe disease course is paramount to optimizing their management by implementing an aggressive therapeutic approach in a timely manner, before irreversible damage takes place. In this study, we propose the SP-RiSc as a reliable tool to estimate the risk of SP conversion, based on demographic, clinical and MRI measures (conventional and non-conventional) collected during the first 2 years of the disease.

We retrospectively assessed a cohort of 262 RRMS patients followed up for a mean of 10 years. As we previously reported,[4] patients who entered the SP phase were distinguished by older age at diagnosis, by a more significant accumulation of focal inflammatory WM and GM damage during the early phase of the disease, and by a larger number of early relapses. This is in line with previous predictive models, but unfortunately this information has limited prognostic use when applied to a single individual. We addressed and overcame these limitations by using the Random Survival Forest, which is a machine learning approach, allowing development of a non-parametric model. In comparison to the more traditional Cox regression approach, the RFS model relaxes the restrictive assumptions, such as the proportional hazard or the normal distribution, and as a result is reliable for handling cases of multicollinearity and of non-linear relationships between response variable and covariates. Indeed, the machine learning approach applied in this study has been already widely used in the clinical field to improve prediction accuracy in cardiovascular events,[17] to stage the esophageal cancer [24] to identify, with high performance, the disease-associated variables in metabolic genomic data [25] or to predict the mortality in rheumatoid arthritis [26]

We demonstrated that the RSF provides higher prediction performance, compared to both the Cox PHM and AIC-Cox model and to the "conventional" Cox regression model. Importantly, the RSF allowed the selection and combination of only those variables more accurately distinguishing patients at higher risk of converting to SPMS. Our model highlighted the early accumulation of focal and

diffuse GM damage as the most important determinants of the conversion to the progressive phase, supporting the crucial role played by the GM pathology in the development of late severe disability [13,15,27,28] In addition, older age at diagnosis was confirmed to independently predict a higher probability of becoming progressive [2,6,7] Interestingly, we also confirmed that male sex, high frequency of early relapses, and larger accumulation of T2 inflammatory lesions in the brain and in the cervical spine early in the disease course are predictors of poor prognosis [7–12] but their MD was found to be higher than the threshold (i.e., 4.21). This indicated a lower prediction ability, compared to other variables, which prevented their inclusion in the SP-RiSc. In line with recent findings from the UCSF cohort [29] this is explained by the strong predictive effect exerted by measures of cortical pathology, which plausibly overshadows the effect of WM damage variables on the outcome. Indeed, our alternative model, not including parameters of GM damage, confirmed the predictive value of early relapses and T2 lesions accumulation. However, by excluding GM variables, the model performed with much lower accuracy and specificity (Table S3 and Figure S2). Overall, the two models with and without using measures of cortical pathology, showed similar sensitivity, which, at least partially, is related to the intrinsic relationship between the brain WM and GM damage load, with reciprocal influence from both pathological perspective and imaging-analysis methods. The predictive model, including only variables of WM pathology, is suitable for being implemented and used in MS tertiary centres, where the GM damage is not routinely radiologically assessed. However, our results demonstrated that measurement of the cortical pathology significantly improves the prediction of the long-term outcome at individual level, and emphasizes the importance of evaluating the cortical damage, in addition to the focal inflammatory WM activity, in order to optimize the patients' management. The early accumulation of GM damage stands out as an essential therapeutic target for maximising the chances of achieving a good control of the disease activity. Previous efforts led to the development of prognostic tools based on demographic and clinical features [29,30] The SP-RiSc provides an individual accurate estimate of the risk of conversion into the progressive phase and it has been innovatively designed by including measures of cortical pathology to improve the prognostic tool's performance. This was confirmed both by the BS and C-index, showing good predictive accuracy of the RSF at different time points, and by the high sensitivity and specificity of the SP-RiSc in both the training and testing sets. At an individual level, a SP- RiSc ≥ of 17.7 indicates a 92% probability of converting to SPMS within 10 years from the disease diagnosis. In contrast, patients with SP-RiSc <17.7 had 87% probability of remaining in the relapsing–remitting phase.

We acknowledge some potential limitations. We did not include in our model the information on the type of symptoms at disease onset. However, the predictive role of the clinical features at presentation has been widely debated, with some studies showing worse prognosis among patients presenting with brainstem symptoms [31] but others indicating no clear effect on the clinical outcome [8] In addition, our analyses results are based on the definition of the clinical onset of the progressive phase, which has an inevitable degree of subjectivity. However, the long follow-up allowed confirmation retrospectively after 1 year the occurrence of the progressive course unrelated to any relapsing activity in all patients, providing high reliability to our findings. We deliberately opted not to use any minimal level of disability for capturing progression, as this approach would allow the uncovering of more comprehensively an element of progressive disease even in the early stage. The occurrence of progression independent of relapsing activity (PIRA) has been recently highlighted as an important component of the disability accumulation since the early phase of the disease [32] We are aware of the potential overlapping between PIRA events and the SP course, which nevertheless plausibly share the same underlying pathological mechanisms. Finally, the relatively small sample size might represent an additional limitation of our study. However, the RSF is advantaged by providing a good performance even when applied to small size dimensional datasets [33,34] Our results have been validated in two independent homogeneous datasets, although we acknowledge that an additional validation, especially on a larger independent cohort with neuroimaging data from different field strength MRI scanners, is important for confirming the score predictive properties and its application in the clinical setting.

Our tool could be implemented in the clinical context (Figure S1), especially in tertiary/academic centres where the use of non-routine GM damage MRI measures is widespread, mainly in the clinical trials scenarios. This will provide a basis for developing in future an SP-RiSc online platform with a simple interface available for the neurologist. The score can be updated with other biological baseline parameters, such as the CSF profile [35,36] in order to further improve the predictive tool's accuracy.

In conclusion, we propose the SP-RiSc as an implemented tool to reliably estimate early in the disease course the individual prognosis, which can be potentially helpful for optimizing the therapeutic strategies.

CONFLICT OF INTEREST

Massimiliano Calabrese received honoraria for research or speaking from Sanofi-Genzyme, Merck-Serono, Biogen Idec, Bayer, Novartis Pharma and funds for travel from Sanofi-Genzyme, Merck-Serono, Biogen Idec, Teva, Novartis Pharma, Roche and Bayer. Francesco Crescenzo received research support from Sanofi-Genzyme. All the other authors have nothing to disclose.

AUTHOR CONTRIBUTIONS

Anna Isabella Pisani: Data curation (lead); Formal analysis (lead); Methodology (lead); Software (equal); Writing-original draft (lead); Writing-review & editing (lead). **Antonio Scalfari:** Formal analysis (equal); Investigation (lead); Writing-original draft (lead); Writing-review & editing (lead). **Francesco Crescenzo:** Investigation (supporting); Writing-original draft (supporting); Writing-review & editing (equal). **Chiara Romualdi:** Data curation (supporting); Methodology (supporting); Supervision (lead); Writing-original draft (supporting); Writing-review & editing (lead). **Massimiliano Calabrese:** Conceptualization

(lead); Investigation (supporting); Supervision (lead); Writing-original draft (supporting); Writing-review & editing (lead).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Anna Isabella Pisani https://orcid.org/0000-0002-1170-8296
Massimiliano Calabrese https://orcid.org/0000-0002-3362-7403

REFERENCES

1. Koch-Henriksen N, Sørensen PS. The changing demographic pattern of multiple sclerosis epidemiology. *Lancet Neurol*. 2010;9(5):520-532.
2. Scalfari A, Neuhaus A, Degenhardt A, et al. The natural history of multiple sclerosis: a geographically based study 10 : relapses and long-term disability. *Brain*. 2010;1914-1929.
3. Fox RJ, Miller DH, Phillips JT, et al. Placebo-controlled phase 3 study of oral BG-12 or glatiramer in multiple sclerosis. *N Engl J Med*. 2012;367(12):1087-1097.
4. Scalfari A, Neuhaus A, Daumer M, Ebers GC, Muraro PA. Age and disability accumulation in multiple sclerosis. *Neurology*. 2011;77(13):1246-1252.
5. Feinstein A, Freeman J, Lo AC. Treatment of progressive multiple sclerosis: what works, what does not, and what is needed. *Lancet Neurol*. 2015;14(2):194-207. https://doi.org/10.1016/S1474-4422(14)70231-5
6. Tutuncu M, Tang J, Zeid NA, et al. Onset of progressive phase is an age-dependent clinical milestone in multiple sclerosis. *Mult Scler*. 2014;19(2):188-198.
7. Koch M, Mostert J, Heersema D, De Keyser J. Progression in multiple sclerosis: further evidence of an age dependent process. *J Neurol Sci*. 2007;255(1-2):35-41.
8. Confavreux C, Vukusic S, Adeleine P. Early clinical predictors and progression of irreversible disability in multiple sclerosis: an amnesic process. *Brain*. 2003;126(4):770-782.
9. Tremlett H, Yousefi M, Devonshire V, Rieckmann P, Zhao Y. Impact of multiple sclerosis relapses on progression diminishes with time. *Neurology*. 2009;73(20):1616-1623.
10. Leray E, Yaouanq J, Le Page E, et al. Evidence for a two-stage disability progression in multiple sclerosis. *Brain*. 2010;133(7):1900-1913.
11. Tintore M, Rovira À, Río J, et al. Defining high, medium and low impact prognostic factors for developing multiple sclerosis. *Brain*. 2015;138(7):1863-1874.
12. Brownlee WJ, Altmann DR, Prados F, et al. Early imaging predictors of long-term outcomes in relapse-onset multiple sclerosis. *Brain*. 2019;142(8):2276-2287.
13. Calabrese M, Romualdi C, Poretto V, et al. The changing clinical course of multiple sclerosis: a matter of gray matter. *Ann Neurol*. 2013;74(1):76-83.
14. Rocca M, Preziosa P, Copetti M, et al. Gray matter damage predicts the accumulation of disability and cognitive impairment 13 years later in patients with multiple sclerosis (S51.005). *Neurology*. 2012;78(Meeting Abstracts 1):S51.005.
15. Scalfari A, Romualdi C, Nicholas RS, et al. The cortical damage, early relapses, and onset of the progressive phase in multiple sclerosis. *Neurology [Internet]*. 2018;90(24):e2107-e2118.Available from http://www.neurology.org/lookup/doi/10.1212/WNL.0000000000005685
16. Polman CH, Reingold SC, Edan G, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". *Ann Neurol*. 2005;58(6):840-846. https://doi.org/10.1002/ana.20703
17. Hsich E, Gorodeski EZ, Blackstone EH, Ishwaran H, Lauer MS. Identifying important risk factors for survival in patient with systolic heart failure using random survival forests. *Circ Cardiovasc Qual Outcomes*. 2011;4(1):39-45.
18. Lublin FD, Reingold SC. Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. *Neurology*. 1996;46(4):907-911.
19. Association WM. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* [Internet]. 2013;310(20):2191-2194. https://doi.org/10.1001/jama.2013.281053
20. Miller DH, Barkhof F, Berry I, Kappos L, Scotti G, Thompson AJ. Magnetic resonance imaging in monitoring the treatment of multiple sclerosis: concerted action guidelines. *J Neurol Neurosurg Psychiatry*. 1991;54(8):683-688.
21. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. *Classif Regres Trees*. 2017;2000:1-358.
22. Ishwaran H, Kogalur UB, Gorodeski EZ, Minn AJ, Lauer MS. High-dimensional variable selection for survival data. *J Am Stat Assoc*. 2010;105(489):205-217.
23. Mogensen UB, Ishwaran H, Gerds TA. Evaluating random forests for survival analysis using prediction error curves. *J Stat Softw* [Internet]. 2012;50(11):1-23.Available from http://www.ncbi.nlm.nih.gov/pubmed/25317082%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4194196
24. Ishwaran H, Blackstone EH, Apperson-Hansen C, Rice TW. A novel approach to cancer staging: application to esophageal cancer. *Biostatistics*. 2009;10(4):603-620.
25. Dietrich S, Floegel A, Troll M, et al. Random Survival Forest in practice: a method for modelling complex metabolomics data in time to event analysis. *Int J Epidemiol*. 2016;45(5):1406-1420.
26. Lezcano-Valverde JM, Salazar F, León L, et al. Development and validation of a multivariate predictive model for rheumatoid arthritis mortality using a machine learning approach. *Sci Rep*. 2017;7(1):1-10.
27. Filippi M, Preziosa P, Copetti M, et al. Gray matter damage predicts the accumulation of disability 13 years later in MS. *Neurology*. 2013;81(20):1759-1767.Available from: http://n.neurology.org/content/81/20/1759.abstract
28. Fisniku LK, Chard DT, Jackson JS, et al. Gray matter atrophy is related to long-term disability in multiple sclerosis. *Ann Neurol*. 2008;64(3):247-254.
29. Cree BAC, Hollenbach JA, Bove R, et al. Silent progression in disease activity–free relapsing multiple sclerosis. *Ann Neurol*. 2019;85(5):653-666.
30. Manouchehrinia A, Zhu F, Piani-Meier D, et al. Predicting risk of secondary progression in multiple sclerosis: a nomogram. *Multiple Sclerosis Journal*. 2019;25(8):1102-1112.
31. Misicka E, Sept C, Briggs FBS. Predicting onset of secondary-progressive multiple sclerosis using genetic and non-genetic factors. *J Neurol*. 2020;267(8):2328-2339. https://doi.org/10.1007/s00415-020-09850-z
32. Kappos L, Wolinsky JS, Giovannoni G, et al. Contribution of relapse-independent progression vs relapse-associated worsening to overall confirmed disability accumulation in typical relapsing multiple sclerosis in a pooled analysis of 2 randomized clinical trials. *JAMA Neurol*. 2020;77(9):1132-1140.
33. Biau G, Scornet E. A random forest guided tour. *TEST* [Internet]. 2016;25(2):197-227. https://doi.org/10.1007/s11749-016-0481-7
34. Schmid M, Welchowski T, Wright MN, Berger M. Discrete-time survival forests with Hellinger distance decision trees. *Data Min Knowl Disc*. 2020;34(3):812-832. https://doi.org/10.1007/s10618-020-00682-z
35. Magliozzi R, Howell OW, Nicholas R, et al. Inflammatory intrathecal profiles and cortical damage in multiple sclerosis. *Ann Neurol*. 2018;83(4):739-755.

36. Magliozzi R, Scalfari A, Pisani AI, Ziccardi S, Marastoni D, Pizzini FB, The CSF profile linked to cortical damage predicts multiple sclerosis activity. *Ann Neurol* [Internet]. 2020. 88(3):562–573. https://doi.org/10.1002/ana.25786

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Pisani AI, Scalfari A, Crescenzo F, Romualdi C, Calabrese M. A novel prognostic score to assess the risk of progression in relapsing–remitting multiple sclerosis patients. *Eur J Neurol*. 2021;28:2503–2512. https://doi.org/10.1111/ene.14859