MULTIPLE SCLEROSIS JOURNAL | MSJ

Original Research Paper

Multiple Sclerosis Severity Score (MSSS) improves the accuracy of individualized prediction in MS

Tomas Kalincik*, Ilya Kister*, Tamar E Bacon, Charles B Malpas, Sifat Sharmin, Dana Horakova, Eva Kubala-Havrdova, Francesco Patti, Guillermo Izquierdo, Sara Eichau, Serkan Ozakbas, Marco Onofrj, Alessandra Lugaresi, Alexandre Prat, Marc Girard, Pierre Duquette, Pierre Grammond, Patrizia Sola, Diana Ferraro, Raed Alroughani, Murat Terzi, Cavit Boz, Francois Grand'Maison, Roberto Bergamaschi, Oliver Gerlach, Maria J Sa, Ludwig Kappos, Elisabetta Cartechini, Jeannette Lechner-Scott, Vincent van Pesch, Vahid Shaygannejad, Franco Granella, Daniele Spitaleri, Gerardo Iuliano, Davide Maimone, Julie Prevost, Aysun Soysal, Recai Turkoglu, Radek Ampapa, Helmut Butzkueven and Gary Cutter; for MSBase Study Group

Multiple Sclerosis Journal

1–10

DOI: 10.1177/13524585221084577

Abstract

Background: The MSBase prediction model of treatment response leverages multiple demographic and clinical characteristics to estimate hazards of relapses, confirmed disability accumulation (CDA), and confirmed disability improvement (CDI). The model did not include Multiple Sclerosis Severity Score (MSSS), a disease duration-adjusted ranked score of disability.

Objective: To incorporate MSSS into the MSBase prediction model and compare model accuracy with and without MSSS.

Methods: The associations between MSSS and relapse, CDA, and CDI were evaluated with marginal proportional hazards models adjusted for three principal components representative of patients' demographic and clinical characteristics. The model fit with and without MSSS was assessed with penalized r2 and Harrell C.

Results: A total of 5866 MS patients were started on disease-modifying therapy during prospective follow-up (age 38.4 ± 10.6 years; 72% female; disease duration 8.5 ± 7.7 years). Including MSSS into the model improved the accuracy of individual prediction of relapses by 31%, of CDA by 23%, and of CDI by 24% (Harrell C) and increased the amount of variance explained for relapses by 49%, for CDI by 11%, and for CDA by 10% as compared with the original model.

Conclusion: Addition of a single, readily available metric, MSSS, to the comprehensive MSBase prediction model considerably improved the individual accuracy of prognostics in MS.

Keywords: Multiple sclerosis, Multiple Sclerosis Severity Score (MSSS), relapse prediction, prognostics

Date received: 30 October 2021; revised: 20 January 2022; accepted: 3 February 2022

Correspondence to:
I Kister
Neurology, NYU School of Medicine, 240 East 38th Street, New York, NY 10026, USA.
ilya.kister@nyulangone.org

Tomas Kalincik
Charles B Malpas
Sifat Sharmin
CORe, Department of Medicine, The University of Melbourne, Melbourne, VIC, Australia/MS Centre, Department of Neurology, The Royal Melbourne Hospital, Melbourne, VIC, Australia

Ilya Kister
Tamar E Bacon Neurology, NYU School of Medicine, New York, NY, USA

Dana Horakova
Eva Kubala-Havrdova
Department of Neurology, Charles University in Prague, Prague, Czech Republic

Francesco Patti
GF Ingrassia Department, University of Catania, Catania, Italy

Guillermo Izquierdo
Sara Eichau
Hospital Universitario Virgen Macarena, Sevilla, Spain

Serkan Ozakbas
Dokuz Eylul University, Izmir, Turkey

Marco Onofrj
Department of Neuroscience, Imaging, and Clinical Sciences, University G. d'Annunzio, Chieti, Italy

Introduction

Severity score represents disease duration–adjusted mean rank of disability in multiple sclerosis (MS) patients from the reference population.[1,2] Severity rank was validated as a group-level predictor of disability in MS[1,3,4] and explained a large proportion of the individual variance in MS disability outcomes.[5] The MSBase predictive model of treatment response is a powerful algorithm for predicting MS clinical outcomes conditional on treatment exposure, demographic, and clinical variables. However, the model did not include severity rank among its predictor variables.[6] We hypothesized that incorporating the Multiple Sclerosis Severity Score (MSSS) into the MSBase predictive model would improve the model's power for individual-level prediction and validate the

Alessandra Lugaresi
IRCCS Istituto delle Scienze Neurologiche di Bologna, Bologna, Italy/Dipartimento di Scienze Biomediche e Neuromotorie, Università di Bologna, Bologna, Italy

Alexandre Prat
Marc Girard
Pierre Duquette
Universite de Montreal and CHUM, Montreal, QC, Canada

Pierre Grammond
CISSS Chaudi're-Appalache, Levis, QC, Canada

Patrizia Sola
Neurology Unit, Department of Neuroscience, Azienda Ospedaliera Universitaria, Modena, Italy

Diana Ferraro
Neurology Unit, Department of Neuroscience, Azienda Ospedaliera Universitaria, Modena, Italy/Department of Biomedical, Metabolic and Neurosciences, University of Modena and Reggio Emilia, Modena, Italy

Raed Alroughani
Division of Neurology, Department of Medicine, Amiri Hospital, Sharq, Kuwait

Murat Terzi
Medical Faculty, 19 Mayis University, Samsun, Turkey

Cavit Boz
KTU Medical Faculty Farabi Hospital, Trabzon, Turkey

Francois Grand'Maison
Neuro Rive-Sud, Greenfield Park, QC, Canada

Roberto Bergamaschi
IRCCS Mondino Foundation, Pavia, Italy

Oliver Gerlach
Department of Neurology, Zuyderland Medical Centre, Sittard-Geleen, The Netherlands/School for Mental Health and Neuroscience, Department of Neurology, Maastricht University Medical Centre, Maastricht, The Netherlands

Maria J Sa
Hospital S. João, Porto, Portugal; University Fernando Pessoa, Porto, Portugal

Ludwig Kappos
Research Center for Clinical Neuroimmunology and Neuroscience(RC2NB) and MS Center, Departments of Head, Spine and Neuromedicine, Clinical Research and Biomedicine, University Hospital Basel, University of Basel, Basel, Switzerland

Elisabetta Cartechini
UOC Neurologia, Azienda Sanitaria Unica Regionale Marche—AV3, Macerata, Italy

prognostic value of MSSS against the background of other characteristics that are already included in the model.

Methods

The MSBase Registry is a global collaborative research group that prospectively collects outcomes data from MS treatment centers using an Internet-based, physician-owned, and operated system. The structure and operations of MSBase have been described in detail (www.msbase.org).[7] MSBase holds approval from Melbourne Health Human Research Ethics Committee and approval or exemption at each contributing center as required (WHO International Clinical Trials Registry Platform (ACTRN12605000455662)).

The data were extracted from the MSBase registry on 18 January 2019. Inclusion/exclusion criteria for this study were as follows: diagnosis of MS or clinically isolated syndrome based on the 2005 or 2010 revised McDonald criteria;[8,9] commencement of a DMT during the prospectively recorded follow-up (irrespective of the previous exposure to DMTs); minimum pre-DMT follow-up of 6 months (except patients with <6-month disease duration); minimum on-treatment prospective follow-up of 6 months; availability of the minimum dataset (i.e. patient sex, year of birth, year of the first clinical presentation, disease course, treating center, and ⩾2 on-treatment clinical visits with recorded Extended Disability Status Scale (EDSS) scores); disability score (EDSS) recorded between 6 months before and 1 month after the index DMT commencement, but not within 30 days after a previous relapse. Patients with inactive primary progressive MS were excluded.[10] Flow diagram of patient disposition is shown in Supplemental Figure 1. Baseline was defined as the date of a start of a new disease-modifying therapy, with a recorded EDSS score and followed by ⩾6 months of recorded clinical follow-up, including two EDSS scores separated by ⩾6 months. MSSS was calculated at all eligible baselines, using the reference dataset.[1]

Three principal components summarizing patient demographic and clinical characteristics were calculated at each baseline, as per the methodology of the MSBase predictive model of treatment response.[6] In brief, the model used principal component analysis to reduce the dimensionality of the matrix of demographic and clinical characteristics that were used to predict relapses, confirmed disability accumulation (CDA), and confirmed disability improvement (CDI). The analysis defined three principal components based on 55 patient characteristics, which included sex, age, disease duration and course, inaugural symptoms, EDSS, EDSS trajectory and slope, functional scores, current and past DMT, time on DMT, relapse rate and severity, brain and spinal MRI lesion burden and activity, and CSF findings (for a full list of variables, see Table 3 in Kalincik et al.[6]). Principal components were largely reflective of: (1) disability and previous exposure and response to DMTs; (2) history of relapses (including their frequency, severity, and recovery), age, disease phenotype; and (3) the initial multiple sclerosis symptoms. The principal components were used to adjust the predictive models of MSSS for the known prognostic demographic and clinical characteristics.

The three main outcomes of the model were: (1) first postbaseline relapse, defined as an occurrence of new symptoms or worsening of existing symptoms in the absence of febrile disease; (2) CDA, defined as an increase in EDSS score by 1.5 steps if the previous EDSS was 0, by 1 step if the previous EDSS was 1–5.5 and by 0.5 steps if the previous EDSS was ⩾6, confirmed over ⩾6 months (with the confirmation EDSS recorded >30 days following a previous relapse) and sustained for the remainder of the follow-up; and (3) CDI, defined as a decrease in EDSS by 1.5 steps if EDSS was 1.5 (1 step if EDSS was 2–6, and 0.5 step if EDSS was ⩾6.5), confirmed over ⩾6 months and sustained for the remainder of the follow-up. Any CDA and CDI events that occurred between baseline and the first postbaseline visit, as well as any event within 6 months after baseline, were excluded to eliminate the effects of therapeutic lag and recent prebaseline relapses on disability.[11] As the associations between MSSS and the outcomes of interest adjusted for the three principal components were independent of treatment[12] (Supplemental Table 1), we combined all treatment datasets into a single composite dataset allowing only one entry per patient. The earliest recorded eligible baseline was retained for each patient to maximize the available postbaseline follow-up.

We first reported cumulative hazards of post-baseline relapse, CDA and CDI per MSSS decile (MSSS range from 0 to 10, with the first decile including the least disabled 10% of patients (relative to the subset of patients with similar disease duration), and the 10th decile encompassing the 10% of the most disabled patients. Then, the associations between MSSS and the first postbaseline relapse, CDA, and CDI were evaluated with marginal Cox proportional hazards models adjusted for three principal components representative of patients' demographic and clinical

characteristics. The accuracy of the predicted risks of relapses, CDA, and CDI were compared between models with and without MSSS in a nonoverlapping validation cohort (10% of the eligible cohort) using Harrell C (concordance index for censored data, interpreted as the proportion of observations where higher predicted risk concurred with a shorter time to event). Because Harrell C is nonparametric and only provides a relative evaluation of the overall performance of the model, we also developed a more granular and quantitative assessment of model accuracy using the following approach: first, we established the predicted hazards of each of the three outcomes of interest (relapse, CDA, and CDI) at 2- and 4-year benchmarks. We then stratified patients into 10% strata of predicted hazard, ranging from 0% to 100%. We calculated the proportions of patients in each 10% stratum who experienced an event within the 2- and 4-year timeframe. Because cumulative hazard at a single timepoint is closely related to the expected proportion of patients with an event, we can directly compare predicted hazards with observed proportions of patients with the event. Finally, we compared the model fit with and without MSSS with penalized r^2.

Results

The study sample consisted of 5866 treated MS patients (age (mean, (SD)) was 38.4 (10.6) years; 72% were female; disease duration was 8.5 (7.7) years; disease subtype was relapsing-remitting in 86%; median EDSS was 2 (quartiles 1–3)). Disease-modifying therapies are shown in Supplemental Table 2. The characteristics of patients in each of the 10 MSSS deciles at the treatment start are shown in Table 1. The sample is weighted toward milder disease as compared to the reference MSSS population from the pretreatment era,[1] likely a reflection of the well-documented shift in disease severity in the past decades.[13] Reassuringly, the trends for age with severity decile were minimal and largely accounted for by the adjustment of MSSS for MS duration. The percentage of men within each decile increased with increasing decile severity in line with prior analyses of the MSBase dataset in which men tended to have faster disability progression than women.[14]

The original MSBase predictive model of individual treatment response enables prediction that is specific to MS DMTs,[6] but the prognostic value of MSSS was shown to be ubiquitous across the studied therapies: 1-point rank increase in MSSS would, on average, be associated with 1.5–2.8× increase in the cumulative hazard of relapses, 1.5–2× increase in the cumulative hazard of CDA, and a 2.2–3× increase in the

Table 1. Characteristics of patients in each of the 10 MSSS deciles.

	MSSS-1	MSSS-2	MSSS-3	MSSS-4	MSSS-5	MSSS-6	MSSS-7	MSSS-8	MSSS-9	MSSS-10
Number of patients	855	470	1027	407	828	956	260	583	230	250
Age in years, mean (SD)	34.7 (10.1)	40.3 (9.6)	34.1 (9.9)	38.4 (9.9)	35.1 (10.3)	34.8 (10.8)	40.6 (10.6)	36.2 (11.0)	37.9 (11.9)	38.6 (11.5)
Female %	72.2	73.4	71.0	71.5	72.2	67.1	65.8	66.6	61.7	64.8
Disease duration in years, mean (SD)	6.5 (8.2)	12.4 (8.2)	4.8 (6.9)	9.4 (7.7)	3.8 (5.7)	3.1 (5.6)	7.6 (5.5)	2 (3.3)	3.6 (4.5)	1.7 (2.2)
Duration of follow-up in months, mean (SD)	107.4 (71.8)	147.2 (73.7)	106.0 (74.2)	145.3 (82)	104.8 (69.3)	99.1 (64.8)	147.8 (70.7)	92.9 (63)	112.4 (61.9)	89.5 (56.2)
Baseline EDSS, median; mean (SD)	0; 0.3 (0.6)	1.5; 1.7 (0.7)	1; 1.5 (0.8)	2; 2.5 (1.1)	1.5; 2.1 (1)	2; 2.5 (1.1)	3.5; 4 (1.2)	3; 3.2 (1)	4; 4.5 (1.2)	5; 5.2 (1.1)
Annual relapse rate, mean (SD)	1.5 (1.6)	1.7 (2.3)	1.5 (1.4)	1.7 (2)	1.7 (1.6)	1.7 (1.6)	1.9 (2)	1.7 (1.4)	2 (1.7)	1.9 (1.5)

EDSS: Extended Disability Status Scale; MSSS: Multiple Sclerosis Severity Scale; SD: standard deviation.
MSSS-1 refers to MSSS scores from 0 to 1 (first decile); MSSS-2 from 1 to 2 (second decile), and so on.

Jeannette Lechner-Scott
School of Medicine and Public Health, The University of Newcastle, Newcastle, NSW, Australia

Vincent van Pesch
Cliniques Universitaires Saint-Luc, UCLouvain, Brussels, Belgium

Vahid Shaygannejad
Isfahan University of Medical Sciences, Isfahan, Islamic Republic of Iran

Franco Granella
Department of Medicine and Surgery, University of Parma, Parma, Italy

Daniele Spitaleri
Azienda Ospedaliera di Rilievo Nazionale San Giuseppe Moscati Avellino, Avellino, Italy

Gerardo Iuliano
Ospedali Riuniti di Salerno, Salerno, Italy

Davide Maimone
Neurology Unit, Piazza S. Maria di Gesù 5, Catania, Italy

Julie Prevost
CSSS Saint-Jérôme, Saint-Jerome, QC, Canada

Aysun Soysal
Bakirkoy Education and Research Hospital for Psychiatric and Neurological Diseases, Istanbul, Turkey

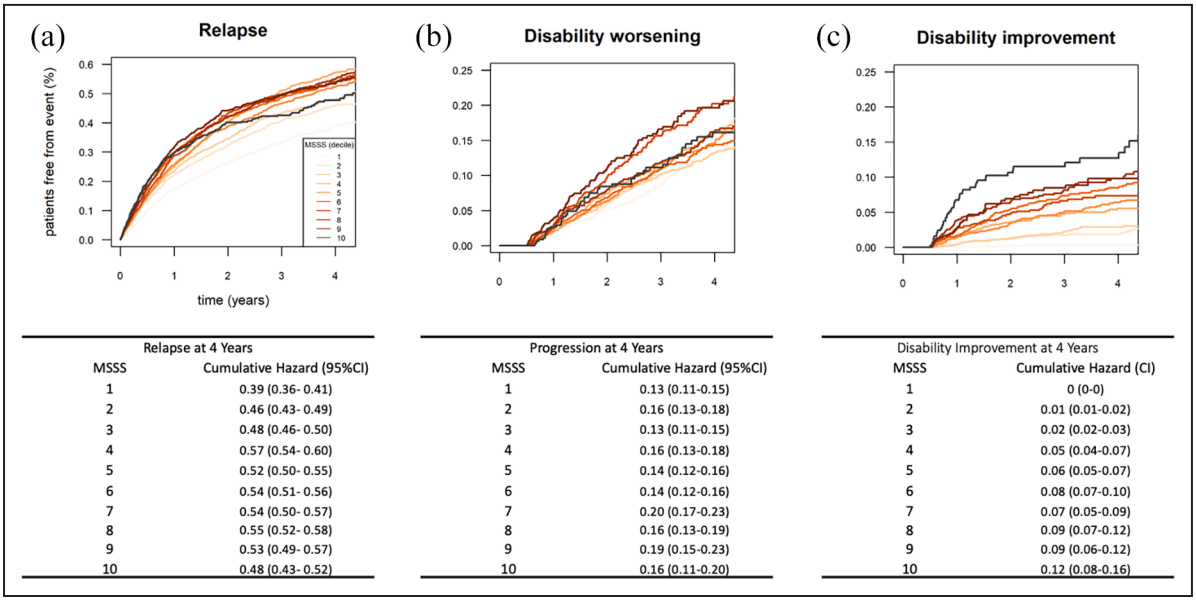
Recai Turkoglu
Haydarpasa Numune Training and Research Hospital, Istanbul, Turkey

Radek Ampapa
Nemocnice Jihlava, Jihlava, Czech Republic

Helmut Butzkueven
Central Clinical School, Monash University, Melbourne, VIC, Australia

Gary Cutter
Department of Biostatistics, UAB School of Public Health, Birmingham, AL, USA

*Tomas Kalincik and Ilya Kister contributed equally to this study.



Relapse at 4 Years	
MSSS	Cumulative Hazard (95%CI)
1	0.39 (0.36- 0.41)
2	0.46 (0.43- 0.49)
3	0.48 (0.46- 0.50)
4	0.57 (0.54- 0.60)
5	0.52 (0.50- 0.55)
6	0.54 (0.51- 0.56)
7	0.54 (0.50- 0.57)
8	0.55 (0.52- 0.58)
9	0.53 (0.49- 0.57)
10	0.48 (0.43- 0.52)

Progression at 4 Years	
MSSS	Cumulative Hazard (95%CI)
1	0.13 (0.11-0.15)
2	0.16 (0.13-0.18)
3	0.13 (0.11-0.15)
4	0.16 (0.13-0.18)
5	0.14 (0.12-0.16)
6	0.14 (0.12-0.16)
7	0.20 (0.17-0.23)
8	0.16 (0.13-0.19)
9	0.19 (0.15-0.23)
10	0.16 (0.11-0.20)

Disability Improvement at 4 Years	
MSSS	Cumulative Hazard (CI)
1	0 (0-0)
2	0.01 (0.01-0.02)
3	0.02 (0.02-0.03)
4	0.05 (0.04-0.07)
5	0.06 (0.05-0.07)
6	0.08 (0.07-0.10)
7	0.07 (0.05-0.09)
8	0.09 (0.07-0.12)
9	0.09 (0.06-0.12)
10	0.12 (0.08-0.16)

Figure 1. Hazard of a relapse (a), confirmed disability accumulation (CDA) (b), and confirmed disability improvement (CDI) (c) for each Multiple Sclerosis Severity Score (MSSS) decile. The upper figures show results of unadjusted analysis for hazards of relapse (a), CDA (b), and CDI (c) for each MSSS decile. The tables under respective figures provide numerical values of cumulative hazards at year 4.

Table 2. Comparison of MSSS-enhanced MSBase prediction model with the original MSBase prediction model (without MSSS).

	Relapse β ± robust SE, *p*	CDA β ± robust SE, *p*	CDI β ± robust SE, *p*
MSSS	0.5 ± 0.01, <0.0001	0.44 ± 0.03, <0.0001	1.01 ± 0.06, <0.0001
MSSS × time (interaction)	−0.00065 ± 0.00003, <0.0001	−0.0003 ± 0.00003, <0.0001	−0.0004 ± 0.00004, <0.0001
D1	0.48 ± 6.34, 0.94	29.3 ± 15.0, 0.05	14.3 ± 20.3, 0.48
D2	24.9 ± 7.36, 0.001	-84.9 ± 16.8, <0.0001	95.3 ± 24.8, 0.0001
D3	33.6 ± 34.4, 0.33	5.56 ± 78.4, 0.94	293 ± 101, 0.01
Harrell's C Index (95% confidence interval	0.87 (0.86–0.88)	0.85 (0.83–0.88)	0.91 (0.87–0.95)
Improvement in model fit (difference in penalized R^2)	0.49	0.11	0.10

CDA: confirmed disability accumulation; CDI: confirmed disability improvement; MSSS: Multiple Sclerosis Severity Scale; SE: standard error.
The table shows that MSSS was a significant predictor of all three outcomes of interest in the enhanced MSBase prediction model. Improvement in model fit after MSSS was incorporated in the model is shown in the last row (assessed as difference in penalized R^2). D1, D2, and D3 are the three principal components of the model.[6]

cumulative hazard of CDI across all therapies (Supplementary Table 1). We therefore conducted all analyses with the combined dataset of MSBase patients who were newly started on a DMT during prospective follow-up, irrespective of DMT identity. Unadjusted hazards of relapse for patients in each MSSS decile as a function of follow-up time (up to 4 years) are shown in Figure 1(a). Overall, there was a tendency for the increasing hazard of relapse over a 4-year period with increasing decile (excepting the 10th decile). Thus, the relapse hazard for the first MSSS decile was 0.39, while for the ninth decile, it was 0.52. For CDI, there was a consistent increase in the cumulative hazard of event from the least severe (0–0.01) to the most severe (0.9–1.12) deciles (Figure 1(c)). For CDA, the trend was less consistent on the whole, but the lowest severity deciles also had the lowest hazards of disability accumulation (Figure 1(b)).

The results of the MSSS-enhanced MSBase prediction model as compared with the original prediction model without MSSS are shown in Table 2. In the enhanced model, MSSS showed a high correlation

with the hazard of relapse, CDA, and CDI. Improvement in model fit was greatest for relapse hazard—the difference in penalized r^2 of 0.49. For CDA, the difference in penalized r^2 was 0.11, and for CDI—the difference was 0.10. Thus, MSSS was an independent predictor of relapses and disability on a group level in a model that already accounts for dozens of clinical and demographic factors. However, even the predictive power of the enhanced model decreased with the duration of follow up as can be seen from the highly significant MSSS × time interaction term for all three outcomes of interest.

The improvement of the overall accuracy of individualized prediction in the model was assessed with Harrell C. The concordance for relapse events improved from 56% (95% confidence interval 54%–57%) in a baseline model without MSSS to 87% (95% confidence interval 86%–88%) in the model that included MSSS—a 31% improvement in prediction accuracy. The concordance index for CDA improved from 63% (95% confidence interval 61%–66%) for the original model to 85% (95% confidence interval 83%–88%) in the model version that incorporated MSSS—a 23% improvement. The improvement in concordance for CDI events improved from 67% (95% confidence interval 63%–71%) to 91% (95% confidence interval 87%–95%)—a 24% improvement. Results of the more granular assessment of model accuracy in which the predicted hazards of each of the three outcomes of interest (relapse, CDA, and CDI) at 2- and 4-year benchmarks were compared to the expected proportion of patients with events of interest are summarized in Table 3. The model reproduces the expected gradient of observed progression events with the increasing predicted hazard but tends to overestimate the hazard throughout the range. This can be seen for both CDA (column 1 for 0–2 years and column 2 for 0–4 years in Table 3) and CDI (columns 3 for 0–2 years and 4 for 0–4 years). For relapses, the model increasingly overestimated the risk for higher deciles (columns 5 and 6). Overall, the model was conservative (overestimating the risk of events) and was more accurate in predicting changes in disability than relapses.

Figure 2(a)–(c) shows the predicted and actual outcomes with regard to disability and relapses for three individuals with MS from the MSBase registry whose demographic and disease-related information is provided in the legend. Figure 2 illustrates the ability of the model to differentiate among patients with similar characteristics, such as patients #1 (top row) and patient #3 (bottom row), who are both in their early 40s and both have a low EDSS score of 2. Because patient #1 has been living with MS for 11 years, her MSSS is only 2.1 (i.e. she is less disabled than the majority of patients with disease duration of 11 years), whereas patient #3, who had MS for only 1 year, has MSSS of 5.9 (i.e. his disability is higher than most patients within one year of disease onset). The predicted hazard of relapse is much higher for patient #3 than for patient #1, while the hazard of CDA is lower for patient #3 than for patient #1. This example demonstrates how the enhanced MSBase prediction model can effectively discriminate between expected outcomes in two patients of similar age and disability. The model also captures the possibility of disability improvement even in patients with relatively fast disease progression. Thus, patient #2 has a relatively high MSSS of 8 (she accumulated considerable disability in only 6 years of disease), yet the model predicts equal hazards of CDI as for CDA for her. Indeed, her subsequent course shows improvement in EDSS over a 4-year period.

Discussion

Many clinical and radiographic prognostic markers of more severe disease course have been described in MS: vascular[15] and psychiatric comorbidities[16]; frequent relapses during early disease[17]; incomplete recovery from the initial relapse[18]; high T2 and T1 lesion burden on brain MRI[19,20]; multiple Gadolinium-enhancing lesions in the brain; spinal cord lesions[21]; early cerebellar involvement[22]; and others. The associations of these markers with disease outcomes have been established at the group level. To improve their prognostic yield and facilitate use in individuals, recent studies combined several variables. For example, EDSS ≥ 3, pyramidal signs in the first year of disease, and age > 35 at symptom onset carry a 32% risk of developing substantial disability within 10 years of MS onset[23]; fortunately, only a small minority of patients has all these additive risk factors for "aggressive MS" course. Another scoring system, the modified Rio score, has been successfully validated in patients treated with interferon-beta.[24] These efforts notwithstanding, individualized prediction remains, in the words of a recent review, "a major unmet need in MS."[25]

To address this unmet need, a prototype of an individual model of treatment response was developed using data from the MSBase registry.[6] The model leverages dozens of individual demographic and clinical characteristics to estimate future hazards of relapses, disability worsening, and disability improvement. The discriminating ability of this model is limited by the relatively wide variance of the estimates of

Table 3. Individualized prediction of relapses, CDA, and CDI.

	CDA		CDI		Relapse	
	Years 0–2	Years 0–4	Years 0–2	Years 0–4	Years 0–2	Years 0–4
Predicted hazard	Patients with observed events/patients with predicted risk (percentage)		Patients with observed events/patients with predicted risk (percentage)		Patients with observed events/patients with predicted risk (percentage)	
0–10%	175/5039 (3%)	80/1998 (4%)	97/4836 (2%)	98/4294 (2%)	–	–
10–20%	29/253 (11%)	229/2855 (8%)	36/374 (10%)	55/700 (8%)	–	–
20–30%	4/26 (15%)	59/343 (17%)	8/32 (25%)	30/202 (15%)	–	–
30–40%	–	24/127 (19%)	–	9/40 (23%)	151/632 (24%)	–
40–50%	–	7/46 (15%)	–	1/6 (17%)	955/2877 (33%)	0/1 (0%)
50–60%	–	4/17 (24%)	–	–	700/1736 (40%)	152/516 (29%)
60–70%	–	1/8 (13%)	–	–	204/453 (45%)	708/1845 (38%)
70–80%	–		–	–	41/114 (36%)	840/1830 (46%)
80–90%	–	–	–	–	16/32 (50%)	467/983 (48%)
90–100%	–	–	–	–	7/12 (58%)	218/420 (52%)
⩾100	–	–	–	–	4/10 (40%)	119/271 (44%)

CDA: confirmed disability accumulation; CDI: confirmed disability improvement.
The table compared expected hazards of relapse, CDA or CDI recorded during 2 and 4 years with proportions of patients in whom respected event was observed after the prediction baseline. To illustrate the use of the table, consider the patients for whom the expected rate of disability progression within the first 2 years was 0%–10% (first row, second column: "CDA, 0–2 years"). Among the patients with the predicted 0%–10% cumulative hazard of CDA events, the percentage of patients who had observed progression within 2 years of baseline was 3% (175 of 5039 patients), which is within the range predicted by the model. Among the patients in whom the model assessed the cumulative hazard of CDA to be 10%–20% (second row), the percentage of patients with observed progression was 11% (29 of 253), which is also within the predicted range. Of patients with 20%–30% cumulative hazard of CDA only 15% (4 of 26) experienced disability progression, which is less than expected by the model. Because the subsets of patients within each decile of risk beyond 30% were less than 10 patients, the data for these are not shown in the table.

hazards of disability progression and relapses for individual patients. The current work aimed to enhance the accuracy of individual prediction of the model by incorporating the readily available disability rank normalized by disease duration—the MSSS[1] into the model. Our initial analyses have established that the additional predictive value of MSSS in the MSBase model does not depend on treatment,[12] possibly because the effectiveness of a given DMT across different rates of disability accrual is roughly proportional (i.e. prognostic value of MSSS is independent of the effectiveness of medications). Therefore, in the current work, we combine all treatments into a single composite dataset allowing only one entry per patient to test the hypothesis that MSSS adds meaningfully to the individual-level prognostics in a comprehensive prediction model.

We show that the addition of MSSS to the model improved the accuracy of individual prediction of relapses by 31% (from 56% in a baseline model to 87% in the model including MSSS), of CDA—by 23% (from 63% in the original model to 86% in the enhanced model) and of CDI—by 25% (from 67% to 91%). The incorporation of MSSS also increased the amount of variance explained for relapses by 49%, for CDI—by 11%, for CDA—by 10% as compared with the original model. These gains in accuracy and in explained variance with the addition of just a single variable to a model that already accounts for 55 combined variables—including EDSS and disease duration—can be partly explained by the combination of these variables into principal components (to reduce the dimensionality of the model), and partly by the fact that MSSS is not a duration-adjusted disability, but ranking of duration-adjusted disability. MSSS contains information about the distribution of scores for patients with similar disease duration in the reference cohort and, as such, is much more prognostically informative than duration-adjusted disability.[1] MSSS is analogous to normative percentiles on a child's growth chart, which reflect distributions of height/weight in children of the same age across the population and are therefore more valuable for gauging child's development and predicting growth than the actual height/weight at any timepoint.

Clinically, the predictions made by our prognostic model are conservative, that is, they tend to overestimate the risk of events. This is, on the whole, preferable to underestimating the severity of a disease that exerts such high cost in disability-adjusted life-years (DALYs),[26] and for which potent therapies are available that can slow down the disease, especially when

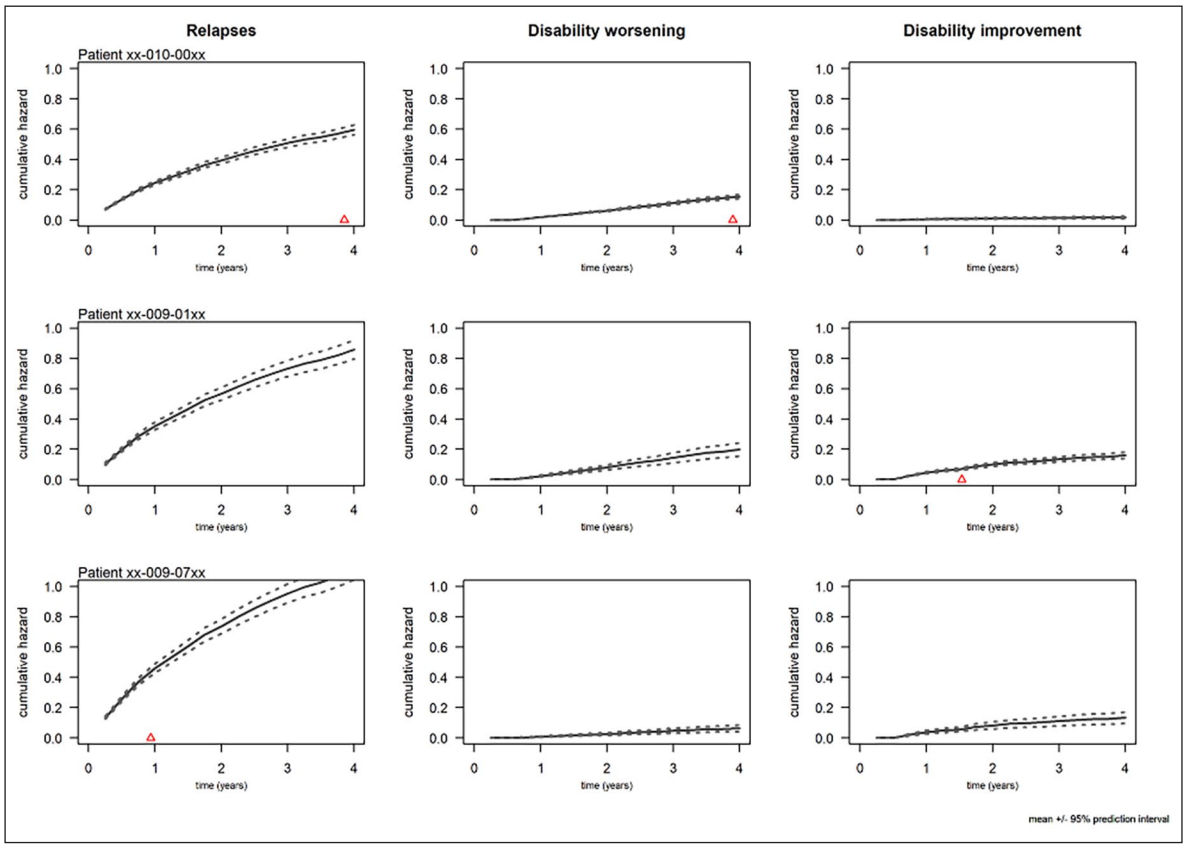


Figure 2. Examples of model predictions for three patients in MSBase. Each row represents predicted and actual outcomes for a patient with relatively mild disease (MSSS 2.1, top row), severe disease (middle row, MSSS of 8.0) and "average" disease course (bottom row, MSSS 5.9). The first patient (top panel) is a 41-year-old woman with relapsing MS for 11 years. Her EDSS at baseline was 2 (MSSS: 2.1). She suffered only one relapse prior to baseline. The model predicts that her hazard of relapse over the next 4 years (left upper graph) is approximately 0.5; hazard of disability worsening is 0.2 (middle-upper graph) and hazard of disability improvement is slightly less than 0.1 (right upper graph). The red triangle in the "relapse" and "disability worsening" graphs indicate when actual relapse and disability progression were observed—both events occurred toward the end of the 4-year follow-up. The absence of the red triangle in "disability improvement" graph means that disability improvement was not observed at any point during the 4-year follow-up. The second patient (middle panel) is a 30-year-old woman who had MS for 6 years, and reached EDSS of 5.5 (MSSS: 8.0). She suffered one relapse in the prior year. The model predicts cumulative hazard of relapse is 0.8, of disability worsening 0.2, and of disability improvement 0.2 over 4 years. This patient did not experience relapses or disability progression during follow-up but did have disability improvement after 1.5 years of follow-up. The third patient (bottom panel) is a 40-year-old man with relapsing MS who had EDSS = 2 after 1 year of disease (MSSS: 5.9). He suffered two relapses in the year preceding the baseline. The model predicts high hazard of relapses, but low hazard of disability accumulation or improvement. The patient had a relapse after about 1 year of follow-up, and his EDSS remained stable.

started early.[27,28] On the other hand, false-positive prediction of worse outcomes may subject patients to higher risks of treatment complications and have a negative impact on patients' mental wellbeing. As individualized prediction algorithms in MS come of age, clinicians would need to take into account limitations of prognostic algorithms and acknowledge an inevitable element of the unpredictability of MS courses.

Our results confirm the value for MSSS for individual-level prognostics, but also expand our understanding of MSSS as a group-level predictor of disability outcomes.[1,3,4] We show that—on a group level—hazard of relapses, CDI and, to a lesser extent, CDA are higher in patients with higher MSSS. The increased probability of disability improvement as well as the risk of disability worsening with higher MSSS may appear counter-intuitive. One possible explanation is that "high-MSSS" patients comprise a heterogeneous group: those with rapid disease course who are at a high risk of disability progression, and those without rapid disease who are ranked at a high MSSS at a given time due to fluctuation in

disability and would be expected to show disability improvement due to regression to the mean phenomenon. These considerations illustrate the danger of using MSSS as a stand-alone individual predictor. On the other end of the spectrum are patients with very low MSSS—especially those in the first decile—who have a very low risk of disability worsening or relapses (Figure 1). This group of patients with intrinsically slow disease may remain stable on less intense treatment. Reanalysis of datasets from randomized clinical trials by MSSS decile, especially those that included a placebo arm, would be informative to test this hypothesis as originally proposed Herbert.[29] Moreover, we advocate that future randomized clinical trials in MS should include prespecified analyses of drug efficacy stratified on baseline MSSS.

Our study is made possible by a very large, multinational patient dataset and a comprehensive predictive model that enables an analyst to account for the prognostic value of known clinical and demographic predictors. We are aware of the limitations of observational data (cataloged in Kalincik and Butzkueven[7], Kister et al.[30]) and used several strategies to mitigate some of the potential biases. Thus, we disregarded EDSS worsening and improvement events within 6 months from treatment start, and between baseline and the first postbaseline visit, to eliminate the effect of recent prebaseline relapses on disability and of therapeutic lag.[11] We addressed the lack of calibration by examining the frequency of events within predicted hazard strata. We did not, however, validate this version of the model on an independent cohort as was done with the original version.[6] Our results may also not be generalizable to untreated patients as the prediction model has been developed and tested on treated MS patients only. Finally, our model can presently be deployed for patients whose structured data has already been recorded in an MS registry, such as MSBase; it would be unrealistic to expect clinicians to input dozens of variables needed to calculate the outputs for the model. The next steps in the model development would be to reduce the number of input variables, which would eliminate extensive data entry and the need for dimensionality reduction. Furthermore, because the predictive accuracy of the existing model deteriorates with longer follow-up, future work should try to find ways to improve prediction over a longer time. The 4-year follow-up period in our study is longer than the typical 2- to 3-year duration of clinical trials in relapsing-remitting MS but is still short relative to the duration of a disease that spans decades.

In conclusion, we show that MSSS considerably improves the accuracy of individualized prognostics in MS. We advocate that disability rank scores be incorporated into future prognostic models of MS.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: T.K. served on scientific advisory boards for BMS, Roche, Sanofi Genzyme, Novartis, Merck and Biogen, steering committee for Brain Atrophy Initiative by Sanofi Genzyme, received conference travel support and/or speaker honoraria from WebMD Global, Novartis, Biogen, Sanofi-Genzyme, Teva, BioCSL, and Merck, and received research or educational event support from Biogen, Novartis, Genzyme, Roche, Celgene, and Merck. I.K. served on the scientific advisory board for Biogen Idec, Genentech, Alexion, and EMDSerono; received consulting fees from Roche; and received research support from Guthy-Jackson Charitable Foundation, National Multiple Sclerosis Society, Biogen Idec, Serono, Genzyme, and Genentech/Roche; he receives royalties from Wolters Kluwer for "Top 100 Diagnosis in Neurology" (co-written with Jose Biller). E.K.-H. has received honoraria/research support from Biogen, Merck Serono, Novartis, Roche, and Teva; has served as a member of advisory boards for Actelion, Biogen, Celgene, Merck Serono, Novartis, and Sanofi Genzyme; has been supported by the Czech Ministry of Education research, project PROGRES Q27/LF1.
L. K.'s institution (University Hospital Basel) has received the following exclusively for research support: steering committee, advisory board, and consultancy fees (Actelion, Bayer HealthCare, Biogen, BMS, Genzyme, Janssen, Japan Tobacco, Merck, Novartis, Roche, Sanofi, Santhera, Shionogi, and TG Therapeutics); speaker fees (Bayer HealthCare, Biogen, Merck, Novartis, Roche, and Sanofi); license fees for Neurostatus products; and grants (Bayer HealthCare, Biogen, European Union, InnoSwiss, Merck, Novartis, Roche, Swiss MS Society, and Swiss National Research Foundation). G.C.: Data and Safety Monitoring Boards: Astra-Zeneca, Avexis Pharmaceuticals, Biolinerx, Brainstorm Cell Therapeutics, Bristol Meyers Squibb/Celgene, CSL Behring, Galmed Pharmaceuticals, GreenValley Pharma Ltd, Mapi Pharmaceuticals LTD, Merck, Merck/Pfizer, Opko Biologics, OncoImmune, Neurim, Novartis, Ophazyme, Sanofi-Aventis, Reata Pharmaceuticals, Teva pharmaceuticals, VielaBio Inc, Vivus, NHLBI (Protocol Review Committee), NICHD (OPRU oversight committee). Consulting or Advisory Boards: Biodelivery Sciences International,

Biogen, Click Therapeutics, Genzyme, Genentech, GW Pharmaceuticals, Immunic, Klein-Buendel Incorporated, Medimmune/Viela Bio, Medday, Merck/Serono, Neurogenesis LTD, Novartis, Osmotica Pharmaceuticals, Perception Neurosciences, Recursion/Cerexis Pharmaceuticals, Regeneron, Rekover Pharmaceuticals, Roche, and TG Therapeutics. G.C. is employed by the University of Alabama at Birmingham and President of Pythagoras, Inc., a private consulting company located in Birmingham AL. A.L. has served as a Biogen, Merck Serono, Novartis, Roche, Sanofi/Genzyme, and Teva Advisory Board Member. She received congress and travel/accommodation expense compensations or speaker honoraria from Biogen, Merck, Mylan, Novartis, Sanofi/Genzyme, Teva, and Fondazione Italiana Sclerosi Multipla. Her institutions received research grants from Novartis. V.v.P. has received travel grants from Merck, Biogen, Sanofi, Bristol Meyer Squibb, Almirall, and Roche. His institution has received research grants and consultancy fees from Roche, Biogen, Sanofi, Merck, Bristol Meyer Squibb, Janssen Almirall, and Novartis Pharma. J.L.-S. has accepted travel compensation from Novartis, Biogen, and Merck. Her institution receives the honoraria for talks and advisory board commitment as well as research grants from Biogen Idec, Sanofi Genzyme, Merck, Roche, TEVA, and Novartis. H.B.'s research is funded by an Australian National Health Medical Research Council Investigator Grant. His institution receives funding from Biogen, Roche, Merck, Alexion, and Novartis for speaker engagements, study steering and advisory committee service. He is on the editorial board of Multiple Sclerosis and Related Disorders and the Steering committee of the Brain Health Initiative (Oxford Health Policy Forum). T.E.B., C.B.M., S.S., and O.G. have no disclosures.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by an unrestricted investigator-initiated grant from Genentech. MSBase would like to acknowledge financial contributions to support the MSBase Registry from Biogen, Novartis, Merck, Roche, and Sanofi Genzyme. Authors acknowledge funding from NHMRC (1140766, 1129189) to develop the original prediction model.

ORCID iDs

Tomas Kalincik https://orcid.org/0000-0003-3778-1376
Ilya Kister https://orcid.org/0000-0003-3549-949X
Dana Horakova https://orcid.org/0000-0003-1915-0036
Francesco Patti https://orcid.org/0000-0002-6923-0846
Sara Eichau https://orcid.org/0000-0001-9159-3128
Alessandra Lugaresi https://orcid.org/0000-0003-2902-5589
Raed Alroughani https://orcid.org/0000-0001-5436-5804
Ludwig Kappos https://orcid.org/0000-0003-4175-5509
Jeannette Lechner-Scott https://orcid.org/0000-0002-3850-447X
Vahid Shaygannejad https://orcid.org/0000-0002-6226-0161

Supplemental Material

Supplemental material for this article is available online.

References

1. Roxburgh RH, Seaman SR, Masterman T, et al. Multiple Sclerosis Severity Score: Using disability and disease duration to rate disease severity. *Neurology* 2005; 64(7): 1144–1151.
2. Kister I, Chamot E, Salter AR, et al. Disability in multiple sclerosis: A reference for patients and clinicians. *Neurology* 2013; 80: 1018–1024.
3. Daumer M, Neuhaus A, Herbert J, et al. Prognosis of the individual course of disease: The elements of time, heterogeneity and precision. *J Neurol Sci* 2009; 287(Suppl. 1): S50–S55.
4. Hughes S, Spelman T, Trojano M, et al. The Kurtzke EDSS rank stability increases 4 years after the onset of multiple sclerosis: Results from the MSBase Registry. *J Neurol Neurosurg Psychiatry* 2012; 83(3): 305–310.
5. Kister I, Bacon TE and Cutter GR. Short-term disability progression in two multiethnic multiple sclerosis centers in the treatment era. *Ther Adv Neurol Disord* 2018; 11: 1756286418793613.
6. Kalincik T, Manouchehrinia A, Sobisek L, et al. Towards personalized therapy for multiple sclerosis: Prediction of individual treatment response. *Brain* 2017; 140(9): 2426–2443.
7. Kalincik T and Butzkueven H. The MSBase registry: Informing clinical practice. *Mult Scler* 2019; 25(14): 1828–1834.
8. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann Neurol* 2011; 69(2): 292–302.

9. 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.
10. Lublin FD, Reingold SC, Cohen JA, et al. Defining the clinical course of multiple sclerosis: The 2013 revisions. *Neurology* 2014; 83(3): 278–286.
11. Roos I, Leray E, Frascoli F, et al. Delay from treatment start to full effect of immunotherapies for multiple sclerosis. *Brain* 2020; 143: 2742–2756.
12. Kister I, Bacon T, Sharmin S, et al. Multiple Sclerosis Severity Score (MSSS) helps predict relapses and recovery from disability in patients treated for multiple sclerosis in the MSBase model. In: *Presented at 35th Congress of the European Committee for Treatment and Research in Multiple Sclerosis (ECTRIMS)*, Stockholm, September 2019, https://onlinelibrary.ectrims-congress.eu/ectrims/2019/stockholm/278795/ilya.kister.multiple.sclerosis.severity.score.28msss29.helps.predict.relapses.html?f=listing%3D3%2Abrowseby%3D8%2Asortby%3D1%2Amedia%3D1
13. Sorensen PS, Sellebjerg F, Hartung HP, et al. The apparently milder course of multiple sclerosis: Changes in the diagnostic criteria, therapy and natural history. *Brain* 2020; 143(9): 2637–2652.
14. Ribbons KA, McElduff P, Boz C, et al. Male sex is independently associated with faster disability accumulation in relapse-onset MS but not in primary progressive MS. *PLoS ONE* 2015; 10(6): e0122686.
15. Marrie RA, Rudick R, Horwitz R, et al. Vascular comorbidity is associated with more rapid disability progression in multiple sclerosis. *Neurology* 2010; 74: 1041–1047.
16. McKay KA, Tremlett H, Fisk JD, et al. Psychiatric comorbidity is associated with disability progression in multiple sclerosis. *Neurology* 2018; 90(15): e1316–e1323.
17. Scalfari A, Neuhaus A, Daumer M, et al. Early relapses, onset of progression, and late outcome in multiple sclerosis. *JAMA Neurol* 2013; 70(2): 214–222.
18. Confavreux C, Vukusic S and Adeleine P. Early clinical predictors and progression of irreversible disability in multiple sclerosis: An amnesic process. *Brain* 2003; 126(Pt 4): 770–782.
19. Brex PA, Ciccarelli O, O'Riordan JI, et al. A longitudinal study of abnormalities on MRI and disability from multiple sclerosis. *N Engl J Med* 2002; 346: 158–164.
20. Minneboo A, Uitdehaag BM, Jongen P, et al. Association between MRI parameters and the MS severity scale: A 12 year follow-up study. *Mult Scler* 2009; 15(5): 632–637.
21. 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.
22. Le M, Malpas C, Sharmin S, et al. Disability outcomes of early cerebellar and brainstem symptoms in multiple sclerosis. *Mult Scler* 2021; 27(5): 755–766.
23. Malpas CB, Manouchehrinia A, Sharmin S, et al. Early clinical markers of aggressive multiple sclerosis. *Brain* 2020; 143(5): 1400–1413.
24. Sormani MP, Rio J, Tintorè M, et al. Scoring treatment response in patients with relapsing multiple sclerosis. *Mult Scler* 2013; 19(5): 605–612.
25. Brown FS, Glasmacher SA, Kearns PKA, et al. Systematic review of prediction models in relapsing remitting multiple sclerosis. *PLoS ONE* 2020; 15(5): e0233575.
26. GBD 2016 Multiple Sclerosis Collaborators. Global, regional, and national burden of multiple sclerosis 1990-2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol* 2019; 18(3): 269–285.
27. Harding K, Williams O, Willis M, et al. Clinical outcomes of escalation vs early intensive disease-modifying therapy in patients with multiple sclerosis. *JAMA Neurol* 2019; 76(5): 536–541.
28. Simonsen CS, Flemmen HØ, Broch L, et al. Early high efficacy treatment in multiple sclerosis is the best predictor of future disease activity over 1 and 2 years in a Norwegian Population-Based Registry. *Front Neurol* 2021; 12: 693017.
29. Charlson R, Herbert J and Kister I. Severity grading in multiple sclerosis: A proposal. *Int J MS Care* 2016; 18(5): 265–270.
30. Kister I, Chamot E, Cutter G, et al. Increasing age at disability milestones among MS patients in the MSBase Registry. *J Neurol Sci* 2012; 318(1–2): 94–99.