Brain Advance Access published September 10, 2015

BRAIN
A JOURNAL OF NEUROLOGY

Defining reliable disability outcomes in multiple sclerosis

Tomas Kalincik,[1,2] Gary Cutter,[3] Tim Spelman,[1] Vilija Jokubaitis,[1] Eva Havrdova,[4] Dana Horakova,[4] Maria Trojano,[5] Guillermo Izquierdo,[6] Marc Girard,[7] Pierre Duquette,[7] Alexandre Prat,[7] Alessandra Lugaresi,[8] Francois Grand'Maison,[9] Pierre Grammond,[10] Raymond Hupperts,[11] Celia Oreja-Guevara,[12] Cavit Boz,[13] Eugenio Pucci,[14] Roberto Bergamaschi,[15] Jeannette Lechner-Scott,[16] Raed Alroughani,[17] Vincent Van Pesch,[18] Gerardo Iuliano,[19] Ricardo Fernandez-Bolaños,[20] Cristina Ramo,[21] Murat Terzi,[22] Mark Slee,[23] Daniele Spitaleri,[24] Freek Verheul,[25] Edgardo Cristiano,[26] José Luis Sánchez-Menoyo,[27] Marcela Fiol,[28] Orla Gray,[29] Jose Antonio Cabrera-Gomez,[30] Michael Barnett[31] and Helmut Butzkueven[1,2,32]

Prevention of irreversible disability is currently the most important goal of disease modifying therapy for multiple sclerosis. The disability outcomes used in most clinical trials rely on progression of Expanded Disability Status Scale score confirmed over 3 or 6 months. However, sensitivity and stability of this metric has not been extensively evaluated. Using the global MSBase cohort study, we evaluated 48 criteria of disability progression, testing three definitions of baseline disability, two definitions of progression magnitude, two definitions of long-term irreversibility and four definitions of event confirmation period. The study outcomes comprised the rates of detected progression events per 10 years and the proportions of the recorded events persistent at later time points. To evaluate the ratio of progression frequency and stability for each criterion, we calculated the proportion of events persistent over the five subsequent years once progression was achieved. Finally, we evaluated the clinical and demographic determinants characterising progression events and, for those that regressed back to baseline, determinants of their subsequent regression. The study population consisted of 16 636 patients with the minimum of three recorded disability scores, totalling 112 584 patient-years. The progression rates varied between 0.41 and 1.14 events per 10 years, with the length of required confirmation interval as the most important determinant of the observed variance. The concordance among all tested progression criteria was only 17.3%. Regression of disability occurred in 11–34% of the progression events over the five subsequent years. The most important determinant of progression stability was the length of the confirmation period. For the most accurate set of the progression criteria, the proportions of 3-, 6-, 12- or 24-month confirmed events persistent over 5 years reached 70%, 74%, 80% and 89%, respectively. Regression post progression was more common in younger patients, relapsing-remitting disease course, and after a smaller change in disability, and was inflated by higher visit frequency. These results suggest that the disability outcomes based on 3–6-month confirmed disability progression overestimate the accumulation of permanent disability by up to 30%. This could lead to spurious results in short-term clinical trials, and the issue may be magnified further in cohorts consisting predominantly of younger patients and patients with relapsing-remitting disease. Extension of the required confirmation period increases the persistence of progression events.

1 Department of Medicine, University of Melbourne, Melbourne, Australia
2 Department of Neurology, Royal Melbourne Hospital, Melbourne, Australia
3 Department of Biostatistics, University of Alabama, Tuscaloosa, AL, USA
4 Department of Neurology and Center of Clinical Neuroscience, General University Hospital and Charles University in Prague, Czech Republic

Received February 4, 2015. Revised May 14, 2015. Accepted July 11, 2015.

5 Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari, Bari, Italy
6 Department of Neurology, Hospital Universitario Virgen Macarena, Sevilla, Spain
7 Department of Neurology, Hôpital Notre Dame, Montreal, Canada
8 MS Center, Department of Neuroscience, Imaging and Clinical Sciences, University 'G. d'Annunzio', Chieti, Italy
9 Neuro Rive-Sud, Hôpital Charles LeMoyne, Quebec, Canada
10 Department of Neurology, Hotel-Dieu de Levis, Quebec, Canada
11 Zuyderland Ziekenhus, Sittard-Geleen, The Netherlands
12 Multiple Sclerosis Unit, University Hospital San Carlos, Madrid, Spain
13 Karadeniz Technical University, Trabzon, Turkey
14 Neurology Unit, Azienda Sanitaria Unica Regionale Marche - AV3, Macerata, Italy
15 C. Mondino National Neurological Institute, Pavia, Italy
16 Hunter Medical Research Institute, University Newcastle, Newcastle, Australia
17 Amiri Hospital, Kuwait City, Kuwait
18 Cliniques Universitaires Saint-Luc, Brussels, Belgium
19 Ospedali Riuniti di Salerno, Salerno, Italy
20 Hospital Universitario Virgen de Valme, Seville, Spain
21 Hospital Germans Trias i Pujol, Badalona, Spain
22 Medical Faculty, Department of Neurology, Ondokuz Mayis University, Samsun, Turkey
23 Flinders University and Medical Centre, Adelaide, Australia
24 Azienda Ospedaliera di Rilievo Nazionale, San Giuseppe Moscati, Avellino, Italy
25 Groene Hart Ziekenhuis, Gouda, The Netherlands
26 Hospital Italiano, Buenos Aires, Argentina
27 Department of Neurology, Hospital de Galdakao-Usansolo, Galdakao, Spain
28 Fundacion para la Lucha contra las Enfermedades Neurologicas de la Infancia, Buenos Aires, Argentina
29 Craigavon Area Hospital, Portadown, UK
30 Centro Internacional de Restauracion Neurologica, Havana, Cuba
31 Brain and Mind Centre, Sydney, Australia
32 Department of Neurology, Box Hill Hospital, Monash University, Melbourne, Australia

Correspondence to: Tomas Kalincik,
L4 Centre,
Melbourne Brain Centre at Royal Melbourne Hospital,
Grattan St, Parkville VIC 3050, Australia
E-mail: tomas.kalincik@unimelb.edu.au

Keywords: disability; outcome measures; relapse; clinical trial; prognosis

Abbreviation: EDSS = Expanded Disability Status Scale

Introduction

Prevention of irreversible disability is currently the most important goal of multiple sclerosis disease modifying therapy. However, assessment of disability outcomes in multiple sclerosis therapeutic trials is a complex task in a disease with great individual and time-dependent variability of neurological disability and measurement error. In particular, the design of modern clinical trials with 1–3 year follow-up infers long-term irreversible disability outcomes from short-term confirmed progression events.

In relapsing-remitting multiple sclerosis, accumulation of irreversible disability is often obscured by transient neurological impairment due to relapses, with EDSS change persisting for three or more months but with subsequent regression to baseline (Hirst *et al.*, 2012). Therefore, delayed confirmation of newly acquired disability is imperative to distinguish true irreversible progression from relapse-associated reversible disability or measurement error (Pozzilli and Prosperini, 2008). To estimate the effect of disease modifying therapy on long-term irreversible disability, several definitions of short-term disability progression have been used, with the most common definition based on a one-step increase of the Expanded Disability Status Scale (EDSS) confirmed at least 3 months after onset (Filippini *et al.*, 2013).

Inherent in the definition of confirmed disability progression is the assumption of long-term persistence, i.e. of irreversible disability. However, an observation from pooled data from the placebo-treated cohorts of several pivotal randomized clinical trials suggested that a 3- or 6-month confirmed EDSS increase may not provide an accurate or stable estimate of long-term disease outcomes (Ebers *et al.*, 2008). In addition, other aspects of the progression definition could determine persistence of the identified disability accrual. These include the magnitude of the EDSS change, as well as and the definition of 'baseline'. For the latter, a single baseline EDSS assessment at the start of the observation period is often used, or, alternatively, confirmation of the baseline EDSS may be required, in order to mitigate against measurement error and EDSS fluctuation.

While psychometric properties of the EDSS, including its validity, reliability, sensitivity to change, distribution properties, feasibility, interpretation, and comparability have been evaluated, no comprehensive evaluation of long-term stability of various definitions of short-term EDSS progression has previously been performed (Meyer-Moock *et al.*, 2014).

We therefore used MSBase, a large international prospective observational multiple sclerosis cohort study with relatively short intervals between reported EDSS scores, to compare a total of 48 combined criteria of disability progression. In particular, the analysis evaluated long-term persistence (of at least 5 years) of the identified EDSS progression events.

Patients and methods

Ethics statement

The MSBase registry (Butzkueven *et al.*, 2006) (registered with WHO ICTRP, ID ACTRN12605000455662) was approved by the Melbourne Health Human Research Ethics Committee, and by the local ethics committees in all participating centres (or exemptions granted, according to applicable local laws and regulations). If required, written informed consent was obtained from enrolled patients, in accordance with the Declaration of Helsinki.

Patients and follow-up

Longitudinal clinical data from 25 266 patients from 66 multiple sclerosis centres in 26 countries were extracted from the MSBase registry in December 2013. The inclusion criteria consisted of the diagnosis of multiple sclerosis or clinically isolated syndrome based on the 2005 or 2010 revised McDonald criteria (Polman *et al.*, 2005, 2011) and availability of the minimum dataset (i.e. patient sex, year of birth, year of the first clinical presentation, multiple sclerosis course, treating centre and at least three clinical visits with recorded EDSS scores). The data quality assessment was conducted using a series of procedures to identify any invalid or inconsistent entries, as described elsewhere (Kalincik *et al.*, 2013*a*); only information from centres contributing at least ten active patient records was included, and a date of onset was required for all recorded events.

The analysed data were recorded as part of quality clinical practice, mostly at large tertiary multiple sclerosis centres. The usual data entry practice was real-time or near-real time data entry (at the time of clinical visits). The MSBase protocol stipulates minimum annual updates of the minimum data set, but patients with less frequent visits were not excluded from the analysis. Data entry portal was either the iMed patient record system or the MSBase online data entry system. The on-study follow-up was defined as the time between the first and the last available EDSS entry.

Disability was scored by accredited scorers (Neurostatus certification was required at each centre) using the EDSS, calculated based on the functional system scores. While EDSS scores at any time-points regardless of their relationship to relapses could serve as evidence of EDSS progression, only EDSS scores recorded more than 30 days from the onset of a preceding relapse were used to confirm these progression events (3.2% of the confirmatory EDSS scores were recorded more than 30 but less than 61 days of a relapse). A relapse was defined as occurrence of new symptoms or exacerbation of existing symptoms persisting for at least 24 h, in the absence of concurrent illness or fever, and occurring at least 30 days after a previous relapse (Schumacher *et al.*, 1965). Formal quantification of relapse-associated disability change is not required as part of the MSBase observational protocol.

Relapsing-remitting disease course was defined as multiple sclerosis presenting with bout onset followed by relapses. Primary progressive multiple sclerosis was defined as the disease with at least one year of disease progression from its first clinical manifestation. Secondary progressive disease was identified by treating neurologists based on continuous progression of disability following the relapse-onset disease course. The progressive disease course comprised both active and not active phenotypes defined by Lublin *et al.* (2014).

Definitions of disability progression

Forty-eight definitions of progression events were generated as combinations of the following criteria:

(i) Baseline EDSS. Given that any single EDSS score may be burdened by measurement error (Goodkin *et al.*, 1992), we examined several definitions of baseline EDSS step:
   (a) EDSS at the first recorded visit (i.e. the typical trial definition), or;
   (b) the minimum EDSS confirmed at two or more consecutive visits separated by at least 3 months, prior to an identified progression event, or;
   (c) the lower of either criterion (a) or (b).

(ii) Based on the suggested use of half-step progression above EDSS step 5.5 (Weinshenker *et al.*, 1996), we evaluated the magnitude of EDSS change as:
   (a) 2 strata: Increase in EDSS by 1 point if baseline EDSS was 5.5 or lower, or increase in EDSS by 0.5 point if baseline EDSS was above 5.5, or;
   (b) 3 strata: Increase in EDSS by 1.5 points if baseline EDSS was 0, increase in EDSS by 1 point if baseline EDSS was between 1 and 5.5, or increase in EDSS by 0.5 points if baseline EDSS was above 5.5.

(iii) Persistence of the EDSS progression for entire follow-up duration:
   (a) EDSS progression events where all the subsequent EDSS scores remained at or above the level defined in (ii), or;
   (b) EDSS progression events regardless of the subsequent EDSS scores.

(iv) Confirmation of EDSS progression at two or more consecutive visits separated in time by the minimum of:
   (a) 3 months;
   (b) 6 months;
   (c) 12 months;
   (d) 24 months.

The progression events were confirmed using all EDSS scores recorded during the minimum confirmation period and the first

EDSS score recorded after the minimum confirmation period. Multiple progression events were allowed per patient. Following each identified progression event, baseline disability level was readjusted using the criteria included in the relevant definition of progression (see above) to eliminate detection of prolonged fluctuation in EDSS.

Study outcomes and statistical analysis

Statistical analyses were carried out using R, version 3.0.3 (R Development Core Team, 2011). The point and interval estimates of data distributions were expressed as mean with 95% confidence intervals or margins of error, or median with interquartile range, as appropriate.

The incidence of progression events is assumed to follow a Poisson distribution and was calculated as the number of events identified per patient-decade of follow-up. In addition, the maximum hypothetical progression incidence was estimated by including those progression events not fulfilling the predefined criteria due to insufficient follow-up (i.e. due to censoring before the criteria could be fulfilled).

To evaluate the persistence of progression events over time, time to confirmed disability regression was assessed for each event. The proportions of events identified by the evaluated criteria which were followed by a 3-month confirmed regression of disability were visualized with Kaplan-Meier curves. The relationship between the sensitivity of the criteria (i.e. the standardized 10-year event incidence) and the persistence of the identified events (i.e. the proportion of the events without 3-month confirmed disability regression at 5 years post-event) was evaluated in the proportion of progression events for which at least a 5-year clinical follow-up was available.

Independent associations between selected demographic and clinical patient characteristics and the probability of experiencing progression events or reaching predefined disability levels were examined using a series of multivariable proportional hazards models (Andersen-Gill models with cluster term for patient and Efron approximation method for handling ties). The tested variables included sex, age, disease duration (from first clinical symptoms), baseline EDSS, disease course (relapsing-remitting, clinically isolated syndrome, secondary progressive, primary progressive), follow-up duration (between the first and the last available EDSS visit), and annualized visit density. Marginal Cox proportional hazards models were used to evaluate independent determinants of 3-month confirmed regression of disability following the progression events. The relative change in EDSS at the time of progression and the confirmed post-progression EDSS were used in the models of disability regression instead of the baseline EDSS.

Results

Patients

Of the 25 266 patients enrolled in the MSBase registry, 25 140 patients were diagnosed with multiple sclerosis or clinically isolated syndrome; 25 101 patients had complete minimum datasets without identified errors; and 16 636 patients had at least three recorded visits with EDSS scores and were included in the analysis. The median time between EDSS visits was 6.6 months (interquartile range 4.3–10.1). The majority of the included patients were enrolled in the MSBase registry in 2000 or later (83.5%). The number of included patients per centre is shown in the Supplementary Table 1 and their demographic and clinical characteristics are provided in Table 1 and Supplementary Fig. 1. The cumulative captured follow-up was 112 584 patient-years, with the median per patient follow-up of 5.7 years and nine visits with recorded EDSS scores.

Incidence of progression events

The mean incidence of progression events (for an example of a typical disability course see Supplementary Fig. 2) varied, with respect to the examined definitions, between 0.41 and 1.14 events per patient-decade (margin of error between 0.003 and 0.006; see Fig. 1; for the number of the identified progression events see Supplementary Table 2). Maximum hypothetical progression rates calculated by including the progression events persistent for the duration of the follow-up but with insufficient time to fulfil the predefined criteria of event confirmation are shown in Fig. 1. These represent the upper bounds of the progression incidence. The most pronounced determinant of the progression event incidence was the duration of the required confirmation interval, with 3-month confirmed events being the most common. The 2-strata definition of EDSS progression magnitude resulted in a marginally higher number of identified

Table 1 Characteristics of the study population

	Count (%)	Mean (SD)	Median (quartiles)
Patients (females)	16 636 (70)		
Age at inclusion		37.8 (11.3)	36.9 (29.3, 45.5)
Disease duration at inclusion		6.6 (7.7)	3.7 (1.0, 9.7)
Follow-up duration, years		6.8 (4.9)	5.7 (3.0, 9.6)
Disease course			
At inclusion			
Clinically isolated syndrome	3583 (22)		
Relapsing-remitting	10642 (64)		
Secondary progressive	1261 (8)		
Primary progressive	1150 (7)		
At censoring			
Clinically isolated syndrome	1462 (9)		
Relapsing-remitting	11574 (72)		
Secondary progressive	2450 (15)		
Primary progressive	1150 (7)		
Disability, EDSS step			
At inclusion		2.5 (1.9)	2 (1, 3.5)
At censoring		3.3 (2.4)	2.5 (1.5, 5.5)
Annualized change		+0.10 (0.005)	+0.07 (0, +0.25)
On-study annualized relapse rate		0.28 (0.40)	0.15 (0, 0.44)
Number of on-study visits		11.9 (9.9)	9 (5, 15)

SD = standard deviation.

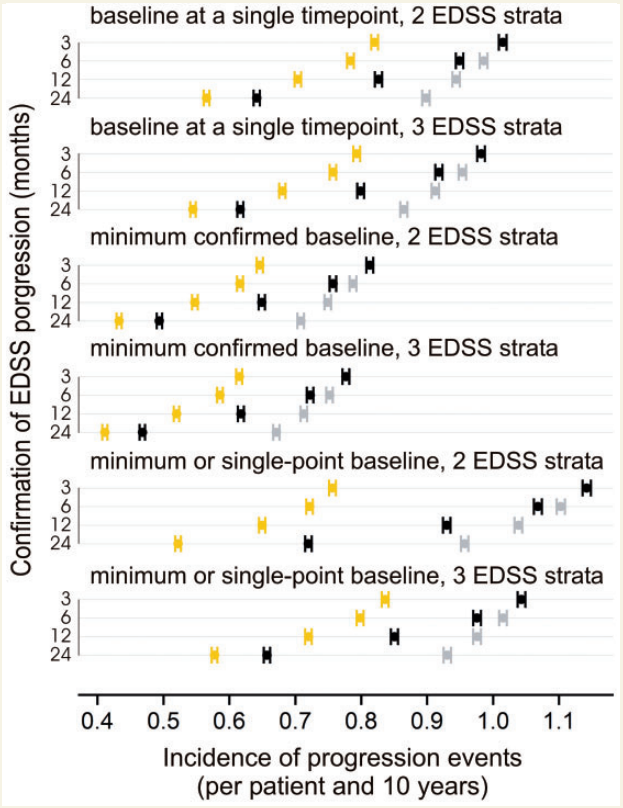

Figure 1 Incidence of disability progression events by the evaluated criteria. The graphs show the observed mean incidence (black), the maximum hypothetical incidence given that sufficient follow-up time to enable confirmation of each event was available (grey) and the incidence of the progression events which were persistent at all subsequent time points (yellow). Mean and 95% CI (Poisson) are shown. The criteria were evaluated in a cohort consisting of 16 636 patients followed over 112 584 patient-years. Two EDSS progression strata were defined as increase in EDSS by 1 step if baseline EDSS was 5.5 or lower, or increase in EDSS by 0.5 otherwise. Three EDSS progression strata were defined as the increase of EDSS by 1.5 steps if baseline EDSS was 0, by 0.5 steps if baseline EDSS was above 5.5, and by 1 step otherwise.

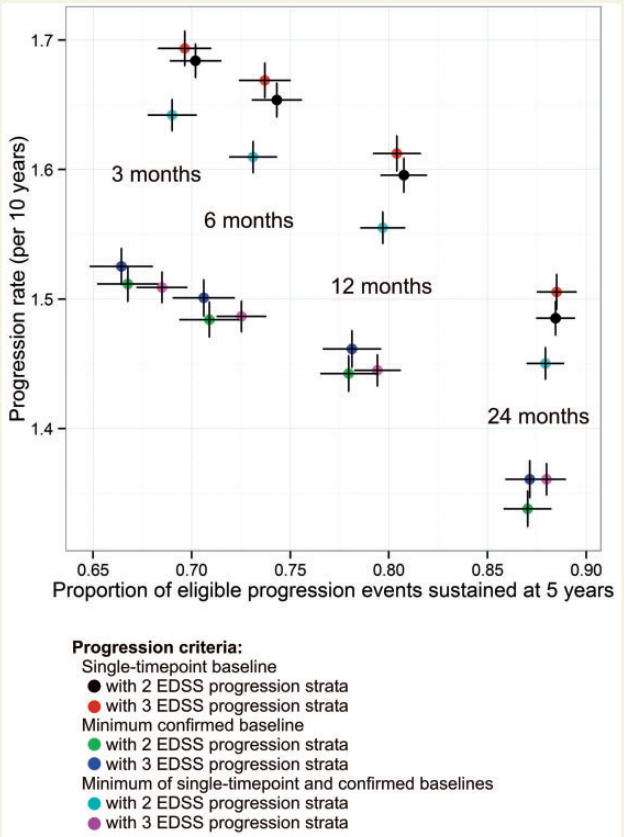

Figure 2 Estimate of the sensitivity and persistence of the evaluated progression criteria. The number of eligible patients was determined as the number of patients with at least 5-year follow-up following their first identified progression event. The number and characteristics of the patients evaluated for each criterion is given in Supplementary Table 3. Two EDSS progression strata were defined as increase in EDSS by 1 step if baseline EDSS was 5.5 or lower, or increase in EDSS by 0.5 otherwise. Three EDSS progression strata were defined as the increase of EDSS by 1.5 steps if baseline EDSS was 0, by 0.5 steps if baseline EDSS above 5.5, and by 1 step otherwise. Error bars represent 95% CI.

progression events than the 3-strata definition. As expected, the requirement of progression being sustained at all subsequent EDSS visits led to a relative decrease in the incidence of progression events. Among the definitions of baseline disability, the definition using the lower of the two potential baseline EDSS scores {either EDSS at the first visit or the 3 months confirmed lowest EDSS [definition (i)c, see 'Materials and methods' section]} predictably identified the highest number of progression events.

Of the 14 129 unique progression events identified by at least one of the tested progression criteria, only 2656 (17.3%) were identified by all 48 criteria simultaneously.

Persistence of progression events

To evaluate the relationship between the incidence of progression events identified by the tested criteria and their persistence over time, we identified those events with at least 5-year available post-event follow-up (for the characteristics of these events see Supplementary Table 3). Figure 2 shows the progression rates per patient-decade and the proportion of these events which continued to fulfil the criteria of progression over the subsequent 5 years (the criteria where persistent progression of disability formed part of the definition are not shown). The two criteria resulting in the combination of the highest disability progression rate and the largest proportion of persistent events at 5 years were those defined by baseline EDSS recorded at a single time-point, and 12- or 24-month confirmed progression of disability. The respective proportions of 5-year persistent progression events were 80–81% [95% confidence interval (CI) 79–82%) and 88–89% (95% CI 87–90%). The criteria using the two and the three EDSS strata resulted in comparable progression persistence. In contrast, the progression criteria typically used in clinical trials, based on 3- or 6-month confirmed progression

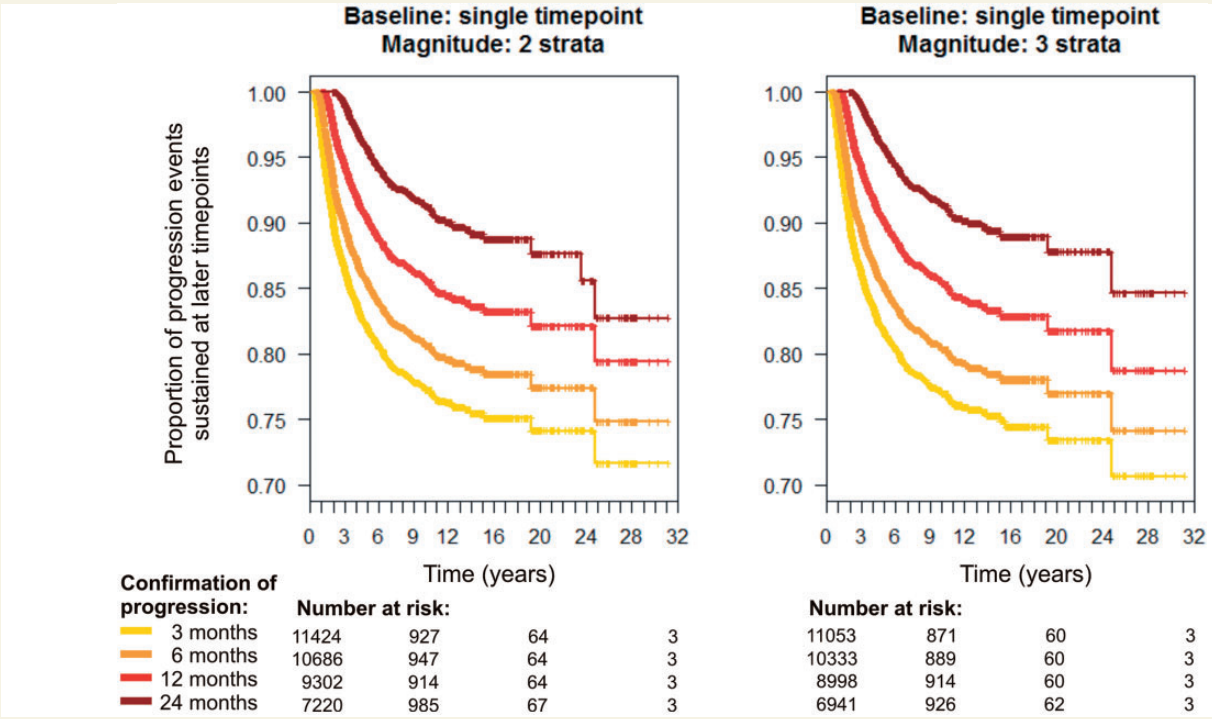

Figure 3 Proportion of persistent progression events. The proportions of progression events for the two most efficient progression definitions and stratified by confirmation period (in colour) sustained over time post-event are shown.

disability and three EDSS strata resulted in 70% (95% CI 68–71%) and 74% (95% CI 72–75%) progression persistence at 5 years, respectively.

To assess persistence of the identified disability progression events over time, we evaluated the proportion of these events in which the original criteria of progression failed to be sustained on two consecutive subsequent visits, separated in time by at least 3 months (i.e. 3-month confirmed regression of disability). Figure 3 demonstrates the Kaplan-Meier curves for the proportions of regressed progression events (with time at reaching progression set as year 0) for the two most efficient criteria sets (see above). With respect to the various definitions of progression, the most powerful determinant of regression probability post-progression was duration of the required confirmation period. The 3-months confirmed progression events were the least persistent, with 22–26% regression rates over the initial 10 years post-progression. In contrast, the 24-month confirmed progression events were the most persistent of the compared criteria, with only 8–9% regression rates over the 10 years post-progression.

Long-term disability outcomes associated with progression and regression events

Among the patients with sufficient post-progression follow-up, we evaluated the EDSS at 5 years from their first progression event. Those in whom the progression events were followed by regression of disability (between 299 and 942 patients, depending on the criterion) showed lower disability at 5 years than those with the progression event persistent over the 5 years (group sizes of between 2036 and 2386 patients). The corresponding EDSS scores were 2–2.5 (1.5, 4–4.5) versus 5.5–6 (3.5–4, 6.5), respectively [median (quartiles)] for the two most efficient disability progression criteria. These values correspond to the respective increase in EDSS by 0.5–1 (0–0.5, 1–1.75) versus 2.5 (1.5–3.5), [median (quartiles)].

Determinants of disability progression and regression

Hazard ratios for the potential determinants of progression events derived from a series of multivariable proportional hazards models are shown in Fig. 4. While male sex was associated with an increased risk of experiencing progression events for all tested criteria, female sex was associated with a more likely recovery of the events confirmed at 3 months. Older age was associated with a higher risk of progression events with a decreased likelihood of recovery. The association with age was stronger for the criteria with longer confirmation interval. Interestingly, disease duration was not independently associated with the probability of progression or regression. Lower baseline EDSS was associated with a higher probability of progression events, and the association varied depending on the definition of progression magnitude (with the

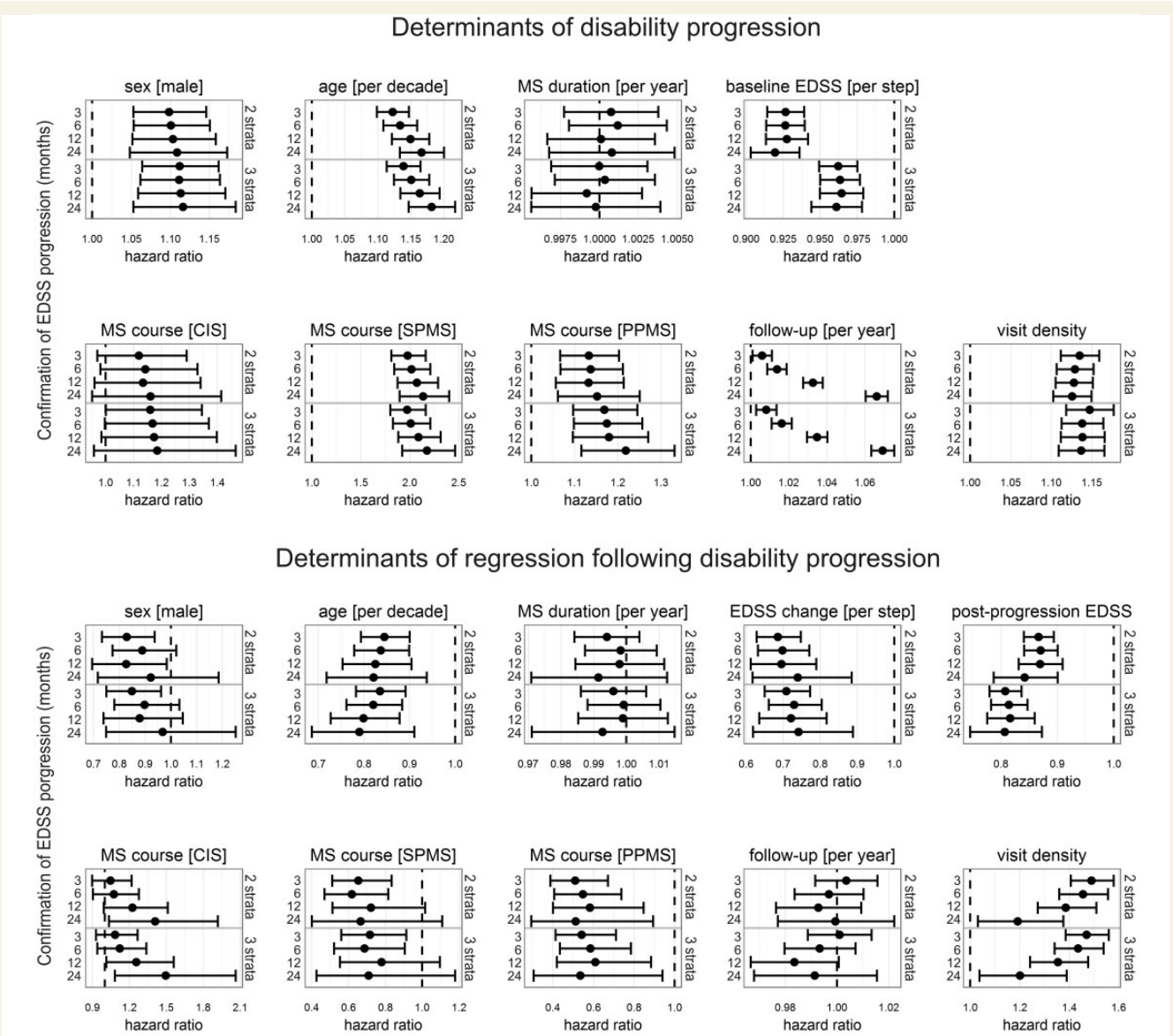

Figure 4 Determinants of disability progression events and their subsequent regression. The graphs represent hazard ratios and their 95% CIs from a series of multivariable Anderson-Gill models of the cumulative hazard of multiple progression events and their subsequent regression (for the numbers of progression events detected by each of the criteria see Supplementary Table 2). Each row represents a single multivariable model of one progression definitions, with the evaluated determinants shown in separate panels to facilitate assessment of the identified associations with respect to the definitions. Two EDSS progression strata were defined as increase in EDSS by 1 step if baseline EDSS was 5.5 or lower, or increase in EDSS by 0.5 otherwise. Three EDSS progression strata were defined as the increase of EDSS by 1.5 steps if baseline EDSS was 0, by 0.5 steps if baseline EDSS above 5.5, and by 1 step otherwise. Baseline EDSS was calculated as per the criteria included in the definitions. Reference value for the disease course was relapsing-remitting multiple sclerosis. Visit density was calculated as the mean number of visits per year. CIS = clinically isolated syndrome; PPMS = primary progressive multiple sclerosis; SPMS = secondary progressive multiple sclerosis.

stronger association observed for the 2-strata paradigm). Both greater progression-related change in EDSS and higher post-progression EDSS score were associated with a decreased likelihood of disability regression. Progressive disease course, in particular the secondary progressive course, was positively associated with the incidence of progression events, and in most instances was negatively associated with a subsequent recovery. As expected, the association with follow-up duration was more pronounced for the progression events with later confirmation but not for the regression events. Importantly, the frequency of EDSS assessments was positively associated with the incidence of progression events and their regression for all evaluated criteria. The association between the EDSS frequency and the probability of regression was mitigated by the longer confirmation of the progression events.

Discussion

In this analysis of multiple sclerosis disability change from the prospective, observational MSBase cohort study, we have shown that the disability metrics based on short-term confirmed disability progression overestimate the long-term accumulation of irreversible disability. This bias can be mitigated by extending the minimum confirmation time from the 3–6 months, used in most of the previous intervention trials, to 12 or 24 months, with only little effect on the sensitivity of the progression criteria.

Optimizing the definition of disability progression

The definitions of disability outcomes used most commonly in the clinical trials of multiple sclerosis therapies are those of 1-step EDSS progression confirmed over 3- or 6-month period (Filippini *et al.*, 2013). These metrics are used as estimators of irreversible, long-term accumulation of disability available within the limited timeframe of the treatment trials. However, information from the placebo arms of 31 randomized clinical trials suggested that only a small proportion of the 3- and 6-month confirmed progression events reflect permanent disability, especially in relapsing-remitting multiple sclerosis (Ebers *et al.*, 2008). For instance, data from the pivotal trial of interferon β-1b showed that the on-trial 1-step progression of EDSS confirmed at 3 months was only poorly predictive of disability outcomes at 16 years (EDSS step 6 or secondary progressive multiple sclerosis) (Goodin *et al.*, 2012). Liu and Blumhardt (2000) demonstrated that even as early as the end of a clinical trial conducted over 2 years, half of those patients who experienced 3- or 6-month confirmed disability progression have already reverted to a non-progressed status.

Healy *et al.* (2011) showed that modelling EDSS scores directly may provide higher power to detect relative treatment effects than modelling confirmed progression events. However, the absolute change in EDSS (particularly when relying on a small number of compared time-points) is burdened by noise introduced by the relatively high inter- and intra-rater variability (Noseworthy *et al.*, 1990) as well as variance introduced by reversible deterioration of neurological signs due to relapse (Lublin *et al.*, 2003), and its use in intervention trials has been discouraged (European Medicines Agency, 2012). Instead, the European Medicines Agency recommended the following criteria: (i) sustained progression of disability based on 1-step EDSS progression (for EDSS $\leqslant$ 5.5) or 0.5-step EDSS progression (for EDSS $>$ 5.5) confirmed at two consecutive examinations at least 6 months apart, or (ii) accumulation of a specified degree of disability.

We have demonstrated that the most important determinant of progression event stability (defined as the lack of confirmed regression following a progression event) was the minimum required confirmation time. For the most stable set of progression criteria (based on a single baseline EDSS time-point and requiring 1.5-point progression where baseline EDSS step was 0), the proportion of events regressing within 5 years of progression decreased with longer confirmation time (30%, 26%, 20%, and 11% for 3-, 6-, 12-, and 24-month confirmation periods, respectively). This is in agreement with a previous study which reported that incidence of 1–2-step EDSS change confirmed at 12 months is a more reliable disability outcome than the outcomes routinely used in clinical trials (Ebers *et al.*, 2008).

As expected, the 2-strata paradigm of progression magnitude (with the EDSS increase of 1 point for EDSS 0-5.5, or 0.5 points above EDSS 5.5) resulted in marginally higher detection of progression events than the 3-strata paradigm. However, it should be noted that the stability of EDSS improves at the higher levels of disability (Weinshenker *et al.*, 1996; Hohol *et al.*, 1999; Ravnborg *et al.*, 2005) and thus the requirement of a relatively larger step progression in patients with milder disability may be necessary to improve the accuracy of the definition of progression events.

It is worth noting that only 17.3% of the all detected events were identified by all the tested criteria simultaneously. This implies that various aspects of the progression criteria impact markedly on the sensitivity of these criteria, in particular, the various definitions of event confirmation and baseline disability. Out of the baseline EDSS definitions, the definition using the lower of the single EDSS value recorded at the first visit or the 3-month confirmed minimum EDSS resulted in the highest detection of progression events. However, when examined in the subset of events with the subsequent minimum 5-year follow-up, it was less sensitive than the definition based on a single time point, which also yielded a marginally higher stability of the identified events (Fig. 2). This, together with the practicality of obtaining the baseline EDSS during a single visit, favours this definition above others.

The significance of accurate identification of the progression events with long-term persistence was demonstrated by the evaluation of the 5-year post-progression disability outcomes. Those progression events followed by regression of disability were associated only with a minor change in EDSS at 5 years, unlike the persistent progression events, which resulted in marked accumulation of disability and higher overall EDSS scores. Therefore, accurate identification of persistent progression events enables more accurate evaluation of patients' disability trajectories.

Determinants of progression and regression of disability

We confirmed the associations of male sex, older age and progressive multiple sclerosis course with higher probability of disability accrual reported by previous studies (Confavreux *et al.*, 2003; Leray *et al.*, 2010). We also observed an increased probability of progression events in

patients with lower EDSS score, an effect most likely attributed to the EDSS structure. While the lower EDSS steps are based on mild to moderate changes in multiple functional systems, higher EDSS steps are determined by quantitated locomotor performance and self-care functions and therefore possess improved stability. As expected, longer follow-up duration was required to optimize detection of the events with longer confirmation period. Importantly, all disability events were more incident to the patients with higher frequency of EDSS visits. Thus the follow-up duration and visit frequency represent potential confounders of disability outcomes and need to be taken into consideration in the design of observational studies as well as clinical trials.

As shown previously, improvement of disability is a well-known scenario in multiple sclerosis, either in relation to remission after a relapse or due to a more prolonged recovery of neurological function (Tremlett *et al.*, 2012). Our observed associations of younger age, non-progressive disease course, lower post-progression EDSS and lower progression-related disability accrual with increased probability of recovery from disability progression events were in keeping with the results of a study conducted in an untreated multiple sclerosis cohort (Tremlett *et al.*, 2012). The association of female sex with relatively higher chance of recovery from the progression events confirmed over 3 months most probably signifies that a number of these events represent relapses, which are known to be more frequent among females (Kalincik *et al.*, 2013*b*). The notion that the incidence of progression events and their regression was more closely associated with patient age than with disease duration is complementary to our observation of a similar interaction between age, disease duration and relapse frequency (Kalincik *et al.*, 2013*b*). The overall follow-up duration had no impact on the probability of recovery, whose definition requiring confirmation over 3 months was constant across various progression definitions. Similar to the incidence of progression events, the recovery was vulnerable to the confounding owed to the variable visit density—particularly the recovery from the events confirmed over a relatively short time.

Study limitations

The main limitation of the present study overlaps with the limitations of the EDSS. While EDSS is based on neurological examination and is therefore clinically relevant and accessible to neurologists, it is burdened with a relatively low intra- and inter-rater reliability contributed to by the subjective components of clinical assessment, particularly at the lower end of the scale (Amato *et al.*, 1988; Goodkin *et al.*, 1992). The scale is asymmetrical, assigning a relatively larger weight to locomotion, and non-linear (for review see Amato and Portaccio, 2007). Evaluation of the contribution of the functional system scores to the overall EDSS sensitivity and accuracy was beyond the scope of this project. Our study involved data recorded in 63 centres over long follow-up periods, and this probably led to inflation of EDSS score variance. On the other hand, we aimed at mitigating the EDSS variance through the requirement of Neurostatus certification at each participating MSBase centre and at diminishing its impact through the size of the studied population. Importantly, the central aim of this study was to assess the accuracy (i.e. long-term persistence) of disability outcomes assessed in quality clinical practice; therefore, the higher variance most probably resulted in more conservative conclusions in relation to the stability of the disability outcomes. While relapses represent important source of variability in disability, they often lead to accumulation of permanent disability (Lublin *et al.*, 2003). To reflect this, we allowed the initial change in disability to be recorded regardless of its relation to preceding relapses but required confirmation of this initial event with an EDSS score recorded outside a post-relapse period. It should be noted that the analysed cohort comprised patients treated with a variety of disease modifying therapies. While it is expected that various disease modifying agents exert differential effects on the incidence of EDSS progression events (in particular the events associated with relapses), analysis of this potential effect was not the aim of this study and will be addressed in future studies. A further limitation in relation to extrapolation of our conclusions to randomized clinical trials is that these are often restricted to individuals with lower EDSS scores and have frequent observations (e.g. every 3 months). As we have demonstrated that the latter would result in lower long-term persistence of identified sustained progression events, our evaluations of long-term progression persistence in a cohort with median inter-visit interval of 6.6 months could be optimistic. Finally, the lack of a minimum required follow-up time may have introduced a bias that would underestimate sensitivity of the definitions, in particular those using longer confirmation periods. The maximum magnitude of this bias can be estimated from the maximum hypothetical event incidence (shown in Fig. 1), whose trends mirror the trends observed in the incidence of the confirmed events.

Conclusion

Progression of EDSS score confirmed at 3 or 6 months is an outcome feasible for use in 2–3-year intervention trials; however, it may result in identification of temporary disability changes in 30% or 26% of the identified events, respectively. While 12- or 24-month confirmed disability progression is not free from this bias, it provides more accurate evaluation of irreversible disability accrual with 20% and 11% of the detected events owed to temporary EDSS changes, respectively. We therefore suggest implementation of longer disability confirmation periods in the design of observational studies but also of prospective clinical trials. This is not impractical as most modern trials include open-label extension studies, and these observations can be used

to define 12- and 24-month confirmations of disability progression events which occurred during the randomised stages of these trials.

Acknowledgements

From Hospital S. Joao, Porto, Portugal, Dr Maria Edite Rio; from University of Florence, Florence, Italy, Dr Maria Pia Amato; from Aarhus University Hospital, Aarhus C, Denmark, Dr Thor Petersen; from Liverpool Hospital, Liverpool, Australia, Dr Suzanne Hodgkinson; from Assaf Harofeh Medical Center, Beer-Yaakov, Israel, Dr Shlomo Flechter; Sfrom the MS-Centrum Nijmegen, Nijmegen, The Netherlands Dr Cees Zwanikken; from Francicus Ziekenhuis, Roosendaal, The Netherlands, Ms Leontien den Braber-Moerland; from Centro Internacional de Restauracion Neurologica, Havana, Cuba. Dr Jose Antonio Cabrera-Gomez; from INEBA, Buenos Aires, Argentina, Dr Maria Laura Saladino; from Hospital Fernandez, Buenos Aires, Argentina, Dr Norma Deri; from the Westmead Hospital, Sydney, Australia, Dr Steve Vucic; from the Geelong Hospital, Geelong, Australia, Dr Cameron Shaw; from Jewish General Hospital, Montreal, Canada, Dr Fraser Moore; from Jeroen Bosch Ziekenhuis, Den Bosch, The Netherlands, Dr Erik van Munster; from Buenos Aires, Argentina, Dr Aldo Savino; from Veszprem Megyei Csolnoky Ferenc Korhaz, Veszprem, Hungary, Dr Imre Piroska; from Royal Brisbane Hospital, Brisbane, Australia, Dr Pam McCombe from the Bombay Hospital Institute of Medical Sciences, Mumbai, India, Dr Bhim Singhal; from Mater Dei Hospital, Malta; Dr Norbert Vella; from Hopital Tenon, Paris, France, Dr Etienne Roullet; from the Neurology Clinical Center, Skopje, Macedonia, Dr Tatjana Petkovska-Boskova; from St Vincent's Hospital, Melbourne, Australia; from the Hospital Ecoville, Curibita, Brazil, Dr Walter Oleschko Arruda; from New York University Langone Medical Center, New York, USA, Dr Joseph Herbert and Dr Iliya Kister; from Isfahan University of Medical Sciences, Isfahan, Iran, Dr Vahid Shaygannejad from Central Clinical Emergency Military Hospital, Bucharest, Romania, Dr Carmen-Adella Sirbu; from Semmelweis University, Budapest, Hungary, Dr Magdolna Simo and Dr Anna Iljicsov; from The Alfred Hospital and Monash University, Melbourne, Australia, Dr Olga Skibina; from the Instituto de Neurociencias, Cordoba, Argentina, Dr Elizabeth Alejandra Bacile; from University of Debrecen, Debrecen, Hungary Dr Tunde Csepany; from HIGA Gral. San Martin, La Plata, Argentina, Dr Santiago Vetere; from Royal Victoria Hospital, Belfast, UK, Dr Gavin McDonnell; from Royal Hobart Hospital, Hobart, Australia, Dr Bruce Taylor; from Hospital Angeles de Las Lomas, Mexico City, Mexico, Dr Eli Skromne; from Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari, Italy, Dr Damiano Paolicelli, Dr Pietro Iaffaldano, Dr Vita Direnzo and Dr Mariangela D'Onghia; from FLENI, Buenos Aires, Argentina, Dr Jorge Correale and Dr Celica Ysrraelit; from the Hopital Notre-Dame, Canada, Dr Alexandre Prat, Ms Elaine Roger and Mr Pierre Despault; from the Royal Melbourne Hospital, Australia, Dr Mark Marriott, Dr Anneke Van der Walt, Dr John King and Dr Trevor Kilpatrick; from Box Hill Hospital, Melbourne, Australia, Ms Jodi Haartsen; from University 'G. d'Annunzio', Italy, Dr Giovanna De Luca, Dr Valeria Di Tommaso, Dr Daniela Travaglini, Dr Erika Pietrolongo, Dr Maria di Ioia, Dr Deborah Farina and Dr Luca Mancinelli; from Hospital Italiano, Argentina, Dr Juan Ignacio Rojas and Dr Liliana Patrucco; from Ospedale di Macerata, Italy, Dr Elisabetta Cartechini and Dr Giorgio Giuliani; from John Hunter Hospital, Australia, Dr David Williams and Dr Lisa Dark; from Craigavon Area Hospital, Portadown, UK, Dr Stella Hughes; from the MSBase Operations Team, Dr Jill Byron, Ms Lisa Morgan, Ms Eloise Hinson and Mr James Milesi; and from Rodanotech, Geneva, Switzerland: Mr Samir Mechati, Mr Matthieu Corageoud, Mr Alexandre Bulla.

Funding

The work was supported by the NHMRC Early Career Fellowship (1071124), NHMRC Practitioner Fellowship (1080518), NHMRC Project Grants (1083539 and 1032484), NHMRC Centre for Research Excellence (Grant ID 1001216) and the MSBase Foundation. The MSBase Foundation is a not-for-profit organization that receives support from Merck Serono, Biogen, Novartis Pharma, Bayer-Schering, Sanofi-Aventis and BioCSL. The study was conducted separately and apart from the guidance of the sponsors.

Disclosure statement

A.L. served as a Bayer Pharma, Biogen Idec, Genzyme, Merck Serono Advisory Board Member. She received travel grants and honoraria from Bayer Pharma, Biogen Idec, Merck Serono, Novartis, Sanofi and Teva, research grants from BayerPharma, Biogen Idec, Merck Serono, Novartis, Sanofi and Teva, travel and research grants from the Associazione Italiana Sclerosi Multipla. A.P. did not declare any competing interests. C.B. received conference travel support from Biogen Idec, Novartis, Bayer-Schering, Merck-Serono and Teva and has participated in clinical trials by Sanofi Aventis, Roche and Novartis. C.O.-G. received honoraria as consultant on scientific advisory boards from Biogen-Idec, Bayer-Schering, Merck-Serono, Teva and Novartis; has participated in clinical trials/other research projects by Biogen-Idec, GSK, Teva and Novartis. C.R. received research funding and compensation for travel from Biogen-Idec and Novartis. D.H. received speaker honoraria and consulting fees from Biogen Idec, Merck Serono, Teva and Novartis, as well as support for research activities

from Biogen Idec. D.S. received honoraria as a consultant on scientific advisory boards by Bayer-Schering, Novartis and Sanofi-Aventis and compensation for travel from Novartis, Biogen Idec, Sanofi Aventis, Teva and Merck-Serono. E.C. received honoraria as consultant on scientific advisory boards by Biogen-Idec, Bayer-Schering, Merck-Serono, Genzyme and Novartis; has participated in clinical trials/other research projects by Merck-Serono, Roche and Novartis. E.P. served on scientific advisory boards for Genzyme and Biogen-Idec; he has received honoraria and travel grants from Sanofi Aventis, UCB, Lundbeck, Novartis, Bayer Schering, Biogen Idec, Merck Serono, Genzyme and Teva; he has received travel grants from Associazione Marchigiana Sclerosi Multipla e altre malattie neurologiche. E.H. received speaker honoraria and consultant fees from Biogen Idec, Merck Serono, Novartis, Genzyme and Teva, as well as support for research activities from Biogen Idec and Merck Serono. F.G.M. received honoraria or research funding from Biogen Idec, Genzyme, Novartis, Teva Neurosciences, Mitsubishi and ONO Pharmaceuticals. F.V. is an advisory board member for Teva Biogen Merck Serono and Novartis. G.C. has served on scientific advisory boards for and/or received funding for travel from Innate immunity, Klein-Buendel Incorporated, Genzyme, Medimmune, Novartis, Nuron Biotech, Spiniflex Pharmaceuticals, Somahlution, Teva Pharmaceuticals; receives royalties from publishing Evaluation of Health Promotion and Disease Prevention (The McGraw Hill Companies, 1984); has received honoraria from GlaxoSmithKline, Novartis, Advanced Health Media Inc, Biogen Idec, EMD Serono Inc, EDJ Associates Inc, the National Heart, Lung, and Blood Institute, National Institute of Neurological Diseases and Stroke, National Marrow Donor Program, Consortium of Multiple Sclerosis Centers; Mt. Sinai School of Medicine, and Teva Pharmaceuticals; and has served on independent data and safety monitoring committees for Apotek, Ascendis, Biogen-Idec, Cleveland Clinic, Glaxo Smith Klein Pharmaceuticals, Gilead Pharmaceuticals, Modigenetech/Prolor, Merck/Ono Pharmaceuticals, Merck, Neuren, PCT Bio, Teva, Vivus, NHLBI (Protocol Review Committee), NINDS, NMSS, and NICHD (OPRU oversight committee). G.I. had travel/accommodations/meeting expenses funded by Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi Aventis and Teva Guillermo Izquierdo received speaking honoraria from Biogen-Idec, Novartis, Sanofi, Merck Serono and Teva. H.B. has served on scientific advisory boards for Biogen Idec, Novartis and Sanofi-Aventis and has received conference travel support from Novartis, Biogen Idec and Sanofi Aventis. He serves on steering committees for trials conducted by Biogen Idec and Novartis, and has received research support from Merck Serono, Novartis and Biogen Idec. J.L.-S. has accepted travel compensation from Novartis, Biogen and Merck Serono. Her institution receives the honoraria for talks and advisory board commitment and also clinic support from Bayer Health Care, Biogen Idec, CSL, Genzyme Sanofi, Merck Serono and Novartis. J.A.C.-G. did not declare any competing interests. J.L.S.-M. has accepted travel compensation from Novartis and Biogen, speaking honoraria from Biogen-Idec, Novartis, Sanofi, Merck Serono, Almirall, Bayer and Teva and has participated in a clinical trial by Biogen-Idec. M.G. received consulting fees from Teva Canada Innovation, Biogen Idec, Novartis and Genzyme Sanofi; lecture payments from Teva Canada Innovation, Novartis and EMD Serono; has also received a research grant from Canadian Institutes of Health Research. M.F. received speaker honoraria and research funding from Novartis, Biogen Idec, Genzyme, Merck-Serono and Bayer. M.T. received speaking honoraria from Biogen-Idec, Bayer-Schering, Sanofi Aventis, Merck-Serono, Teva and Novartis; has received research grants from Biogen-Idec, Merck-Serono and Novartis. M.S. did not declare any competing interests in relation to this work. M.B. has served on scientific advisory boards for Biogen-Idec, Novartis and Genzyme and has received conference travel support from Biogen-Idec and Novartis. He serves on steering committees for trials conducted by Novartis. His institution has received research support from Biogen-Idec, Merck-Serono and Novartis. M.T. received travel grants from Merck Serono, Novartis, Bayer-Schering, Merck-Serono and Teva; has participated in clinical trials by Sanofi Aventis, Roche and Novartis. O.G. received honoraria as consultant on scientific advisory boards for Genzyme, Biogen Idec, Merck Serono and Novartis; has received travel grants from Biogen Idec, Merck Serono and Novartis; has participated in clinical trials by Biogen Idec and Merck Serono. P.D. did not declare any competing interests. P.G. is a Novartis, Teva-neuroscience, Biogen Idec and Genzyme advisory board member, consultant for Merck Serono, received payments for lectures by Merck Serono, Teva-Neuroscience and Canadian Multiple sclerosis society, and received grants for travel from Teva-Neuroscience and Novartis. R.A. received honoraria from Biologix, Bayer, Merck Sorono, Genpharm, GSK and Novartis, and served on advisory board for Biologix, Bayer, Novartis, Genzyme, Genpahrm and Merck Sorono. R.H. received honoraria as consultant on scientific advisory boards from Merck-Serono, Biogen-Idec, Genzyme-Sanofi and Teva, research funding from Merck-Serono and Biogen-Idec and speaker honoraria from Sanofi-Genzyme and Novartis. R.F.-B. received speaking honoraria from Biogen-Idec, Novartis, Merck Serono and Teva. R.B. received speaker honoraria from Bayer Schering, Biogen Idec, Genzyme, Merck Serono, Novartis, Sanofi-Aventis, Teva; research grants from Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi-Aventis, Teva; congress and travel/accommodation expense compensations by Almirall, Bayer Schering, Biogen Idec, Genzyme, Merck Serono, Novartis, Sanofi-Aventis, Teva. T.S. received compensation for travel from Biogen Idec. T.K. received conference travel support and consultancy/speaker honoraria from Novartis, Biogen Idec, Sanofi Aventis, Genzyme,

Teva, BioCSL and Merck Serono. V.J. has received conference travel support from Novartis. V.V.P. has served on advisory boards for Biogen Idec, Novartis Pharma and Sanofi-Genzyme; has received travel grants and consultancy fees from Biogen Idec, Bayer Schering, Sanofi Aventis, Merck Serono, Sanofi-Genzyme and Novartis Pharma; has received research grants from Bayer Schering.

Supplementary material

Supplementary material is available at *Brain* online.

Web resource

European Medicines Agency. Guideline on clinical investigation of medicinal products for the treatment of Multiple Sclerosis. Draft. 2012 (cited 2014 21/10/2014]; Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000425.jsp&mid=WC0b01ac0580034cf5

References

Amato MP, Fratiglioni L, Groppi C, Siracusa G, Amaducci L. Interrater reliability in assessing functional systems and disability on the Kurtzke scale in multiple sclerosis. Arch Neurol 1988; 45: 746–8.

Amato MP, Portaccio E. Clinical outcome measures in multiple sclerosis. J Neurol Sci 2007; 259: 118–22.

Butzkueven H, Chapman J, Cristiano E, Grand'Maison F, Hoffmann M, Izquierdo G, et al. MSBase: an international, online registry and platform for collaborative outcomes research in multiple sclerosis. Mult Scler 2006; 12: 769–74.

Confavreux C, Vukusic S, Adeleine P. Early clinical predictors and progression of irreversible disability in multiple sclerosis: an amnesic process. Brain 2003; 126 (Pt 4): 770–82.

Ebers GC, Heigenhauser L, Daumer M, Lederer C, Noseworthy JH. Disability as an outcome in MS clinical trials. Neurology 2008; 71: 624–31.

Filippini G, Del Giovane C, Vacchi L, D'Amico R, Di Pietrantonj C, Beecher D, et al. Immunomodulators and immunosuppressants for multiple sclerosis: a network meta-analysis. Cochrane Database Syst Rev 2013; 6: CD008933.

Goodin DS, Traboulsee A, Knappertz V, Reder AT, Li D, Langdon D, et al. Relationship between early clinical characteristics and long term disability outcomes: 16 year cohort study (follow-up) of the pivotal interferon beta-1b trial in multiple sclerosis. J Neurol Neurosurg Psychiatry 2012; 83: 282–7.

Goodkin DE, Cookfair D, Wende K, Bourdette D, Pullicino P, Scherokman B, et al. Inter- and intrarater scoring agreement using grades 1.0 to 3.5 of the Kurtzke Expanded Disability Status Scale (EDSS). Multiple Sclerosis Collaborative Research Group. Neurology 1992; 42: 859–63.

Healy B, Chitnis T, Engler D. Improving power to detect disease progression in multiple sclerosis through alternative analysis strategies. J Neurol 2011; 258: 1812–9.

Hirst CL, Ingram G, Pickersgill TP, Robertson NP. Temporal evolution of remission following multiple sclerosis relapse and predictors of outcome. Mult Scler 2012; 18: 1152–8.

Hohol MJ, Orav EJ, Weiner HL. Disease steps in multiple sclerosis: a longitudinal study comparing disease steps and EDSS to evaluate disease progression. Mult Scler 1999; 5: 349–54.

Kalincik T, Spelman T, Trojano M, Duquette P, Izquierdo G, Grammond P, et al. Persistence on therapy and propensity matched outcome comparison of two subcutaneous interferon beta 1a dosages for multiple sclerosis. PLoS One 2013a; 8: e63480.

Kalincik T, Vivek V, Jokubaitis V, Lechner-Scott J, Trojano M, Izquierdo G, et al. Sex as a determinant of relapse incidence and progressive course of multiple sclerosis. Brain 2013b; 136 (Pt 12): 3609–17.

Leray E, Yaouanq J, Le Page E, Coustans M, Laplaud D, Oger J, et al. Evidence for a two-stage disability progression in multiple sclerosis. Brain 2010; 133 (Pt 7): 1900–13.

Liu C, Blumhardt LD. Disability outcome measures in therapeutic trials of relapsing-remitting multiple sclerosis: effects of heterogeneity of disease course in placebo cohorts. J Neurol Neurosurg Psychiatry 2000; 68: 450–7.

Lublin FD, Baier M, Cutter G. Effect of relapses on development of residual deficit in multiple sclerosis. Neurology 2003; 61: 1528–32.

Lublin FD, Reingold SC, Cohen JA, Cutter GR, Sorensen PS, Thompson, AJ, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. Neurology 2014; 83: 278–86.

Meyer-Moock S, Feng YS, Maeurer M, Dippel FW, Kohlmann T. Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) in patients with multiple sclerosis. BMC Neurol 2014; 14: 58.

Noseworthy JH, Vandervoort MK, Wong CJ, Ebers GC. Interrater variability with the Expanded Disability Status Scale (EDSS) and Functional Systems (FS) in a multiple sclerosis clinical trial. The Canadian Cooperation MS Study Group. Neurology 1990; 40: 971–5.

Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol 2011; 69: 292–302.

Polman CH, Reingold SC, Edan G, Filippi M, Hartung HP, Kappos L, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the 'McDonald Criteria'. Ann Neurol 2005; 58: 840–6.

Pozzilli C, Prosperini L. Clinical markers of therapeutic response to disease modifying drugs. Neurol Sci 2008; 29 (Suppl 2): S211–3.

Ravnborg M, Blinkenberg M, Sellebjerg F, Ballegaard M, Larsen SH, Sorensen PS. Responsiveness of the multiple sclerosis impairment scale in comparison with the expanded disability status scale. Mult Scler 2005; 11: 81–4.

Schumacher GA, Beebe G, Kibler RF, Kurland LT, Kurtzke JF, McDowell F, et al. Problems of experimental trials of therapy in multiple sclerosis: report by the panel on the evaluation of experimental trials of therapy in multiple sclerosis. Ann N Y Acad Sci 1965; 122: 552–68.

R Development Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2011.

Tremlett H, Zhu F, Petkau J, Oger J, Zhao Y. Natural, innate improvements in multiple sclerosis disability. Mult Scler 2012; 18: 1412–21.

Weinshenker BG, Issa M, Baskerville J. Meta-analysis of the placebo-treated groups in clinical trials of progressive MS. Neurology 1996; 46: 1613–9.