Multiple Sclerosis Severity Score

Using disability and disease duration to rate disease severity

R.H.S.R. Roxburgh, FRACP*; S.R. Seaman, PhD*; T. Masterman, PhD; A.E. Hensiek, MRCP; S.J. Sawcer, PhD; S. Vukusic, PhD; I. Achiti, PhD; C. Confavreux, PhD; M Coustans, PhD; E. le Page, PhD; G. Edan, PhD; G.V. McDonnell, MD; S. Hawkins, PhD; M. Trojano, PhD; M. Liguori, PhD; E. Cocco, PhD; M.G. Marrosu, PhD; F. Tesser, MD; M.A. Leone, MD; A. Weber, Dip Biol; F. Zipp, MD; B. Miterski, MD; J.T. Epplen, MD; A. Oturai, PhD; P. Soelberg Sørensen, PhD; E.G. Celius, PhD; N. Téllez Lara, PhD; X. Montalban, PhD; P. Villoslada, PhD; A.M. Silva, MD; M. Marta, MD; I. Leite, MD; B. Dubois, PhD; J. Rubio, PhD; H. Butzkueven, PhD; T. Kilpatrick, PhD; M.P. Mycko, MD; K.W. Selmaj, MD; M.E. Rio, PhD; M. Sá, PhD; G. Salemi, MD; G. Savettieri, MD; J. Hillert, PhD; and D.A.S. Compston, PhD

Abstract—*Background:* There is no consensus method for determining progression of disability in patients with multiple sclerosis (MS) when each patient has had only a single assessment in the course of the disease. *Methods:* Using data from two large longitudinal databases, the authors tested whether cross-sectional disability assessments are representative of disease severity as a whole. An algorithm, the Multiple Sclerosis Severity Score (MSSS), which relates scores on the Expanded Disability Status Scale (EDSS) to the distribution of disability in patients with comparable disease durations, was devised and then applied to a collection of 9,892 patients from 11 countries to create the Global MSSS. In order to compare different methods of detecting such effects the authors simulated the effects of a genetic factor on disability. *Results:* Cross-sectional EDSS measurements made after the first year were representative of overall disease severity. The MSSS was more powerful than the other methods the authors tested for detecting different rates of disease progression. *Conclusion:* The Multiple Sclerosis Severity Score (MSSS) is a powerful method for comparing disease progression using single assessment data. The Global MSSS can be used as a reference table for future disability comparisons. While useful for comparing groups of patients, disease fluctuation precludes its use as a predictor of future disability in an individual.

NEUROLOGY 2005;64:1144–1151

*These authors contributed equally to this work.

From the Neurology Department (Drs. Roxburgh, Sawcer, and Compston, and A.E. Hensiek), Cambridge University, UK; Auckland City Hospital (Dr. Roxburgh), New Zealand; Max Planck Institute for Psychiatry (Dr. Seaman), Munich, Germany; Division of Neurology (Drs. Masterman and Hillert), NEUROTEC, Karolinska Institute at Huddinge University Hospital, Stockholm, Sweden; EDMUS Coordinating Centre (Drs. Vukusic, Achiti, and Confavreux), Service de Neurologie A, Hôpital Neurologique, Lyon, France; Clinique Neurologique (Drs. Coustans, le Page, and Edan), Centre Hospitalier Universitaire de Rennes, Hôpital de Pontchaillou, France; Northern Ireland Regional Neurology Service (Drs. McDonnell and Hawkins), Royal Victoria Hospital, Belfast; Department of Neurological and Psychiatric Sciences (Drs. Trojano and Liguori), University of Bari, Italy; Department of Neuroscience (Drs. Cocco and Marrosu), Centro Sclerosi, Multipla University of Cagliari, Italy; Clinica Neurologica (Drs. Tesser and Leone), Università del Piemonte Orientale, "A. Avogadro," Novara, Italy; Clinica Neurologica (Drs. Tesser and Leone), Ospedale Maggiore della Carità, Novara, Italy; Institut für Neuroimmunologie (A. Weber and Dr. Zipp), Klinische und Experimentelle Neuroimmunologie, Universitätsklinikum Charité, Neurowissenschaftliches Zentrum, Berlin; Abteilung für Humangenetik (Drs. Miterski and Epplen), Ruhr-Universität Bochum, Germany; Department of Neurology (Drs. Oturai and Soelberg Sørensen), Copenhagen University Hospital, Rigshospitalet, Denmark; Department of Neurology (Dr. Celius), Ullevål University Hospital, Oslo, Norway; Clinical Neuroimmunology Unit (Drs. Téllez Lara and Montalban), Vall D'Hebron, University Hospital, Barcelona, Spain; Multiple Sclerosis Centre (Dr. Villoslada), Department of Neurology, Clinica Universitaria de Navarra, Pamplona, Navarra, Spain; Neuroimmunology Group (Drs. Silva, Marta, and Leite), Department of Neurology, Hospital Geral Santo Antonio, Porto, Portugal; University Hospital Gasthuisberg (Dr. Dubois), University of Leuven, Belgium; The Walter and Eliza Hall Institute of Medical Research (Drs. Rubio, Butzkueven, and Kilpatrick), IG Royal Parade, Parkville, Victoria, Australia; The Cooperative Research Centre for Discovery of Genes for Common Human Diseases (Dr. Kilpatrick), Richmond, Victoria, Australia; Department of Neurology (Drs. Mycko and Selmaj), Medical University of Lodz, Poland; Neurology Department (Drs. Rio and Sá), Hospital de S. Joáo, Porto, Portugal; and Department of Neuropsichiatry (Drs. Salemi and Savettieri), Palermo, Italy.

The Cambridge Multiple Sclerosis Genetics Group and GAMES project were supported by the Wellcome Trust (grant 057097) supplemented by the National Multiple Sclerosis Society of the USA, the Multiple Sclerosis Society of Great Britain and Northern Ireland, and Multiple Sclerosis International Federation. Other groups received financial or clinical assistance from the following: Bari, the Italian Multiple Sclerosis Society; Belfast, MS Society, Action MS and the Royal Group Hospitals Trust; Berlin, the Bundesministerium für Bildung und Forschung and the Ernst Schering Research Foundation; Bochum, Drs. M. Haupts, S. Sindern, and D. Pöhlau, and "Gemeinnützige Hertie-Stiftung" (project no. 1.319.110/02/05); Copenhagen, the Danish Multiple Sclerosis Society; Leuven, Belgian Charcot Foundation, University Hospital Research Council, Leuven; Lyons, Biogen, Schering, Serono and Teva pharmaceutical companies, ARSEP and LFSEP; Stockholm, the Swedish Medical Research Council, the Swedish Association of Neurologically Disabled, Karolinska Institutet, Swedish Brain Foundation and the Swedish Society of Medicine; Tasmania, CRC for Discovery of Genes for Common Human Diseases, the National Multiple Sclerosis Society of the USA, and B. Taylor.

Received November 6, 2003. Accepted in final form December 9, 2004.

Address correspondence and reprint requests to Dr. R. Roxburgh, Neurology Dept., Cambridge University, Hills Rd., Cambridge CB2 2QQ, UK; e-mail: richard@roxburgh.net.nz

There is considerable individual variation in the clinical course of multiple sclerosis (MS). This partly reflects the random involvement of separate anatomic pathways that vary in their thresholds for clinical expression, but may also depend on more fundamental differences in the etiology, disease mechanisms, and responses to treatment. In all these contexts, clinical research is handicapped by lack of a validated method for measuring the rate at which disability accumulates in the individual (for simplicity, referred to hereafter as "disease progression"). Here, we describe a system that utilizes information about disease duration more effectively, and therefore allows improved precision in assessing cohorts for disease severity.

Two approaches have been used to measure differences in progression of MS. One involves following patients longitudinally and measuring the time taken to reach a given level of disability. The other assesses patients once only during the course of their disease—a method that depends on the hypothesis, which we test in this article, that such assessments reflect disease severity as a whole. The issue is how best to adjust single measurements of disability for disease duration because, if raw disability scores are compared without this correction, information is squandered and larger samples needed to detect real differences in rates of disease progression. The most validated and widely accepted measure of disability is the Expanded Disability Status Scale score (EDSS).[1] This is ordinal but, despite numbered steps, intervals between individual half points are not linear. Because there is no simple relation between EDSS and disease duration, correcting for this parameter is not straightforward. One solution has been to derive a ratio of the EDSS and time from onset (in years) as the Progression Index (PI). The fact that disability, as measured by the EDSS, is not linear limits this approach. The Multiple Sclerosis Severity Score (MSSS) corrects EDSS for duration by using an arithmetically simple method to compare an individual's disability with the distribution of scores in cases having equivalent disease duration. By applying this method to a database of 9,892 patients we created a reference table for future comparisons. Simulation shows that the MSSS is a useful method for identifying factors that influence disease progression using single assessment data.

Methods. *Clinical resources.* Each group involved in the Genetic Analysis of Multiple Sclerosis in Europeans (GAMES) consortium[2] was invited to contribute clinical data. Details relating to date of birth, sex, clinical course (primary progressive/relapsing-remitting or secondary progressive), age at first symptoms, single or sequential EDSS scores, time from first symptoms to EDSS assessment, methods of recruitment, and previous exposure to immunomodulatory drugs were collected. Ethical approval was obtained for each group from local committees.

Additional material related to this article can be found on the *Neurology* Web site. Go to www.neurology.org and scroll down the Table of Contents for the April 12 issue to find the title link for this article.

In most cohorts, only a single assessment was made on each patient. Where multiple assessments had been made, one group (Spain) submitted the first, another (Bari) the last, and others (Sardinia and Rennes) selected one at random for inclusion in the Global MSSS. Two groups (Rennes and Lyons) made serial data available. These were used to assess stability of MSSS scores over time, and to determine whether EDSS measurements made early in the disease course predicted later disability. A few of the smaller single assessment datasets included observations made at times of relapse. Serial datasets other than the Spanish specifically excluded them.

For most populations, patients were selected from existing databases of cases referred to centers specializing in the management of MS (Sweden, n = 1,071; Norway, n = 391; Denmark, n = 460; Belgium, n = 153; Germany [Bochum], n = 589; Germany [Berlin], n = 359; France [Rennes], n = 1,630; Italy [Bari], n = 1,741; Italy [Novara], n = 198; Italy [Sardinia], n = 1,378; and Portugal, n = 215). Some cohorts were supplemented by self-referral or access to databases used previously for epidemiologic or genetic research (Cambridge, n = 1,021; Spain, n = 267; and Australia, n = 176). Two datasets were strictly population-based (Northern Ireland [Belfast], n = 243; and France [Lyons], n = 1,975). All patients had a diagnosis of clinically or laboratory supported definite MS.[3] Three small groups of patients (from Poland, n = 140; Portugal, n = 130; and Sicily, n = 42) were excluded. These groups were only able to provide complete data on a small proportion of their total MS patients, raising the possibility that they might be particularly vulnerable to selection bias.

Method for deriving Multiple Sclerosis Severity Scores. The MSSS algorithm is a simple method for adjusting disability for disease duration. Patients were stratified by the number of whole years from first symptoms to EDSS assessment. To reduce the impact of stochastic fluctuations over time, each year was analyzed with the two on either side. Thus, for example, year 5 results were generated from data for all patients with onset of symptoms attributable to MS from 3 to 7 years previously. Within each year EDSS scores were ranked and the average of the lowest and highest ranks for each possible EDSS value (0, 1, 1.5 . . . 9.5) was calculated. These averages were then normalized by dividing by 1 + the number of available assessments for that year. The normalized values were multiplied by 10 to provide a range from 0 to 10 (for easier comparison with raw EDSS), and rounded to two decimal places. Therefore, MSSS is the decile of the EDSS within the range of patients who have had the disease for the same disease duration. We used this algorithm to construct three tables. The Global MSSS table was derived from combined data excluding the Lyons patients. One Local MSSS was derived for the Rennes, France dataset and a second Local MSSS for the combined Italian data. These were used to test whether the MSSS generated from local data were substantially more powerful than the Global MSSS.

Comparison of datasets and natural history data. To investigate how representative our samples were for MS, we compared median EDSS and 25th to 75th centile ranges in the 15 contributing datasets. Disability in the overall combined dataset was compared with published data from the Ontario natural history study.[4,5] In cohorts providing serial observations we chose a disability assessment at random within the disease course.

Stability of EDSS rankings over time and exploration of earliest useful EDSS data. Reliance on single point assessments assumes that data are representative of disease severity and not unduly influenced by sampling variation or disease fluctuation. Particular concerns have been expressed with respect to EDSS scores recorded early in the clinical course. Therefore, we assessed in individual patients how well disability measured at one point of the disease correlated with disability measured later in the disease course. In the longitudinal datasets, using Spearman's rank correlation coefficient, we compared EDSS rankings in years 0, 1, 2, 3, 4, and 5 with EDSS rankings for the same individuals 5 ($\pm$1), 10 ($\pm$1), and 15 ($\pm$2) years later.

Simulation studies to compare power of MSSS with other measures of disease progression. The ability of the MSSS and previous scales to detect a risk factor influencing disease severity was compared using simulations based on a variety of scenarios, i.e., assumed models for simulating datasets. In all scenarios, we assumed that half the patients were exposed (Group E) and the other half not exposed (Group U) to a risk factor such as, for

Table 1 *Models used to simulate the genetic effect on disease progression*

Model	Increase in EDSS as a result of exposure
I	0.5 throughout the disease course
II	0.5 up to 15 y and then by 1.0
III	0.5 only after 15 y
IV	2.0 throughout the disease course
V	0.5 for the first 5 y, 1.0 for the next 5 y, 1.5 for the next 5 y

EDSS = Expanded Disability Status Scale.

example, possession of the mutant allele at a genetic susceptibility locus. The scenarios were constructed by assuming the following: either 500 (i.e., 250 in Group E and 250 in Group U) or 1,000 (i.e., 500 in each group) patients in the study; a distribution of EDSS over duration either similar to that seen in Rennes, France or the combined Italian data (the most divergent of the larger datasets in terms of their distributions of disability over time); one of five models (table 1) for the effect of exposure on progression; and penetrance for the risk factor of 0, 0.1, 0.2, . . . , 1 (i.e., 11 points). Thus, there were a total of $2 \times 2 \times 5 \times 11 = 220$ scenarios. Here, penetrance means the probability that an exposed individual is affected by that exposure. For example, in the scenario where penetrance was Model 1 (in which EDSS increased by 0.5 throughout the course of the disease) and penetrance 0.3, durations and EDSS values would be sampled in the same way for exposed and unexposed individuals and then each exposed individual would have a probability of 0.3 of his or her sampled EDSS being increased by 0.5. For each individual, the Global MSSS, Local MSSS, and progression index (PI) were calculated.

Five thousand datasets were simulated for each of the 220 scenarios. For each dataset, the Wilcoxon rank sum test was used to test for a significant difference between Groups E and U in Global MSSS scores, Local MSSS scores, raw EDSS scores, and PI (both with and without exclusion of the first 5 years). For each of the scenarios, the power of the tests based on Global and Local MSSS, raw EDSS, and PI was calculated as the proportion of the 5,000 datasets in which the corresponding test was significant ($p < 0.05$). The power of a method for categorical analysis[6] was calculated using the same approach.

Results. *Demographics of datasets.* Altogether, data were collected on 11,867 patients from 16 research groups based in 11 countries (table 2). Sixty-six percent were Northern European; 33% were from Southern Europe and 1% from Tasmania (mainly of Northern European origin). The sex ratio was 2.2 F:M. Eleven percent had primary progressive disease. The average age at disease onset was 30.6 years and the average disease duration at measurement of EDSS was 11.7 years.

Around 500 patients were included for each of the first 11 years of disease duration: at least 200 patients were available up to year 20; and >100 patients to year 27. Combining data from each year with the two adjacent years contributed >1,000 data points for the generation of Global MSSS scores annually to year 20, and >500 per annum for years 20 to 26 (figure 1). To compare datasets, we examined progression of the EDSS in each group as a function of time (figure 2). Contributing datasets were somewhat different from each other. Therefore, when comparing the power of Global MSSS against Local MSSS we specifically chose two of the most disparate large samples as the basis for our simulations to see whether this disparity would have a substantial influence on power. Our com-

Table 2 *Demographic and treatment data for the 16 participating cohorts*

Country	Location	Number of patients	Female, %	Mean age, y	Primary progressive, %	Immunomodulatory drugs other than corticosteroids prior to EDSS,* %
United Kingdom	Cambridge	1,021	74.6	26.3	8.2	<5
	Belfast	243	70.0	31.2	12.8	5
Scandinavia	Sweden	1,071	70.9	31.6	11.1	20
	Norway	391	69.8	31.7	22.5	<5
	Denmark	460	65.7	31.6	14.8	19
France	Rennes	1,630	64.7	31.0	25.0	75
	Lyons	1,975	60.6	31.7	6.9	25
Belgium	Leuven	153	64.4	30.7	4.8	25
Germany	Bochum	589	67.7	29.5	18.8	NA
	Berlin	359	73.0	30.8	4.2	20
Italy	Bari	1,741	67.6	28.2	12.6	37
	Novara	198	67.7	32.1	3.0	26
	Sardinia	1,378	68.4	28.8	9.0	35
Portugal	Oporto	215	70.7	29.1	8.4	60
Spain	Barcelona	267	65.5	30.5	15.0	0
Australia	Tasmania	176	71.0	32.4	7.4	50
Total		11,867	68.4	30.6	10.8	27

* Figures for use of immunomodulatory drugs are estimates only: see text for details.

EDSS = Expanded Disability Status Scale; NA = data not available.

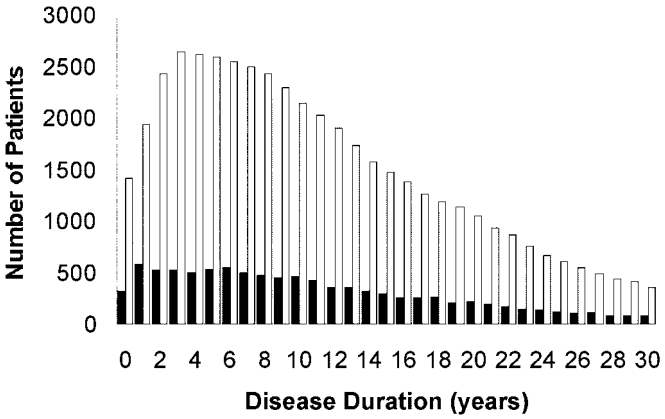

Figure 1. Total number of patients and data points available for each year of duration. The filled bars represent the numbers of patients included at each disease duration. We combined data from each year with the two years on either side and the sample size for each year is represented by clear bars.

bined sample was similar in progression to data from the Ontario natural history study[4,5] (table 3).

Validation of single point data: stability of patient disability rankings over time. A total of 8,498 assessments were available on the 1,975 patients from the Lyons dataset and 7,557 on the 1,630 patients from Rennes. Measurements made in the first year (year 0) were not reliable indicators of disease severity, having low correlation (table 4) with measurements later in the disease course. Therefore, we excluded data from year 0 from any further calculations including construction of the Global MSSS and our power simulations. Correlations for year 1 were markedly greater than for year 0. Correlations between measurements taken at year 1 and 5, 10 or 15 years later, were similar to correlations 5, 10, and 15 after year 5. This indicates that single point assessment data from as early as year 1 can be used to represent disease severity.

Table 3 *Comparison of overall disease progression with the longitudinal follow-up study in London, Ontario, Canada*

Duration, y	Proportion of patients with EDSS greater than or equal to:	Present study, %	Ontario, %
5	3	48	37*
	6	14	14†
	8	4	2*
10	3	67	56*
	6	30	32†
	8	5	5*
15	3	75	80†
	6	38	50†
	8	11	10*
20	3	85	86*
	6	58	64*
	8	14	17*

Figures cited here are either estimated from diagrams*[4,5] or stated in the text of the articles cited.†

EDSS = Expanded Disability Status Scale.

Calculation of Global MSSS. We derived a Global MSSS table from the combined data on 9,892 patients, excluding the cases from Lyons (assessed using the DSS), for years 1 to 30 and EDSS points 0 to 9.5 (figure 3). For example, an individual with symptoms for 10 years and an EDSS of 4 has a Global MSSS score of 5.28. Another patient with symptoms for 20 years and the same EDSS would have a Global MSSS score of 2.99. The algorithm used to derive the Global MSSS ensured that, for any given year, scores increase with higher values for EDSS. This is shown by reading from left to right across columns of the table. Although there is no a priori obligation for the MSSS associated with a given value of EDSS to decrease

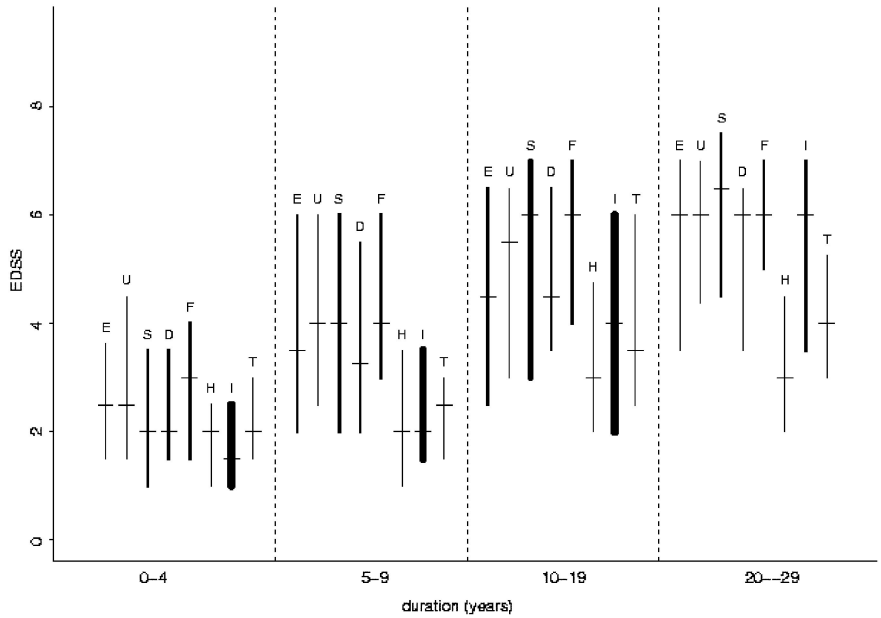

Figure 2. Data for each period of 5 years are combined. Vertical lines represent the 25th to 75th centile range for each dataset. The median is shown as a horizontal bar. The width of the vertical lines is proportional to the number of patients represented: E = England, U = Belfast, S = Scandinavia, D = Germany, F = France + Belgium, H = Spain + Portugal, I = Italy, T = Tasmania.

Table 4 *Correlation between individual disability rankings at different time points in the Lyons and Rennes cohorts*

		Lyons			Rennes		
First assessment	Follow-up assessment	No. of observations	Spearman correlation coefficient	Mean and 80% range for change in MSSS	No. of observations	Spearman correlation coefficient	Mean and 80% range for change in MSSS
Year 0	+5	178	0.27 (0.13–0.40)	−3.2 (−8.8–+0.7)	69	0.47 (0.26–0.63)	−0.4 (−3.4–+3.2)
	+10	95	0.14 (−0.06–0.33)	−3.8 (−9.0–+0.8)	38	0.55 (0.28–0.74)	−0.2 (−3.3–+3.3)
	+15	90	0.22 (0.02–0.41)	−3.7 (−8.7–+0.6)	29	0.07 (−0.30–0.43)	−1.0 (−5.1–+3.2)
Year 1	+5	137	0.65 (0.55–0.74)	−0.8 (−3.2–+1.5)	70	0.66 (0.50–0.77)	0.0 (−2.0–+3.0)
	+10	66	0.57 (0.38–0.71)	−1.7 (−5.6–+1.1)	22	0.47 (0.06–0.75)	0.2 (−2.0–+3.6)
	+15	50	0.67 (0.48–0.80)	−1.5 (−5.6–+1.1)	21	0.40 (−0.04–0.71)	−1.2 (−5.0–+2.5)
Year 2	+5	127	0.67 (0.56–0.76)	−0.8 (−3.7–+1.5)	90	0.49 (0.31–0.63)	0.1 (−3.0–+2.5)
	+10	62	0.42 (0.19–0.60)	−2.2 (−6.6–+1.4)	29	0.67 (0.40–0.83)	−0.7 (−3.8–+1.9)
	+15	49	0.55 (0.31–0.72)	−1.4 (−5.2–+1.3)	23	0.66 (0.34–0.84)	−1.0 (−4.3–+2.9)
Year 3	+5	105	0.85 (0.79–0.90)	−0.5 (−2.1–+1.5)	97	0.73 (0.62–0.81)	−0.1 (−2.1–+2.2)
	+10	57	0.76 (0.62–0.85)	−1.0 (−3.6–+1.3)	34	0.57 (0.29–0.76)	−0.7 (−3.6–+1.9)
	+15	42	0.68 (0.48–0.82)	−1.0 (−4.4–+1.6)	25	0.76 (0.53–0.89)	−1.5 (−4.1–+1.2)
Year 4	+5	112	0.77 (0.68–0.83)	−0.3 (−2.4–+2.2)	86	0.75 (0.63–0.83)	0.0 (−1.7–+1.8)
	+10	53	0.81 (0.69–0.88)	−0.7 (−3.4–+1.6)	31	0.57 (0.26–0.77)	0.2 (−2.3–+2.4)
	+15	35	0.59 (0.32–0.77)	−1.6 (−5.4–+1.7)	16	0.80 (0.51–0.93)	−0.7 (−2.9–+1.4)
Year 5	+5	89	0.76 (0.66–0.84)	−0.1 (−2.6–+2.6)	71	0.72 (0.59–0.82)	0.1 (−1.7–+2.4)
	+10	54	0.81 (0.69–0.89)	−0.4 (−2.9–+2.0)	32	0.65 (0.39–0.82)	−0.1 (−2.8–+2.6)
	+15	29	0.77 (0.56–0.89)	−0.6 (−3.1–+1.8)	18	0.33 (−0.16–0.69)	−0.2 (−3.5–+3.6)
Year 10	+5	66	0.91 (0.85–0.94)	−0.1 (−1.6–+1.8)	57	0.79 (0.66–0.87)	0.4 (−0.9–+2.1)
	+10	28	0.71 (0.46–0.86)	0.1 (−2.3–+1.5)	19	0.66 (0.30–0.86)	0.5 (−1.7–+3.1)
	+15	17	0.79 (0.51–0.92)	−0.7 (−2.4–+1.5)	11	0.66 (0.10–0.90)	2.2 (−0.1–+4.5)

Spearman correlation coefficient (95% CI) and the mean and range of values by which MSSS was observed to change over the intervening periods are shown. +5 years includes second assessments between 4 and 6 years, +10 years includes assessments between 9 and 11 years, +15 years includes assessments between 13 and 17 years. The total number of patients in this table does not match the total number of patients in the Rennes and Lyons datasets in table 2 because this table only shows those patients who happened to have assessments at these time intervals.

MSSS = Multiple Sclerosis Severity Scale.

with increasing disease duration (reading vertically down each column), this was in fact observed in the vast majority of cases. Derivation of the Global MSSS is illustrated for year 10 in table E-1 (on the *Neurology* Web site at www.neurology.org). A program is available for download from http://www-gene.cimr.cam.ac.uk/MSgenetics/GAMES/MSSS that calculates Global and Local MSSS values and performs the Kruskal-Wallis tests based on these.

We aimed only to include assessments made when the EDSS was stable. However, some of the contributing groups were unable to identify which measurements were made at times of relapse. We estimate that these make up less than half a percent of assessments used to construct the Global MSSS and hence are unlikely to affect its power and applicability.

The stability of a person's MSSS score over time is also shown in table 4, with mean and 80% range for changes over 5, 10, and 15 years. The mean change in MSSS was about zero. The 80% range for the change in MSSS was about ±2, considerably less than ±5.5, which would be expected if scores were taken at random, reflecting the moderate to strong correlation between EDSS measurements made at different time points on the same person. However, the ranges are not negligible, indicating that an individual's MSSS scores can alter quite a lot over time. Thus, although differences in mean MSSS (or EDSS) between two groups of patients reflect real differences in mean disease progression, one should be cautious about using MSSS to predict future disease severity in any single patient: any such prediction will be subject to considerable uncertainty.

Simulation studies to compare power of MSSS with other measures of disease progression. To test the ability of the MSSS to detect factors that influence disease progression, we simulated groups exposed and not exposed to a putative risk factor, and compared sensitivity of the MSSS with other published methods. Figure 4 shows results from the simulation in which 500 patients were exposed and 500 not exposed to a factor modifying disease progression by 0.5 EDSS points (Model 1, see table 1), based on the Italian cohort. MSSS was more powerful in detecting simulated differences between exposed and unexposed groups than were EDSS (MSSS [G] vs EDSS), PI with and without exclusion of the first 5 years (MSSS [G] vs PI and PI2), and categorical

	0	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5	EDSS
1	0.67	2.44	4.30	5.87	7.08	7.93	8.64	9.09	9.35	9.50	9.63	9.74	9.84	9.90	9.94	9.97	9.98	9.98	9.99	
2	0.53	2.01	3.69	5.24	6.46	7.27	7.98	8.58	8.95	9.18	9.38	9.59	9.79	9.88	9.93	9.97	9.99	9.99	9.99	
3	0.45	1.77	3.34	4.82	6.00	6.81	7.54	8.14	8.55	8.83	9.07	9.35	9.63	9.77	9.86	9.92	9.97	9.98	9.99	
4	0.35	1.45	2.87	4.27	5.41	6.24	6.98	7.65	8.12	8.42	8.70	9.08	9.47	9.68	9.80	9.88	9.95	9.98	9.99	
5	0.30	1.28	2.60	3.90	4.95	5.79	6.58	7.26	7.75	8.08	8.38	8.83	9.32	9.60	9.76	9.86	9.95	9.98	9.99	
6	0.25	1.13	2.33	3.54	4.55	5.38	6.14	6.81	7.33	7.66	7.98	8.50	9.08	9.45	9.68	9.81	9.93	9.97	9.99	
7	0.24	1.04	2.10	3.17	4.13	4.96	5.75	6.46	6.98	7.32	7.65	8.24	8.91	9.33	9.59	9.76	9.90	9.95	9.99	
8	0.21	0.94	1.92	2.93	3.81	4.57	5.36	6.10	6.61	6.95	7.32	7.97	8.71	9.21	9.55	9.74	9.89	9.96	9.99	
9	0.21	0.88	1.76	2.65	3.45	4.17	4.93	5.64	6.14	6.50	6.90	7.65	8.53	9.09	9.47	9.70	9.87	9.95	9.99	
10	0.19	0.78	1.53	2.34	3.10	3.79	4.55	5.28	5.77	6.14	6.58	7.39	8.31	8.92	9.34	9.61	9.83	9.94	9.99	
11	0.17	0.71	1.40	2.13	2.82	3.46	4.21	4.94	5.42	5.82	6.30	7.18	8.15	8.79	9.24	9.52	9.78	9.92	9.98	
12	0.16	0.64	1.28	1.98	2.64	3.25	3.94	4.63	5.13	5.54	6.03	6.92	7.93	8.63	9.13	9.43	9.71	9.88	9.97	
13	0.13	0.57	1.14	1.80	2.44	3.05	3.70	4.38	4.91	5.32	5.80	6.74	7.83	8.55	9.03	9.34	9.65	9.85	9.96	
14	0.11	0.49	1.03	1.70	2.33	2.91	3.55	4.26	4.82	5.23	5.70	6.56	7.59	8.34	8.86	9.20	9.57	9.82	9.95	
15	0.10	0.45	0.99	1.64	2.26	2.82	3.44	4.14	4.68	5.09	5.51	6.33	7.41	8.17	8.70	9.11	9.51	9.78	9.94	
16	0.09	0.38	0.85	1.42	1.99	2.56	3.17	3.86	4.41	4.81	5.18	6.00	7.14	7.97	8.54	9.04	9.49	9.75	9.94	
17	0.05	0.32	0.76	1.28	1.77	2.30	2.95	3.65	4.17	4.55	4.94	5.74	6.89	7.77	8.38	8.99	9.52	9.79	9.96	
18	0.04	0.26	0.66	1.12	1.57	2.09	2.70	3.37	3.89	4.27	4.62	5.43	6.62	7.54	8.23	8.94	9.51	9.78	9.96	
19	0.05	0.28	0.63	1.00	1.39	1.89	2.50	3.19	3.72	4.12	4.49	5.35	6.59	7.51	8.22	8.98	9.57	9.81	9.96	
20	0.05	0.26	0.59	0.94	1.29	1.71	2.29	2.99	3.51	3.93	4.30	5.15	6.43	7.45	8.23	8.98	9.58	9.80	9.95	
21	0.05	0.30	0.66	1.02	1.39	1.77	2.34	2.97	3.43	3.83	4.21	5.09	6.35	7.33	8.08	8.87	9.49	9.77	9.96	
22	0.04	0.23	0.54	0.90	1.28	1.66	2.20	2.82	3.29	3.69	4.09	5.04	6.35	7.35	8.10	8.84	9.42	9.73	9.95	
23	0.05	0.27	0.58	0.91	1.26	1.64	2.19	2.78	3.21	3.69	4.19	5.16	6.47	7.46	8.20	8.87	9.43	9.75	9.95	
24	0.05	0.24	0.52	0.86	1.25	1.63	2.15	2.71	3.09	3.52	4.01	5.03	6.36	7.38	8.15	8.81	9.39	9.74	9.96	
25	0.05	0.23	0.47	0.77	1.15	1.56	2.05	2.53	2.84	3.21	3.74	4.88	6.26	7.24	8.00	8.73	9.35	9.75	9.98	
26	0.05	0.20	0.45	0.78	1.17	1.58	2.08	2.63	2.99	3.40	3.95	5.02	6.39	7.44	8.21	8.89	9.48	9.80	9.96	
27	0.05	0.22	0.48	0.78	1.15	1.56	2.03	2.56	2.91	3.29	3.86	4.93	6.33	7.38	8.14	8.91	9.56	9.85	9.98	
28	0.04	0.17	0.40	0.74	1.16	1.52	1.88	2.39	2.76	3.04	3.46	4.54	5.99	7.07	7.90	8.75	9.45	9.80	9.98	
29	0.03	0.18	0.47	0.80	1.19	1.51	1.79	2.27	2.68	3.01	3.41	4.35	5.68	6.76	7.66	8.62	9.38	9.75	9.96	
30	0.01	0.13	0.45	0.82	1.19	1.45	1.69	2.23	2.75	3.13	3.50	4.35	5.61	6.66	7.54	8.47	9.27	9.67	9.91	
Years																				

=1st Decile
=2nd Decile
=3rd Decile
= 4th Decile
= 5th Decile
= 6th Decile
= 7th Decile
= 8th Decile
= 9th Decile
= 10th Decile

Figure 3. Global Multiple Sclerosis Severity Scores (MSSS) generated from 9,892 European patients. The MSSS for an individual patient is ascertained by finding the column corresponding to the patient's Expanded Disability Status Scale and the row corresponding to the number of years since the onset of multiple sclerosis. Deciles are color coded to show the pattern of disease progression at different disease durations.

analysis of results (MSSS [G] vs CAT). Similar results were obtained from all five different models tested (see table 1) and when using the Rennes cohort in place of the Italian data. Where effects of the putative risk factor were very weak, no method was particularly powerful. Similarly, when effects were strong, all except the category method had close to 100% power. The MSSS showed most advantage when effects were modeled to be intermediate. The Global MSSS was similar (and often superior) in power to MSSS derived from the datasets used to generate the simulations (MSSS [G] vs MSSS [L]). PI was no more powerful and in most simulations less powerful than raw EDSS.

Discussion. There is no consensus method for measuring progression in MS using single, cross-sectional assessments of disability. Many exploratory methods have been used,[7] none of which is supported by empirical evidence or critical analysis. While it is statistically valid to compare disease progression between groups of patients using raw disability scores, failure to take account of disease duration reduces the information utilized and, therefore, the power to detect real differences between groups. We have taken an approach to measuring disease severity that uses this information by comparing individual EDSS scores with the distribution of EDSS in patients who have had MS for the same length of time. A simple algorithm was developed for deriving decile scores for disability and applied to a cohort of 9,892 patients with disease durations ranging from 1 to >30 years to create the Global MSSS.

Using serial data from two large French cohorts, we show that single point measures of disability are representative of disease course overall. The ranking of a patient's disability at one point correlates with the ranking of disability in the same patient later in the disease course. Previously suggested methods[8] have excluded data from the first 5 years. However, we found that single point assessments made as early as the second year of disease show sufficient correlation with later disability measurements to inform the analysis of disease severity. It has been generally accepted that serial studies are superior to single point assessments of disease progression. However, a longitudinal survey providing serial measurements of EDSS in the same patient but analyzed at only one point, for example, time taken to reach an EDSS of 6.0, is not necessarily more powerful than a study in which each patient is assessed once only, provided appropriate adjustment for disease duration.

In addition to describing an algorithm that may be used for deriving the MSSS from any adequate dataset, we have provided the Global MSSS from which a patient's disease progression can be rated directly if

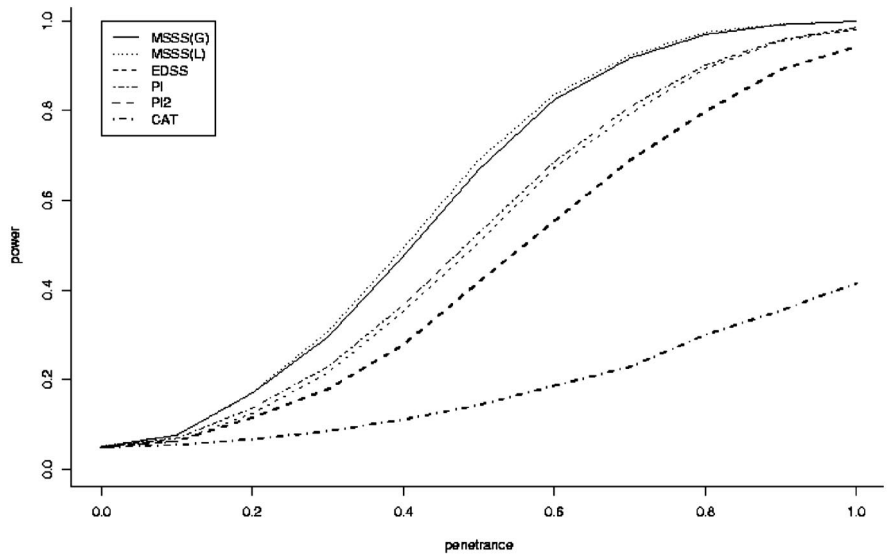

Figure 4. Representative example of the power of Multiple Sclerosis Severity Score (MSSS) and alternative severity methods using Model 1 (exposed patients were at risk of their Expanded Disability Status Scale [EDSS] being increased by 0.5) in 1,000 patients simulated using the Italian dataset. The power (the proportion of 5,000 simulations having a statistically significant difference between exposed and unexposed groups at the 5% level) is shown as a function of penetrance (the proportion of exposed individuals affected). MSSS (L) = MSSS derived from local (Italy) data; MSSS (G) = Global MSSS; PI = progression index; PI2 = PI with first 5 years of assessments discarded; CAT = categorical data.

the EDSS and disease duration are known. Data for the Global MSSS were combined from 16 different research groups representing 11 European nations participating in the GAMES collaboration.[2] The majority came from sources where only a single measure of disability was recorded for each patient. In others, one from a number of serial measurements was selected. The source of data was largely clinic based. Recruitment in these groups may have been influenced by factors affecting disease severity. Thus, for example, more patients exposed to drug therapy would be expected where cases are identified through a department actively involved in the initiation and monitoring of disease modifying treatments. Where enrollment was initially motivated by referral to another research project, the extremes of clinical severity may not have been included. At first sight this pooling of somewhat disparate datasets may be called into question. However, it is important to remember that power of the Global MSSS test based on pooled data does not rely on these being a true representation of disease progression but only on the condition that future datasets resemble these data. The validity of data pooling was justified by the finding that power of the Global MSSS is very similar to that of Local MSSS even at extreme ends of the dataset distribution observed in this study. This implies that the Global MSSS will be applicable to any future study with disease progression that falls within the bounds of the progression patterns collected here. The fact that, in addition, progression rates were similar to those seen in a population-based epidemiologic survey[4,5] implies that the vast majority of MS progression data will satisfy this condition. However, in some situations, local factors may make it preferable to generate a Local MSSS, if necessary stratifying for age[9] or long-term use of immunotherapy before applying a stratified Wilcoxon's rank sum test or, if more than two groups of patients are compared, a Kruskal-Wallis test.[10] Inclusion of a few data from assessments made at times of relapse will have had a negligible effect on the overall Global MSSS and can further be excused along the same lines, namely that future datasets are likely to include some such data.

The average MSSS showed stability over time although, as expected from the unpredictable nature of MS, there were not inconsiderable changes in individual MSSS scores. The implication of this finding is that MSSS is a useful measure for studies of groups of patients but cannot be used to predict very adequately later disability in an individual. Thus one group of patients with a higher mean MSSS than another is likely to maintain a higher MSSS 5, 10, or even 15 years later even though individual MSSS scores within the groups may fluctuate over time.

The simulations show that the test based on Global MSSS is more powerful than previous methods for single-point assessment studies. That is, it is more likely to find a statistically significant difference between two groups whose rates of progression truly differ. The MSSS reduced the number of patients required to detect this difference by up to one half, compared to the next most powerful method.

An examination of the method used for comparing ordinal data (the Wilcoxon's rank sum test[11]) suggests a theoretical basis for why MSSS might be expected to be more powerful than methods that do not take into account disease duration and even those that do such as the PI. Scores from both groups being compared are ranked together and the difference between average ranks in the two groups calculated. Power is lost if disease duration is not taken into account, since patients with longer disease durations tend to have higher ranks, whether or not they carry the risk factor. Using the MSSS, all patients with average disability for disease duration, by definition, score around 5. Patients with more accelerated disease progression receive higher scores and those with less severe disease are assigned lower scores, whatever the disease duration, thus increasing the likelihood of identifying factors that affect

disease progression. Furthermore, the method for determining MSSS scores compensates for variation in sensitivity of the EDSS to detect changes in disability. Thus, at points on the scale that are insensitive to increasing disability, a small change in EDSS will be reflected by a larger change in rankings than at scores where the EDSS is more sensitive to increasing disability. By comparison, adjustment used for the PI does not assist in ranking patients according to disease severity. Instead, it merely consolidates the grouping of patients by disease duration, though in the opposite order from raw EDSS. The increase in power of the MSSS is not at the expense of reduced specificity, which remains unchanged at the nominal value (e.g., 5% if the standard threshold of $p < 0.05$ is used). In fact a positive disease severity effect detected by MSSS is more likely to be a true finding than one found by another method as they may give false positive results because of poorly matched disease durations in the groups being compared.

When comparing disease progression between groups, it is important to avoid confounding due to case ascertainment. For example, a tendency for older patients with mild disease to have a lower referral rate than younger cases with equivalent disability might lead to the spurious conclusion of faster progression in the older group. This is different from the situation in which recruitment is affected by a risk factor only through its influence on EDSS, because, in this case, exposed and nonexposed patients with the same EDSS are just as likely to be recruited.

When analyzing single assessment data, there may be some uncertainty about the precise date of disease onset from which to infer disease duration when anamnestic data are used. This uncertainty applies not just to MSSS but also to any cross-sectional data that utilize information about disease duration and it is exceedingly unlikely that using this additional information would be outweighed by uncertainty of disease onset in some individuals.

Typical applications of the MSSS might be in epidemiologic studies, such as those that correlate disease progression among different family members with MS, and in studies of genetic association, where disease progression is compared between groups with different alleles at a particular locus. A number of rating scales other than the EDSS are currently in use for assessing morbidity in MS. These include global measures such as the Multiple Sclerosis Functional Composite[12] and the Multiple Sclerosis Impact Scale,[13] and scales that assess a particular neurologic system. The core principle of the MSSS—correcting single assessment scores by comparison with the distribution in patients of similar disease duration—could be applied to each.

Acknowledgment

The authors thank Dr. E. Willoughby for assistance with the manuscript.

References

1. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983;33:1444–1452.
2. Sawcer S, Maranian M, Setakis E, et al. A whole genome screen for linkage disequilibrium in multiple sclerosis confirms disease associations with regions previously linked to susceptibility. Brain 2002;125:1337–1347.
3. Poser CM, Paty DW, Scheinberg L, et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. Ann Neurol 1983;13:227–231.
4. Weinshenker BG, Bass B, Rice GP, et al. The natural history of multiple sclerosis: a geographically based study. I. Clinical course and disability. Brain 1989;112:133–146.
5. Weinshenker BG. The natural history of multiple sclerosis. Neurol Clin 1995;13:119–146.
6. Schmidt S, Barcellos LF, DeSombre K, et al. Association of polymorphisms in the apolipoprotein E region with susceptibility to and progression of multiple sclerosis. Am J Hum Genet 2002;70:708–717.
7. Kantarci OH, de Andrade M, Weinshenker BG. Identifying disease modifying genes in multiple sclerosis. J Neuroimmunol 2002;123:144–159.
8. Weinshenker BG, Wingerchuk DM, Liu Q, Bissonet AS, Schaid DJ, Sommer SS. Genetic variation in the tumor necrosis factor alpha gene and the outcome of multiple sclerosis. Neurology 1997;49:378–385.
9. Trojano M, Liguori M, Bosco Zimatore G, et al. Age-related disability in multiple sclerosis. Ann Neurol 2002;51:475–480.
10. Schindel F, Kwasny H. Stratified or multicentre analysis with extended Mantel-Haenszel statistics. Biometrical J 1991;33:755–762.
11. Fisher LD, van Belle G. Biostatistics. New York: Wiley, 1993.
12. Cutter GR, Baier ML, Rudick RA, et al. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. Brain 1999;122:871–882.
13. Hobart J, Lamping D, Fitzpatrick R, Riazi A, Thompson A. The Multiple Sclerosis Impact Scale (MSIS-29): a new patient-based outcome measure. Brain 2001;124:962–973.