J Neurol (2004) 251 : 1472–1480
DOI 10.1007/s00415-004-0567-0

ORIGINAL COMMUNICATION

European validation of a standardized clinical description of multiple sclerosis

Maria Pia Amato
Jerome Grimaud
Iuliana Achiti
Maria Letizia Bartolozzi
Patrice Adeleine
Hans-Peter Hartung
Ludwig Kappos
Alan Thompson
Maria Trojano
Sandra Vukusic
Christian Confavreux
for the Evaluation of the EDMUS system (EVALUED) Study Group

Received: 4 December 2003
Received in revised form: 6 April 2004
Accepted: 2 June 2004

J. Grimaud · A. Thompson
Institute of Neurology
The National Hospital
London, UK

I. Achiti · S. Vukusic · Ch. Confavreux
Service de Neurologie A and EDMUS Coordinating Center
Hôpital Neurologique
Lyon, France

P. Adeleine
Unité de Biostatistiques et Informatique Médicale
Lyon, France

H.-P. Hartung
Neurologische Klinik
Heinrich-Heine-Universität
Dusseldorf, Germany

L. Kappos
Neurologische Universitätspoliklinik
Kantonsspital Basel
Basel, Switzerland

M. Trojano
Department of Neurology
University of Bari
Bari, Italy

M. P. Amato (✉) · M. L. Bartolozzi
Department of Neurological Sciences
University of Florence
Viale Morgagni, 85
50134 Florence, Italy
Tel.: +39 (0)-55-4277836
Fax: +39 (0)-55-4277673
E-Mail: mariapia.amato@unifi.it

M. P. Amato and J. Grimaud contributed equally to the study. The members of the Study Group are listed in appendix.

Abstract *Objectives* The EDMUS system is a clinical database specifically tailored to the description of multiple sclerosis (MS). The EVALUED (Evaluation of the EDMUS system) study is an European project with two objectives: 1) to assess the inter-examiner reliability of the whole EDMUS system; 2) to validate the EDMUS-Grading Scale (EGS), which is a simplified version of the Kurtzke Disability Status Scale (DSS). *Methods* The protocol included 12 neurologists working in pairs within six European centres (Bari, Basel, Florence, London, Lyon, Würzburg). They assessed independently 30 MS patients in their centre, filling in the EDMUS forms. The reliability of the system was assessed on selected key items in the history of the MS onset, the clinical course and the disease course classification. The clinical examination of the patients permitted an assessment of the Kurtzke Expanded Disability Status Scale (EDSS) and the EGS. Level of agreement was measured in terms of kappa and weighted kappa indexes whenever appropriate. *Results* The study included 180 patients with definite or probable MS of whom 37 % were males. Age was 35.8 ± 9.6 years (mean ± SD), disease duration 7.8 ± 5.7 years, and mean EDSS score 4.1 ± 2.2. The disease course was relapsing-remitting in 67 %, secondary progressive in 22 %, and progressive from disease onset in 11 %. For key items of the history, the inter-examiner reliability level ranged from moderate to excellent. Concerning the disability scales, perfect agreement was reached in 59 % for EDSS and 78 % for EGS. The close correlation and linear association ($r = 0.94$, $p < 0.0001$) between both scales demonstrated EGS's construct validity. *Conclusion* The EDMUS system allows a consistent clinical description of MS using a common language. This standardized follow-up of MS patients is valuable especially in studies requiring a critical mass of informative patients.

Key words multiple sclerosis · clinical description · standardisation · database

JON 1567

Introduction

The development of a database using a common language and computerization for the recording, storing and retrieval of data has considerable clinical and research potential [6, 11]. Medical practice is made easier by allowing rapid access to the relevant features of the patient's record. Research is also made more straightforward. Within and among centres using the same standards, selection of appropriate files, exchange of data, and comparison of individual studies are facilitated. Files from various centres can be pooled for common studies. Information from a critical mass of patients thus becomes available. It allows fundamental questions to be addressed with more power, but also encourages new questions that could not have been addressed at the level of a single centre. These considerations taken together are particularly relevant in a disease such as multiple sclerosis (MS) where clinicians and researchers often use varying terminology and for which prevalence is relatively low.

The European Database for Multiple Sclerosis (EDMUS) is a standardised, computerised databasing system which has been conceived and developed within consecutive European Concerted Actions on MS since 1990. It is the result of joint reflections by clinicians and researchers all involved in MS from the whole European Union. A Steering Committee with at least one delegate from each country of the European Union has been set up for this purpose. Continuous interaction with the users has been maintained [6, 8, 11]. Today the EDMUS system is established in more than 200 centres over 28 countries. It is used for the clinical follow-up of patients, independent research projects [9, 13, 14, 19, 31, 36–39] and collaborative multicentre studies [10, 12, 15].

One of the main terms introduced in the system is the EDMUS Grading Scale (EGS) [8]. For many reasons, and despite its well-known limitations [21, 32, 33], the Kurtzke Expanded Disability Status Scale (EDSS) [26] remains the "gold standard" for grading clinical impairment and disability in MS. However, it is far too complex and time-consuming for non-MS specialists and even more so for epidemiological purposes. This is why the EDMUS Steering Committee decided to design a simplified version of the original EDSS which would allow similar grading of the patients but could be administered quickly and assessed for its essential points directly through interviewing the patient.

The evaluation of the EDMUS system (EVALUED) study was a European collaborative multicentre enterprise. The specific objectives of the project were: 1) to assess the inter-examiner reliability of the whole EDMUS system on selected key-items; 2) to assess the inter-examiner reliability of the EDSS and the EGS as well as their correlation.

Material and methods

EDMUS description

The EDMUS system [8] covers identification and demographic data, medical history, key-episodes in the MS course (relapses, onset of the progressive course, time to assignment of successive scores of irreversible disability), biological, electrophysiological and neuro-imaging data, and treatment.

Reliability was assessed on selected key items of the EDMUS system which are pivotal in describing the clinical history of the disease. Regarding onset of the disease, the focus was on the date, disease type and symptoms. For the type of disease onset two categories are described, each with two possibilities: relapsing-remitting with or without sequelae and primary progressive with or without an initial relapse. A relapse is defined as the occurrence of new symptoms, the re-appearance of pre-existing ones, or the worsening of current symptoms lasting at least 24 hours [8]. Fatigue alone or a transient, fever-related worsening of symptoms are not considered to be a relapse. Symptoms that occur within one month of each other are considered as part of the same relapse. Sequelae must produce at least a minimum ambulation-related disability or a significant non ambulation-related disability, such as grade 3 on the EGS. Progression is defined as a continuous worsening of symptoms and signs for at least 6 months, with or without superimposed relapses [8, 34]. Symptoms are distributed into several main categories: "long tracts" symptoms (lower extremity dysfunction, upper extremity dysfunction, sensory symptoms, sphincter disturbance and/or sexual disturbance); "brainstem" symptoms (oculomotor, facial motor, facial sensory, vestibular and/or cochlear, and bulbar symptoms); and "optic neuritis".

For disease evolution, the focus was on the time interval between the onset of the disease and the second neurological episode (which may either be a relapse or the onset of the progressive phase), on the number of relapses and on the conversion to a secondary progressive phase for patients with an initially relapsing-remitting course.

Disability was scored on the EGS [8] and the EDSS and its Functional Systems (FS) [26]. By construction, the EGS is derived from the Kurtzke's Disability Status Scale (DSS) [24, 25] with one point level of disability ranging between 0 and 10, each level having a short and precise description. Grade 0 describes normal status, 1 to 3 minimal signs or no ambulation related disability, 4 to 5 moderate ambulation-related disability, 6 to 7 severe ambulation-related disability, 8 to 9 heavily dependent status, and 10 death due to MS (Table 1).

Study design

Before the beginning of the study, 12 neurologists from six participating centers (Basel – Switzerland; Bari and Florence – Italy; London – United Kingdom; Lyon – France; Würzburg – Germany) underwent a one-day common training session, in order to define clearly and agree upon terminology, criteria and data recording on the EDMUS system. During the study, in each participating centre two neurologists examined a sample of 30 consecutive MS patients. They were not allowed to examine more than five patients a day. Each patient was assessed on the same day, consecutively and independently by each neurologist of the pair, who were required to operate in a random order. Assessment included the collection of all the data concerning the clinical history and the neurological examination with scoring of disability on the EDSS and EGS, and completion of the EDMUS forms. All the participating neurologists refrained from discussing their assessments and ratings until the end of the study.

Patients

The study sample included 180 patients (67 men; 113 women) with definite or probable MS according to Poser's criteria [30]. The mean

Table 1 The EDMUS Grading scale (EGS) reprinted from Confavreux et al. 1992; Journal of Neurology, Neurosurgery and Psychiatry; 55:671–676 with permission

EGS Score	
0	Normal findings on neurological examination.
1	No disability. Minimal signs on neurological examination.
2	Minimal and not ambulation-related disability. Able to run.
3	Unlimited walking distance without rest, but unable to run; or a significant not ambulation-related disability.
4	Walks without aid. Limited walking distance, but > 500 meters without rest.
5	Walks without aid. Walking distance < 500 meters without rest.
6	Walks with uni- or bilateral support. Walking distance < 100 meters without rest.
7	Home restricted. A few steps with wall or furniture assistance. Walking distance < 10 meters without rest.
8	Chair restricted. Unable to take a step. Some effective use of arms.
9	Bedridden and totally helpless.
10	Death due to mutliple sclerosis.

age was 35.8 years (SD 9.6) and mean disease duration was 7.8 years (SD 5.7). The disease course was relapsing-remitting in 121 patients (67.2 %), secondary progressive in 39 (21.7 %) and progressive from MS onset in 20 (11.1 %). The mean EDSS score was 4.1 (SD 2.2). Table 2 shows the characteristics of patients recruited in each study centre (results were analysed using a variance analysis). All subjects gave their informed consent to participate in the study.

Statistical analysis

The graphical methods described by Tukey [40], Altman and Bland [1], and Bland and Altman [3] were used to assess the reliability of continuous variables. For dichotomous variables inter-observer reliability was measured by means of the kappa index, which provides a correction for the extent of the agreement to be expected by chance [16, 27]. For ordinal variables reliability was assessed in terms of the weighted kappa, which provides appropriate weights to take into account the different magnitude of disagreements [5].

In the statistical analysis, evidence for construct validity of the EGS was provided by examining the strength of the relationship between the EGS scores and the EDSS, which was used as the gold standard. Inter-examiner reliability on disability scores was assessed in terms of the Intraclass Correlation Coefficient (ICC) [4]. The agreement level was interpreted conventionally as: < 0 poor, 0–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial, and > 0.80 almost perfect [27]. As reliability estimates are population dependent, 95 % confidence intervals were constructed for the ICC [17]. Percentage of agreement (e. g., frequency of perfect agreement) was also measured in order to facilitate the comparison with results from other reports [2, 18, 20–22, 29, 44].

Results

Reliability of the description of the onset of MS

The date of onset of MS was given precisely by the two examiners within a day in 61 %, and within a year in 93 % of the cases. The median of the difference between the date of onset given by the two examiners was nil (mean 2.63 days, SD 118.1). In all cases of major disagreements, one of the two examiners omitted to register the initial relapse which was not described by the patient or not considered a true relapse by the examiner.

In the assessment of the type of onset according to the four EDMUS categories, the kappa statistics revealed a moderate to substantial agreement in three centres with kappa values ranging from 0.59 to 0.68, and an almost perfect agreement in the remaining three, with kappa values over 0.80 (Table 3, upper part). When the four categories merged into two only, i. e. relapsing-remitting and primary progressive, reliability improved in all the centres with kappa values around or above 0.80 (Table 3, lower part).

As for the detailed symptoms at the onset of MS, the level of agreement ranged from fair to almost perfect. After grouping the symptoms in three main categories ("long tracts", "brainstem" and "optic neuritis"), the kappa values ranged from 0.46 to 1.0 (Table 4).

Table 2 Characteristics of the patients recruited in each study center

	Bari	Basel	Florence	London	Lyon	Würzburg
Mean age ± SD (years)* (range)	31.2±8.1 (20.4–52.6)	37.8±10.5 (23.4–63.7)	35.4±9.0 (24.5–58)	39.6±11.4 (16.2–57.6)	41.3±12.3 (17.0–62.8)	33.8±6.8 (20.9–50.7)
Mean disease duration ± SD (years) (range)	6.7±4.6 (0.9–19.4)	8.5±5. 9 (0.6–24.7)	6.0±5.9 (0–21.7)	8.5±5.1 (1.0–18.3)	9.0±6.3 (0.3–24.7)	7.6±5.1 (1.4–24.1)
Mean EDSS score ± SD** (range)	3.3±2.3 (0.0–8.0)	3.6±1.3 (1.5–6.0)	3.3±2.2 (1.0–7.0)	7.3±0.9 (6.0–9.0)	3.8±1.9 (0.0–8.0)	3.6±2.0 (1.0–8.5)
Type of course**						
– Relapsing-Remitting	70%	65%	70%	10%	49%	88%
– Secondary-Progressive	15%	32%	10%	70%	36%	12%
– Primary Progressive	15%	3%	20%	20%	15%	–

* $p < 0.001$; ** $p < 0.0001$ (variance analysis)
SD standard deviation; *EDSS* Expanded Disability Status Scale

Table 3 Inter-examiner agreement for the description of the onset of multiple sclerosis

	Type of MS onset according to the 2 examining neurologists																KAPPA (ASE)
Examiner N° 1	RO				SO				POR				POP				
Examiner N° 2	RO	SO	POR	POP	RO	SO	POR	POP	RO	SO	POR	POP	RO	SO	POR	POP	
Basel	21	1	·	·	·	3	·	·	·	·	·	·	1	·	·	4	0.84 (0.11)
Bari	23	3	·	·	·	4	·	·	·	·	·	·	·	·	·	·	0.67 (0.17)
Florence	24	1	·	·	·	·	·	·	·	·	1	·	·	·	·	4	0.89 (0.10)
London	12	4	·	·	1	6	·	·	·	·	·	·	·	2	1	4	0.59 (0.12)
Lyon	21	1	·	·	·	2	·	·	·	·	·	·	·	·	1	5	0.85 (0.10)
Würzburg	19	3	·	·	1	7	·	·	·	·	·	·	·	·	·	·	0.68 (0.14)
All centers	120	13	·	·	2	22	·	·	·	·	1	·	1	2	2	17	0.76 (0.05)

Table 3 *Continued*

Examiner N°1	RO/SO		POR/POP		KAPPA (ASE)
Examiner N°2	RO/SO	POR/POP	RO/SO	POR/POP	
Basel	25	·	1	4	0.87 (0.13)
Bari	30	·	·	·	–
Florence	25	·	·	5	1.00 (0.00)
London	23	·	2	5	0.79 (0.14)
Lyon	24	·	·	6	1.00 (0.00)
Würzburg	30	·	·	·	–
All centres	157	·	3	20	0.92 (0.05)

Upper part of the table: the onset of multiple sclerosis is described according to four categories.
Lower part of the table: the onset of multiple sclerosis is described according to two categories.
For each examined patient, the first examiner selected one of the four (upper part of the table) or two (lower part of the table) possibilities for the description of the type of onset of multiple sclerosis. The corresponding selection of the second examiner is indicated in the appropriate columns. The agreement is indicated for each participating centre by the kappa index and the asymptotic standard error (ASE).
RO onset of multiple sclerosis with a relapse followed by remission without sequelae; *SO* onset of multiple sclerosis with a relapse followed by remission with sequelae; *POP* progressive onset of multiple sclerosis without inaugural superimposed relapse; *POR* progressive onset of multiple sclerosis with inaugural superimposed relapse; *RO/SO* onset of multiple sclerosis with a relapse followed by remission with or without sequelae; *POP/POR* progressive onset of multiple sclerosis with or without inaugural superimposed relapse

Table 4 Inter-examiner agreement in the description of symptoms at the onset of multiple sclerosis according to the six participating centers

Centre	Symptoms		
	LT	BS	ON
Basel	0.68 (0.15)	0.46 (0.17)	0.49 (0.16)
Bari	0.57 (0.16)	0.77 (0.12)	0.67 (0.18)
Florence	0.91 (0.09)	0.93 (0.07)	1.00 (0.00)
London	0.77 (0.12)	0.78 (0.12)	0.83 (0.12)
Lyon	0.76 (0.16)	0.76 (0.13)	1.00 (0.00)
Würzburg	1.00 (0.00)	0.79 (0.12)	0.52 (0.24)

The symptoms were grouped in three main categories.
The agreement is indicated for each participating centre by the kappa index and the asymptotic standard error (ASE).
LT long tracts involvement. Refers to any combination of lower extremity dysfunction, upper extremity dysfunction, sensory symptoms, sphincter disturbance, or sexual disturbance; *BS* brainstem involvement. Refers to any combination of facial motor, facial sensory, oculomotor symptoms, vestibular and/cochlear, or bulbar dysfunction; *ON* optic neuritis

Reliability of the description of the course of MS

For the 152 patients who experienced a second neurological episode after MS onset, the time interval between the onset of the disease and the second neurological episode was given precisely within a day in 44.1 %, within a month in 60.5 %, and within a year in 88.8 % of the cases. The same median number of relapses was registered by the two examiners in all but one centre. The difference however was not significant (data not shown).

Both examiners agreed on the occurrence of the secondary progressive phase in 39 patients with an initially relapsing-remitting course, while they disagreed in 17 cases. In terms of the kappa values, agreement on secondary progression turned out to be moderate in two centres, substantial in two others and perfect in the remaining two centres (Table 5). For the 39 patients who were classified by both examiners as secondary progressive cases of MS, the date of conversion to the secondary progressive phase was given precisely within a year in 71.8 %, and within 3 years in 89.7 % of the cases. Inter-examiner agreement in identifying the disease course was moderate in three centres (kappa 0.45 to 0.54) and substantial in the remaining three (kappa 0.76 to 0.79) (data not shown).

Reliability of the scoring of the neurological disability

The degree of inter-examiner reliability in EDSS and EGS scores concerning all patients is shown on Fig. 1 as a grid correlate (Fig. 1). The degree of inter-examiner re-

Table 5 Inter-examiner agreement in the description of the conversion to secondary progression for patients with an initially relapsing-remitting course

	Conversion from relapsing-remitting course to secondary progressive course				
Examiner N°1	yes		no		
Examiner N°2	yes	no	yes	no	Kappa (ASE)
Basel	3	3	–	24	0.62 (0.19)
Bari	7	2	3	18	0.62 (0.15)
Florence	3	–	–	27	1.00 (0.00)
London	15	3	3	9	0.58 (0.15)
Lyon	9	–	–	21	1.00 (0.00)
Würzburg	2	–	3	25	0.52 (0.23)
All centers	39	17		124	0.76 (0.06)

For each examined patient, the first examiner selected one of the two possibilities (yes or no) for the description of the conversion to secondary progression. The corresponding selection of the second examiner is indicated in the appropriate column. The agreement is indicated for each participating centre by the kappa index and the asymptotic standard error (ASE)

liability in EDSS and EGS scores is also shown for each centre separately in Table 6. In terms of the ICC, the agreement level was almost perfect for both scales. Frequency of agreement varied according to the definition of agreement. For the EDSS, it ranged from 43 to 83 % when the perfect identity of scores was required, from 77 to 97 % for differences of no more than 0.5 EDSS point and from 87 % to 100 % for differences of no more than 1.0 EDSS point. For the EGS, the percentage of exact agreement ranged from 50 to 97 % and reached 87–100 % when agreement was defined as a difference in scores of no more than 1 EGS point (Table 6).

Convergent construct validity of the EGS (Fig. 2) is demonstrated by the strength of the relationship with the EDSS applied to the same patient at the same examination: the correlation turned out to be linear with $r = 0.94$ ($p < 0.0001$).

Discussion

The search for a "common language" to describe the pattern and course of MS has become a major aim, because of the increasing number of collaborative studies and clinical trials which require rapid identification of appropriate patients and pooling of data from different centres. To our knowledge, the EVALUED study is the first attempt toward a cross-national standardization in describing the clinical course of the disease and disability outcomes. In the analysis we selected some key items in EDMUS, which turn out to be essential for a uniform description of the illness, especially in terms of disease onset and subsequent evolution. Globally, for most of the items, the degree of agreement ranged between moderate and excellent.

Differences in the assessment of the date of MS onset exceeded one year only in 7 % of the subjects, and in these cases disagreements related to the interpretation of the event "relapse". The difficulty in classifying some minor clinical episodes, particularly in retrospect is well-known. Even with a prospective follow-up, it might be difficult to distinguish a true relapse from transient deterioration of the patient's clinical status due to other factors, such as infections, psychological factors, temperature or fatigue. A good level of concordance was reached in defining the type of onset as described by the EDMUS categories. For the relapsing-remitting onset, all disagreements related to the presence or absence of

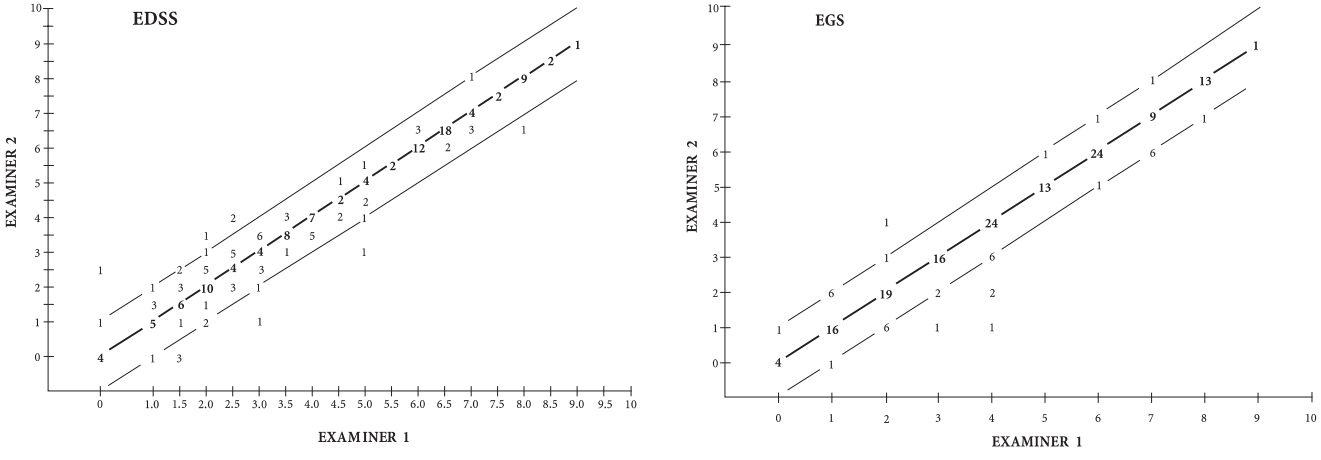

Fig. 1 Inter-examiner reliability in the assessment of the EDSS (left side) and the EGS (right side) scores. Each figure in the graph indicates the number of patients for which examiner N° 1 and examiner N° 2 selected the corresponding score on the abscissae and ordinate axes, respectively. This grid correlate concerns the total group of 180 patients (*EDSS* Kurtzke Expanded Disability Status Scale [26]; *EGS* EDMUS Grading Scale (cf. Table1). The highlighted numbers on the diagonal lines illustrate the frequency of perfect agreement between the two examiners

Table 6 Inter-examiner agreement in the scoring of the neurological disability according to the participating centres

Centres	Cumulative percents of agreement							
	ICC (95% CI)	No difference	≤ 0.5 point difference	≤ 1.0 point difference	≤ 1.5 point difference	≤ 2.0 point difference	≤ 2.5 point difference	≤ 3.0 point difference
EDSS								
Basel	0.96 (0.93–0.99)	56.7	76.7	86.7	96.7	100.0	100.0	100.0
Bari	0.92 (0.87–0.98)	43.3	86.7	96.7	100.0	100.0	100.0	100.0
Florence	0.98 (0.95–1.00)	70.0	96.7	100.0	100.0	100.0	100.0	100.0
London	0.91 (0.80–1.00)	83.3	93.3	96.7	100.0	100.0	100.0	100.0
Lyon	0.84 (0.64–1.00)	46.7	80.0	90.0	96.7	96.7	100.0	100.0
Würzburg	0.97 (0.95–1.00)	53.3	96.7	100.0	100.0	100.0	100.0	100.0
All centres	Not appropriate	58.9	88.3	93.9	98.3	99.4	100.0	100.0
EGS								
Basel	0.97 (0.94–1.00)	73.3		96.7		100.0		100.0
Bari	0.96 (0.93–0.99)	83.3		100.0		100.0		100.0
Florence	1.00 (0.99–1.00)	96.7		100.0		100.0		100.0
London	0.86 (0.75–0.96)	73.3		100.0		100.0		100.0
Lyon	0.99 (0.98–1.00)	93.3		100.0		100.0		100.0
Würzburg	0.84 (0.71–0.97)	50.0		86.7		96.7		100.0
All centres	Not appropriate	78.3		97.2		99.4		100.0

Cumulative percents of agreement: percentage of time both examiners of a given participating centre selected the same EDSS (upper part of the table) or the same EGS (lower part of the table) score when examining the 30 patients of the centre.
ICC Intraclass Correlation Coefficient; *95% CI* 95% Confidence Interval; *EDSS* Kurtzke Expanded Disability Status Scale; *EGS* EDMUS Grading Scale

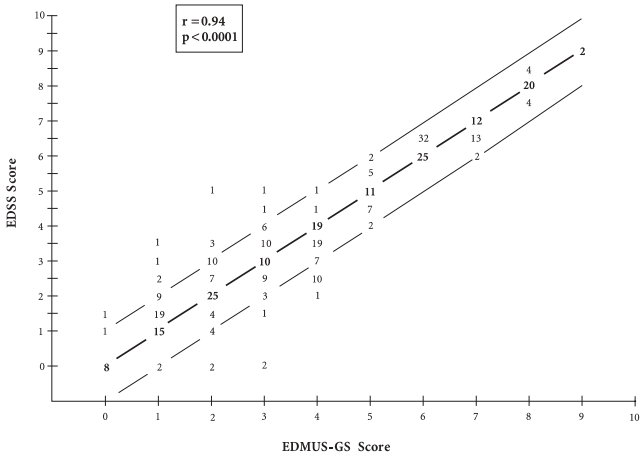

Fig. 2 Relationship between the EGS and EDSS scores for the 180 patients (all examiners, all centres). The highlighted numbers on the diagonal lines illustrate the frequency of perfect agreement between the scores

functional sequelae after the initial episode. In contrast, for the primary progressive onset, disagreements came from a different interpretation of the presence or the absence of a pre-existing relapse. Agreement became excellent when the four categories merged into two. We also found an acceptable reliability level regarding the symptoms at the onset of the disease, especially when we merged them into three categories. As for the description of the disease evolution, we found a good level of concordance concerning the duration of the first inter-attack period and time to convert to secondary progression. The reliability level ranged from moderate to substantial for the interpretation of the disease course. Classification of the disease course has received growing attention in the literature over recent years. On the basis of an international consensus, Lublin and Reingold [28] provided standardized definitions for the courses of MS. They identified a relapsing-remitting course (RR), with or without full recovery, a primary progressive course (PP), a secondary progressive course (SP), with or without occasional relapses, and a progressive-relapsing form (PR), defined as a disease with a progressive onset with later superimposed relapses. The term relapsing-progressive (RP) was rejected since there was no consensus definition. Kremenchutzky et al. [23] evaluated the validity of the terms PR and RP on a large population-based cohort of progressive MS patients seen at the London MS Clinic (Canada). Since there was no difference in terms of distribution of survival in time to reach DSS 3, 6, 8 and 10 when comparing the categories of progressive MS, the authors recommended that the terms RP and PR be dropped and that they be included in the PP or SP groups. The EDMUS system is consistent with these conclusions and retains only the distinction between primary and secondary progressive forms, with or without superimposed relapses.

Only a small number of the many suggested scales for MS are used and none of these fulfils the requirements of the international MS community [21, 22, 35]. Although new more sensitive and multidimensional measures have been proposed, particularly for use in clinical trials [32, 33], Kurtzke's EDSS remains, so far, the most

widely used scoring system and represents a reference criterion for other suggested scales. It has, nonetheless, acknowledged limitations: it is unresponsive, it combines impairment and disability, it has often been shown to have only moderate interrater reliability, it is not entirely objective and its overall score is heavily weighted toward ambulation. The EGS is a simplified version of the DSS, ranging from 0 (no impairment) to 10 (death from MS) 1-point steps. It is derived from the Kurtzke DSS scale, but with a simplified and concise wording, in order to avoid ambiguity, to facilitate clinical practice and to be appropriate for epidemiological purposes. Our results on the reliability of the EDSS largely confirm the findings from previous reports [2, 18, 20–22 29, 44]. We did not assess the levels of agreement for different ranges of the scales because some patients had to be included in different groups according to the scoring of the two examiners and no appropriate statistic tool was available at the time for this purpose. Nevertheless, it seems that the lower portion of the scales yielded the worst agreement level (see Fig. 1). In spite of this, agreement was better for the EGS scale than for the EDSS scale in all the levels of the scale. The EGS therefore is at least as reliable as the EDSS, far more simple to administer and far less time consuming. It reaches the same score with fewer constraints for the patient and the examiner. Although it maintains some of the limitations perceived with the use of the EDSS, it may be a valid tool particularly for clinical practice, even for retrospective records, in standard follow-up and for collaborative epidemiological studies.

These results taken together have been obtained from a cohort of European MS patients which seems representative of the disease when compared with the natural history cohorts available in the literature [7, 41–43]. However, in the interpretation of the data, it must be taken into account that our findings may in part overestimate the reliability level. Since patients were examined twice consecutively on the same day, a memory effect for some variables cannot be ruled out. Besides, this study was performed in specialised MS centres, where neurologists had been trained in the use of the EDMUS system. This being said, if there has been a bias, it has occurred for both EGS and EDSS assessments evenly, which does confirm the reliability of the comparisons we have made between these two scales.

Finally, although the participants had undergone a common training session before the beginning of the study, we found some differences in the agreement level among the participating centres. These differences may be due, in part, to different characteristics of the patients recruited in each centre, and in part to different familiarity of the examiners with the EDMUS system which was in daily use in Lyon and Florence, but not in the other centres.

More generally, this European study has shown that a clinical databasing system such as EDMUS is a reliable and useful tool for the clinical description of MS. This "common language" is of interest for multicentre studies. On the basis of this experience, a new version of the EDMUS software has been developed in order to further improve the reliability of the common language and the flexibility of the system, and to cover the needs of the majority of users.

Appendix

The EVALUED Study Group: Adeleine P, Follet A, Unité de Biostatistiques et Informatique Médicale, Hospices Civils de Lyon, Lyon-France. Amato MP, Bartolozzi ML, Dipartimento di Scienze Neurologiche, Firenze-Italy. Confavreux Ch, Cortinovis-Tourniaire P, Hours M, Hôpital Neurologique, Lyon-France. Hartung HP, Morrisey S, Flachenecker P, Neurologische Universitätsklinik, Würzburg-Germany. Kappos L, Huber S, Lechner-Scott J, Neurologische Universitätspoliklinik, Basel-Switzerland. Livrea P, Trojano M, Avolio C, Cattedra di Neuropathologia e Psicopathologia, Bari-Italy. Thompson AJ, Hobart J, Grimaud J, Institute of Neurology, London-UK.

Acknowledgements We are indebted to Mrs Michèle Canova for technical assistance, and to the patients who contributed so much of their time. Supported by contracts with the Commission of the European Communities DG XII (BMH1-CT93–1529, CIPD-CT97–0227 and BMH4-CT96–0064), and by grants from the Ligue Française contre la Sclérose en Plaques (LFSEP), the Association pour la Recherche sur la Sclérose en Plaques (ARSEP), and the Hospices Civils de Lyon. We are grateful to Biogen France, Schering SA, Serono France and Aventis/Teva Pharma France laboratories for their steady support to the EDMUS project.

References

1. Altman DG, Bland JM (1983) Measurement in medicine: the analysis of method comparison studies. Statistician 32:307–317
2. Amato MP, Fratiglioni L, Groppi C, Siracusa G, Amaducci L (1988) Interrater reliability in assessing functional systems and disability on Kurtzke scale in multiple sclerosis. Arch Neurol 45: 746–748
3. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 8:307–310
4. Cohen J (1960) A coefficient of agreement for nominal scales. Educational and psychological measurement 20: 37–46
5. Cohen J (1968) Weighted Kappa: nominal scale agreement with provision for scaled disagreement on partial credit. Psychol Bull 70:213–220
6. Confavreux C (1994) Establishment and use of multiple sclerosis registers. Ann Neurol 36:S136–S139
7. Confavreux C, Aimard G, Devic M (1980) Course and prognosis of multiple sclerosis assessed by the computerized data processing of 349 patients. Brain 103:281–300

8. Confavreux C, Compston DAS, Hommes OR, McDonald WI, Thompson AJ (1992) EDMUS, a European database for multiple sclerosis. J Neurol Neurosurg Psychiatry 55:671–676
9. Confavreux C, Hours M, Moreau T, Cortinovis-Tourniaire P, Grimaud J for the participants to EDMUS European Concerted Action (1996) Clinical databasing in multiple sclerosis: EDMUS and the European effort. In: Abramsky O, Ovadia H (eds) Frontiers in Multiple Sclerosis Clinical Research and Therapy. Martin Dunitz Ltd, London, pp 299–312
10. Confavreux C, Hutchinson M, Hours M, Cortinovis-Tourniaire P, Moreau T, and the Pregnancy in Multiple Sclerosis Group (1998) Rate of pregnancy-related relapse rate in multiple sclerosis. N Engl J Med 339:285–291
11. Confavreux C, Paty DW (1995) Current status of computerization of multiple sclerosis clinical data for research in Europe and North-America: The EDMUS/MS-COSTAR connection. Neurology 45:573–576
12. Confavreux C, Suissa S, Saddier P, Bourdès V, Vukusic S, and the Vaccines in Multiple Sclerosis Group (2001) Vaccinations and the risk of relapse in multiple sclerosis. N Engl J Med 344: 319–326
13. Confavreux C, Vukusic S, Moreau T, Adeleine P (2000) Relapses and progression of disability in multiple sclerosis. N Engl J Med 343:1430–1438
14. Confavreux C, Vukusic S, Adeleine P (2003) Early clinical predictors and progression of irreversible disability in multiple sclerosis: an amnesic process. Brain 126:770–782
15. Edan G, Miller DH, Clanet M, Confavreux C, Lyon-Caen O, Lubetzki C, Brochet B, Berry I, Rolland Y, Froment JC, Cabanis E, Iba-Zizen MT, Gandon JM, Lai HM, Moseley I, Sabouraud O (1997) Therapeutic effect of mitoxantrone combined with methylprednisolone in multiple sclerosis: a randomised multicenter study of active disease using MRI and clinical criteria. J Neurol Neurosurg Psychiatry 62: 112–118
16. Feinstein AR (1985) Statistical indexes of association. In Feinstein AR (ed) Clinical epidemiology: the architecture of clinical research. WB Saunders Co, Philadelphia
17. Fleiss JL, Schrout PE (1978) Approximate interval estimation for a certain intraclass correlation coefficient. Psychometrika 43:259–262
18. Francis DA, Bain P, Swan AV, Hughes RAC (1991) An assessment of disability rating scales used in multiple sclerosis. Arch Neurol 48:299–301
19. Gayou A, Brochet B, Dousset V (1997) Transitional progressive multiple sclerosis: a clinical and imaging study. J Neurol Neurosurg Psychiatry 63: 396–398
20. Goodkin DE, Cookfair D, Wende K, Bourdette D, Pullicino P, Scherokman B, Whitahm R (1992) 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 42:859–863
21. Hobart J, Lamping DL, Thompson A (1996) Evaluating neurological outcome measures: the bare essentials. J Neurol Neurosurg Psychiatry 60: 127–130
22. Hobart J, Lamping D, Fitzpatrick R, Riazi A, Thompson A (2001) The Multiple Sclerosis Impact Scale (MSIS-29). A new patient-based outcome measure. Brain 124:962–973
23. Kremenchutzky M, Cottrel DA, Rice GPA, Hader W, Baskerville J, Koopman W, Ebers GC (1999) The natural history of multiple sclerosis: a geographically based study. 7. Progressive-relapsing and relapsing-progressive multiple sclerosis: a re-evaluation. Brain 122:1941–1950
24. Kurtzke JF (1961) On the evaluation of disability in multiple sclerosis. Neurology 11(22):686–694
25. Kurtzke JF (1965) Further notes on disability evaluation in multiple sclerosis, with scale modifications. Neurology 15:654–661
26. Kurtzke JF (1983) Rating neurologic impairment in multiple sclerosis and expanded disability status scale (EDSS). Neurology 33:1444–1452
27. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:15–173
28. Lublin FD, Reingold SC, for the National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trial of New Agents in Multiple Sclerosis (1996) Defining the clinical course of multiple sclerosis: results of an international survey. Neurology 46:907–911
29. Noseworthy JH, Vandenvoort MK, Wong CJ, Ebers GC (1990) Interrater variability with the Expanded Disability Status Scale (EDS) and Functional Systems (FS) in a multiple sclerosis clinical trial. The Canadian Cooperation MS Study Group. Neurology 40: 971–975
30. Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, Ebers GC, Johnson KP, Sibley WA, Silberberg DH, Tourtellotte WW (1983) New diagnosis criteria for multiple sclerosis: guidelines for research protocols. Ann Neurol 13:227–231
31. Robertson NP, Clayton D, Fraser M, Deans J, Compston DAS (1996) Clinical concordance in sibling pairs with multiple sclerosis. Neurology 47:347–352
32. Rudick R, Antel J, Confavreux C, Cutter G, Ellison G, Fischer J, Lublin F, Miller A, Petkan J, Rao S, Reingold S, Syndulko K, Thompson A, Wallenberg J, Weinshenker B, Willoughby E (1996) Clinical outcomes assessment in multiple sclerosis. Ann Neurol 40:469–479
33. Rudick R, Antel J, Confavreux C, Cutter G, Ellison G, Fischer J, Lublin F, Miller A, Petkan J, Rao S, Reingold S, Syndulko K, Thompson A, Wallenberg J, Weinshenker B, Willoughby E (1997) Recommendations from the National Multiple Sclerosis Society clinical outcomes assessment task force. Ann Neurol 42:379–382
34. Schumacher GA, Beebe G, Kibler RF, et al. (1983) Problems of experimental trials of therapy in MS: report by the panel on the evaluation of experimental trials of therapy in MS. Ann Neurol 13:227–231
35. Sharrack B, Hughes RAC, Soudain S, Dunn G (1999) The psychometric properties of clinical rating scales used in multiple sclerosis. Brain 122: 141–159
36. Totaro R, Marini C, Cialfi A, Giunta M, Carolei A (2000) Prevalence of multiple sclerosis in the L'Aquila district, Central Italy. J Neurol Neurosurg Psychiatry 68:349–352
37. Trojano M, Avolio C, Manzari C, De Robertis F, Serio G, Livrea P (1995) Multivariate analysis of predictive factors of multiple sclerosis with a validate method to assess clinical events. J Neurol Neurosurg Psychiatry 58: 300–306
38. Trojano M, Avolio C, Simone IL, Defazio G, Manzari C, De Robertis F, Calo A, Livrea P (1996) Soluble intercellular adhesion molecule-1 in serum and cerebrospinal fluid of clinically active relapsing-remitting multiple sclerosis: Correlation with Gd-DTPA magnetic resonance imaging enhancement and cerebrospinal findings. Neurology 47:1535–1541
39. Trojano M, Liguori M, Bosco Zimatore G, Bugarini R, Avolio C, Paolicelli D, Giuliani F, De Robertis F, Marrosu MG, Livrea P (2002) Age-related disability in multiple sclerosis. Ann Neurol 51(4):475–480
40. Tukey JW (1977) Tukey exploratory data analysis. In: Reading, MA: Addison, Weseley

41. Weinshenker BG, Bass B, Rice GPA, Noseworthy JH, Carriere W, Baskerville J, Ebers GC (1989) The natural history of multiple sclerosis: a geographically based study. I. Clinical course and disability. Brain 112: 133–146
42. Weinshenker BG, Bass B, Rice GPA, Noseworthy JH, Carriere W, Baskerville J, Ebers GC (1989) The natural history of multiple sclerosis: a geographically based study. II. Predictive value of the early clinical course. Brain 112:1419–1428
43. Weinshenker BG, Rice GPA, Noseworthy JH, Carriere W, Baskerville J, Ebers GC (1991) The natural history of multiple sclerosis: a geographically based study. III. Multivariate analysis of predictive factors and models of outcome. Brain 114:1045–1056
44. Willoughby EW, Paty DW (1988) Scales for rating impairment in multiple sclerosis: a critique. Neurology 38: 1793–1798