ARTICLE *Multiple Sclerosis* 2006; **12**: 782–786

How similar are commonly combined criteria for EDSS progression in multiple sclerosis?

JJ Kragt[1], JM Nielsen[1], FAH van der Linden[1,2], BMJ Uitdehaag[1,3] and CH Polman[1]

Introduction Measuring disease progression is an important aspect of multiple sclerosis (MS) clinical trials. Commonly applied disability endpoints include time to clinically meaningful Expanded Disability Status Scale (EDSS) change, or the number of patients in whom such a change has occurred. Typically, clinically meaningful EDSS change has been defined as a change of 1.0 point on Kurtzke's EDSS in patients with an entry EDSS score of 5.5 or lower, or 0.5 point in patients with a higher EDSS score. Our goal was to evaluate whether these changes can be considered as similar. Therefore, we compared EDSS changes to corresponding changes in the Guy's Neurological Disability Scale (GNDS), which is a measure of patient perceived disability, and the Multiple Sclerosis Functional Composite (MSFC), which is an examination based quantitative scoring of neurological impairment.

Methods From a large longitudinal database, we selected two groups of patients with a clinically meaningful change in EDSS score according to the usual criteria: patients with EDSS change $\geq$1.0 for baseline EDSS $\leq$5.5 and patients with EDSS change $\geq$0.5 for baseline EDSS $\geq$6.0. We compared changes in GNDS sum score and in MSFC score between both groups.

Results In the group with baseline EDSS $\geq$6.0, GNDS and MSFC changes were higher than in patients with baseline EDSS $\leq$5.5. The difference in change was 1.00 (95% confidence interval (CI): −0.35 to 2.36) for the GNDS and 0.412 (95% CI: 0.300–0.525) for the MSFC.

Conclusion Our results indicate that a 0.5 point EDSS change in patients with baseline EDSS $\geq$6.0 cannot be considered equal to a 1.0 point change in patients with baseline EDSS $\leq$5.5. *Multiple Sclerosis* 2006; **12**: 782–786. http://msj.sagepub.com

Key words: clinical scale; EDSS; GNDS; MSFC; multiple sclerosis

Introduction

Measurement of disability is indispensable in assessing the efficacy of experimental therapeutic agents in multiple sclerosis (MS). Clinical scales are being used as primary or secondary outcome measures for recording disease progression in clinical trials. Despite several methodological limitations, the Expanded Disability Status Scale (EDSS) [1], is still used as a gold standard for measuring impairment and disability in MS.

Sensitivity to detect disease progression, also called responsiveness, is a key attribute for any assessment tool in MS clinical trials [2]. Due to its ordinal and non-continuous nature, the mean change in EDSS is an inappropriate endpoint [3], and therefore, a definition of treatment failure based on change in score from baseline has been introduced – a change in EDSS beyond a certain cut-off that is considered to be relevant and sustained during two consecutive examinations or for a certain length of time [4].

In previous trials [5,6], an EDSS change of 1.0 point sustained for three or more months has been considered as clinically meaningful for patients with a baseline EDSS score of $<$6.0. For patients with a higher EDSS score, a clinically meaningful change has been defined as a 0.5 point EDSS change. This guideline [7] has been defined as a logical consequence of the following characteristics

[1] Department of Neurology, VU University Medical Centre, Amsterdam, the Netherlands
[2] Department of Medical Psychology, VU University Medical Centre, Amsterdam, the Netherlands
[3] Department of Clinical Epidemiology and Biostatistics, VU University Medical Centre, Amsterdam, the Netherlands
Author for correspondence: JJ Kragt, Department of Neurology, VU University Medical Centre, De Boelelaan 1117, PO Box 7057, 1007 MB Amsterdam, the Netherlands. E mail: jj.kragt@vumc.nl
Received 22 September 2005; accepted 24 February 2006

of EDSS: differing staying times at specific EDSS levels and varying reproducibility of EDSS throughout its range.

First, the definition of a clinically meaningful change on EDSS should take into account the variable staying times at specific EDSS levels [8]. The mean staying times proved to be greatest at Disability Status Scale (DSS; as EDSS was formerly called) 1 and 7 and least for DSS 4 and 5. Second, several studies have been performed to assess intra- and inter-rater agreement of the EDSS [9–12]. In these studies, greater variability was observed in the lower part of the scale, showing that agreement depends on baseline EDSS and on definitions of agreement expressed by difference in EDSS scores.

In most clinical trials, patients showing a change of at least 1.0 point in the lower range of EDSS or at least 0.5 point in the upper EDSS range are combined and reported as a single number, thus implicitly assuming that these changes are equal or at least similar. The purpose of the present study was to evaluate whether indeed these changes can be considered similar. Therefore, we selected from a large longitudinal database those patients in whom such EDSS changes, according to one of both criteria, had occurred and compared, between the groups, corresponding changes in two external standards, the Guy's Neurological Disability Scale (GNDS), a measure of patient perceived disability [13], and the Multiple Sclerosis Functional Composite (MSFC), an examination-based quantitative scoring of neurological impairment [14].

Methods

Patients and test procedures

The database consisted of 662 patients with MS [15], who had undergone a series of test procedures as part of a health status assessment program designed to improve individual patient care at the MS Center of the VU University Medical Centre. No criteria for age, gender, disability level or MS subtype were applied during selection of data for analysis.

Patients were first selected on the basis of the availability of repeated EDSS, GNDS and MSFC examinations, with a time interval of at least 265 days. These examinations had been performed during the same visit under carefully standardized conditions by well-trained medical doctors, as described previously [16,17]. To standardize neurological examination as much as possible, we made use of the Neurostatus (Version 2, CD-ROM). The GNDS, which is a patient-based interview that captures the major domains of disabilities in MS, contains 12 subcategories that were scored and summed to create the GNDS sum score, ranging from 0 to 60 [13].

The MSFC consists of three quantitative measures: the Timed 25-foot Walk (T25FW) to assess lower limb disability, the 9-hole Peg Test (9HPT), a measure of upper limb function, and the Paced Auditory Serial Addition Test (PASAT) which estimates cognitive disability. The quantitative results of the three tests are combined into a composite which makes it a sensitive instrument that is able to detect small clinical changes. For creating the MSFC score, Z-scores were calculated for the T25FW, 9HPT and PASAT [14]. Z-scores were obtained using means and standard deviations of an external reference population, consisting of a wide range of MS patients [17]. As the Z-score sign had to be the same for all three tests, the mean of the 9HPT was transformed to its inverse before creating the Z-score and the Z-score of the T25FT was multiplied by -1. The composite score was calculated by adding the three Z-scores and dividing it by three: $\mathrm{MSFC} = (Z_{[1/(9\text{-HPT}),\ \text{average}]} - Z_{[\text{T25FT}]} + Z_{[\text{PASAT}]})/3$ [18]. Inability to perform a test of the MSFC due to MS-related symptoms, was scored with the maximum time allowed for the T25FT (180 seconds) and 9HPT (300 seconds) and with the worst score for the PASAT (0) [17,19].

Analysis

Results were analysed in several ways. First, we selected those patients with a predefined clinically meaningful change in EDSS score according to the usual definition: EDSS change ≥ 0.5 for baseline EDSS ≥ 6.0 (subset A) and EDSS change ≥ 1.0 for baseline EDSS ≤ 5.5 (subset B). After selection, we compared changes in GNDS sum score and MSFC score in subset A to those in subset B. Second, we further subdivided subset B in two groups according to baseline EDSS, which resulted in the formation of three disability strata: EDSS ≥ 6.0 (subset A), EDSS 4.0–5.5 (subset B1) and EDSS 0–3.5 (subset B2), and analysed GNDS and MSFC changes in these three groups. Finally, we explored newly defined EDSS changes in subset A and subset B that would give rise to more or less equal GNDS and MSFC changes.

Statistics

We evaluated differences in GNDS and MSFC changes between groups using Student's t-tests. We calculated point estimates of these differences in change with corresponding 95% confidence intervals (CI). To correct for multiple comparisons, we considered P values <0.01 as significant.

Table 1 Descriptives

	Total group	Subset A (EDSS ≥6.0)	Subset B (EDSS ≤5.5)	Subset B1 (EDSS 4.0–5.5)	Subset B2 (EDSS ≤3.5)
n	282	70	212	51	161
Female/male	180/102	37/33	143/69	30/21	113/48
RR/SP/PP	155/62/45	12/38/19	143/25/25	20/18/12	123/7/13
Age, years (SD)	41.9 (10.2)	45.2 (11.1)	40.8 (9.7)	46.8 (8.6)	39.0 (9.3)
Follow up duration, days (range)	691 (268–1715)	716 (296–1715)	682 (268–1669)	779 (308–1631)	651 (268–1669)

Results

We selected 606 patients (from a database containing 662 patients) who were examined twice with a time interval of at least 265 days. A total of 282 patients experienced a clinically meaningful change in EDSS score, as defined previously. Of these patients, 102 (36%) were male and 180 (64%) female. Mean age at baseline was 41.9 years (standard deviation (SD) 10.2). Most patients were diagnosed as having relapsing-remitting (RR) MS (55%), smaller proportions as having secondary progressive (SP) MS (22%) and primary progressive (PP) MS (16%) [20]. Average time from baseline to follow-up measurement was 691 days, range 268–1715 days.

Of these 282 patients, 70 were severely disabled (subset A). Mean follow-up duration in this subgroup was 716 days (range: 296–1715) and median EDSS change was 0.5 (range: 0.5–2.0). Mean GDNS sum score was 20.48 (SD: 6.78) at baseline and 23.62 (SD: 7.00) at follow-up, resulting in a GNDS change of 3.14 (SD: 4.97). MSFC scores in this group were −0.611 (SD: 0.818) at baseline and −1.128 (SD: 0.865) at follow-up. Therefore, MSFC change measured −0.517 (SD: 0.655). These calculations were also performed in the subgroup of 212 patients who were mildly or moderately disabled (subset B). This subgroup experienced a median EDSS change of 1.5 (range: 1.0–4.0) after an average follow-up duration of 682 days (range: 268–1669). Mean GNDS sum score in this group was 10.83 (SD: 6.32) at baseline and 12.97 (SD: 7.05) at follow-up, which resulted in a GNDS change of 2.14 (SD: 4.96). MSFC score at baseline was 0.462 (SD: 0.435) and at follow-up 0.357 (SD: 0.494), which gave rise to a MSFC change of −0.105 (SD: 0.251). When comparing GNDS changes in both groups, a non-significant trend towards a higher GNDS change in subset A was observed; the difference in change was 1.00 (95% CI: −0.35 to 2.36). Regarding the MSFC, we found a difference in change of 0.412 (95% CI: 0.300–0.525).

The exact same analyses were performed after subdividing subset B in subsets B1 and B2. Differences in changes between subsets B1 and B2 were not significantly different, neither for GNDS nor for MSFC. Descriptives and scores on GNDS and MSFC are shown in more detail in Tables 1 and 2.

Finally, we investigated different cut-off values for EDSS changes that would result in comparable changes in GNDS and MSFC in subsets A and B. For this, we used the change in subset A as reference

Table 2 GNDS and MSFC scores in the different subsets

	Subset A (EDSS ≥6.0)	A versus B	Subset B (EDSS ≤5.5)	Subset B1 (EDSS 4.0–5.5)	B1 versus B2	Subset B2 (EDSS ≤3.5)
Median change in EDSS (range)	0.5 (0.5–2.0)		1.5 (1.0–4.0)	1.5 (1.0–3.5)		1.5 (1.0–4.0)
Baseline GNDS (SD)	20.48 (6.78)		10.83 (6.32)	16.63 (5.78)		8.98 (5.29)
Follow up GNDS (SD)	23.62 (7.00)		12.97 (7.05)	18.94 (6.49)		11.07 (6.10)
Change in GNDS (SD)	3.14 (4.97)		2.14 (4.96)	2.31 (5.33)		2.09 (4.86)
Difference in GNDS change (95% CI)		1.00 (−0.35 to 2.36)			0.23 (−1.80 to 1.35)	
Baseline MSFC (SD)	−0.611 (0.818)		0.462 (0.435)	0.096 (0.436)		0.570 (0.372)
Follow up MSFC (SD)	−1.128 (0.865)		0.357 (0.494)	−0.010 (0.541)		0.466 (0.422)
Change in MSFC (SD)	−0.517 (0.655)		−0.105 (0.251)	−0.107 (0.298)		−0.104 (0.237)
Difference in MSFC change (95% CI)		0.412 (0.300–0.525)			0.003 (−0.083 to 0.088)	

EDSS, Expanded Disability Status Scale; MSFC, Multiple Sclerosis Functional Composite; GNDS, Guy's Neurological Disability Scale.

category, since 0.5 is the minimal detectable EDSS change. Results are shown in Table 3. An EDSS change of $\geq$**2.0** in patients with EDSS $\leq$5.5 resulted in a GNDS change that was roughly the same as that in patients with EDSS $\geq$6.0 who experienced an EDSS change of $\geq$0.5 (GNDS change 4.00 versus 3.14). To level out MSFC changes, a higher EDSS change was needed: $\geq$ **2.5** in patients with EDSS $\leq$5.5 was more or less equal to an EDSS change $\geq$0.5 in patients with EDSS $\geq$6.0 (MSFC change $-$0.407 versus $-$0.517). When evaluating different cut-off values for EDSS changes in the three subsets A, B1 and B2, similar results were obtained (data not shown).

Discussion

The exact definition of a clinically meaningful change in EDSS score is of great importance, especially in MS clinical trials. In most trials, this change has been defined as either a 1.0 point change for patients with an EDSS score at study entry of 5.5 or lower or a change of 0.5 point for more disabled patients (EDSS $\geq$6.0) and statistical analysis plans typically combine these patient groups. Nonetheless, the assumption that these two changes have equal clinical impact has never been properly examined. By using other clinical measurements, both subjective and objective, we were able to compare the changes associated with a 1.0 point and a 0.5 point change on EDSS, respectively, in patients with varying degrees of disability. We found that concomitant GNDS and MSFC changes were considerably higher in patients with severe disability (EDSS $\geq$6.0) who experienced an EDSS change $\geq$0.5 point compared to GNDS and MSFC changes in patients with mild and moderate disability who had at least 1.0 point EDSS change.

In order to compare the aforementioned EDSS changes, we used two other clinical measurements. First, we assessed the subjective clinical impact of EDSS changes by analysing GNDS changes. This instrument has been highly estimated as a measurement to assess patient perceived disability in different functional domains of MS. Second, we compared these EDSS changes in a more quantitative and objective manner by making use of the MSFC. Consequently, both external scales providing similar results, we show that these predefined EDSS changes cannot be considered as equal.

Obviously, the two external measurements have their limitations. Because the GNDS is a more subjective measurement that incorporates the patient's perspective, dissimilarities between GNDS and EDSS changes are a natural consequence. This can be illustrated by the fact that a number of patients experienced a clinically meaningful EDSS worsening, but did *not* deteriorate on GNDS (28% of patients in subset A, 37% of subset B). Concerning MSFC, the clinical meaningfulness of MSFC changes still needs to be clarified. Since this instrument has only recently been implemented in clinical trials, further research is warranted on this topic. Moreover, an explanation for the larger MSFC changes found in patients with severe disability could lie in the fact that these patients were unable to perform one or more of the tests at their follow-up visit and, therefore, were assigned the maximum time or worst score (eg, 0 for the PASAT). Finally, MSFC and EDSS possibly measure different dimensions of disability. This can be demonstrated by the finding that a number of patients who worsened on EDSS did *not* worsen on MSFC (23% in subset A, 32% in subset B).

Another pitfall of this study was the fact that EDSS changes were not confirmed in any manner. In clinical trials, a sustained EDSS change is confirmed by repeated measurements after three or six months. In our study this was not the case.

When using GNDS and MSFC changes from our study to titrate EDSS changes to have a comparable impact on patients with severe versus mild or moderate disability, we found that a 0.5 point EDSS change in patients with a score of 6.0 or higher more or less corresponds to a 2.0 or 2.5 point change in patients with a score of 5.5 or lower. We found no evidence that patients with EDSS of 5.5 or lower should further be separated in mild versus moderate disability in order to fine-tune the impact of EDSS changes.

These results support the need for careful reconsideration of the criteria for clinically meaningful EDSS changes, and the desirability of describing both groups separately in clinical trial reports.

Table 3 Different cut off values for EDSS changes and associated GNDS and MSFC changes

Baseline EDSS score range	Cut off EDSS change	Associated GNDS change	Associated MSFC change
$\geq$6.0	0.5	3.14	$-$0.517
$\leq$5.5	1.0	2.14	$-$0.105
	1.5	2.69	$-$0.117
	2.0	4.00	$-$0.144
	2.5	7.36	$-$0.407
	3.0	7.88	$-$0.501

Acknowledgements

The MS Centre of the VU University Medical Centre is partially funded by a program grant of the Dutch MS Research Foundation. No competing interests were declared.

References

1. **Kurtzke JF**. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983; **33**: 1444–52.
2. **Rudick R, Antel J, Confavreux C, Cutter G, Ellison G, Fischer J** *et al*. Clinical outcomes assessment in multiple sclerosis. *Ann Neurol* 1996; **40**: 469–79.
3. **Wingerchuk DM, Noseworthy JH, Weinshenker BG**. Clinical outcome measures and rating scales in multiple sclerosis trials. *Mayo Clin Proc* 1997; **72**: 1070–79.
4. **Weinshenker BG, Issa M, Baskerville J**. Meta analysis of the placebo treated groups in clinical trials of progressive MS. *Neurology* 1996; **46**: 1613–19.
5. **European Study Group on interferon beta 1b in secondary progressive MS**. Placebo controlled multicentre randomised trial of interferon beta 1b in treatment of secondary progressive multiple sclerosis. *Lancet* 1998; **352**: 1491–97.
6. Randomized controlled trial of interferon beta 1a in secondary progressive MS: clinical results. *Neurology* 2001; **56**: 1496–504.
7. **Goodkin DE**. EDSS reliability. *Neurology* 1991; **41**: 332.
8. **Weinshenker BG, Rice GP, Noseworthy JH, Carriere W, Baskerville J, Ebers GC**. The natural history of multiple sclerosis: a geographically based study. 4. Applications to planning and interpretation of clinical therapeutic trials. *Brain* 1991; **114**: 1057–67.
9. **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**: 746–48.
10. **Francis DA, Bain P, Swan AV, Hughes RA**. An assessment of disability rating scales used in multiple sclerosis. *Arch Neurol* 1991; **48**: 299–301.
11. **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.
12. **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**: 971–75.
13. **Sharrack B, Hughes RA**. The Guy's Neurological Disability Scale (GNDS): a new disability measure for multiple sclerosis. *Mult Scler* 1999; **5**: 223–33.
14. **Cutter GR, Baier ML, Rudick RA, Cookfair DL, Fischer JS, Petkau J** *et al*. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain* 1999; **122**: 871–82.
15. **Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, Ebers GC** *et al*. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann Neurol* 1983; **13**: 227–31.
16. **Hoogervorst EL, van Winsen LM, Eikelenboom MJ, Kalkers NF, Uitdehaag BM, Polman CH**. Comparisons of patient self report, neurologic examination, and functional impairment in MS. *Neurology* 2001; **56**: 934–37.
17. **Kalkers NF, De Groof V, Lazeron RH, Killestein J, Ader HJ, Barkhof F** *et al*. MS functional composite: relation to disease phenotype and disability strata. *Neurology* 2000; **54**: 1233–39.
18. **Fischer JS, Jak AJ, Kniker JE, Rudick RA, Cutter G**. *Administration and scoring manual for the Multiple Sclerosis Functional Composite (MSFC)*. New York, 2001.
19. **Hoogervorst EL, Kalkers NF, Cutter GR, Uitdehaag BM, Polman CH**. The patient's perception of a (reliable) change in the Multiple Sclerosis Functional Composite. *Mult Scler* 2004; **10**: 55–60.
20. **Lublin FD, Reingold SC**. Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. *Neurology* 1996; **46**: 907–11.