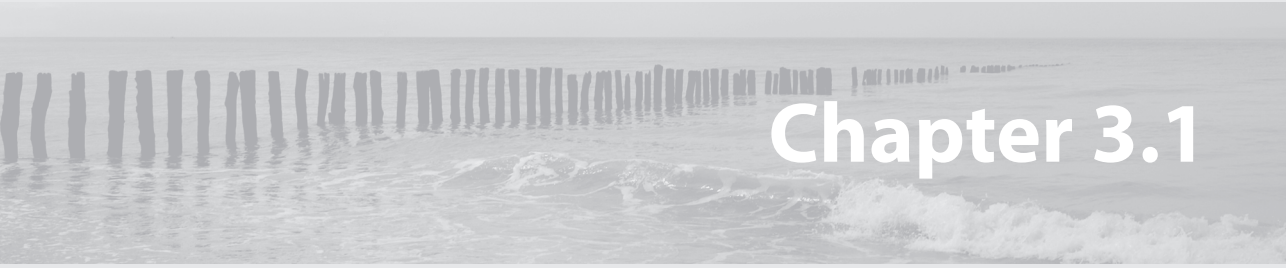

Chapter 3.1

Clinical scales in progressive MS: predicting long-term disability

LVAE Bosma, JJ Kragt, DL Knol, CH Polman, BMJ Uitdehaag

Multiple Sclerosis 2012;18:345-350

ABSTRACT

Objective: To determine which short-term changes on clinical scales including the Expanded Disability Status Scale (EDSS), Timed 25-Foot Walk (T25FW), 9-Hole Peg test (9HPT) and Guy's Neurological Disability Scale (GNDS) are most predictive of long-term outcome of disability as rated by the EDSS in progressive MS.

Methods: From a longitudinal database, all progressive patients, both primary (PP) and secondary (SP), were selected on the basis of at least 2 complete examinations being available within a time interval of 1-2 years (short-term change). All patients who fulfilled the selection criteria were invited for a third visit after an interval of at least 3 years (long-term outcome). We used ordinal logistic regression to see which early changes were most predictive of the long-term EDSS.

Results: 181 patients fulfilled the selection criteria. Early change on EDSS and T25FW were the best predictors of long-term EDSS; both were significant predictors in a 'single predictor' model. Early EDSS change was a slightly stronger single predictor (R^2=0.38, Wald χ^2=42.65, p<0.001) compared to early T25FW change (R^2=0.27, Wald χ^2=12.35, p<0.001). Adding early T25FW change to early EDSS change in a 'combined predictor' model improved prediction (p=0.036).

Conclusion: Both early change on EDSS and T25FW predict long-term EDSS with comparable strength. Early change on T25FW adds significant independent information and improves the prediction model with early EDSS change only. Therefore we support the use of early T25FW examinations in future clinical trials in progressive MS.

INTRODUCTION

Disease modifying drugs have shown effectiveness in the treatment of multiple sclerosis (MS), that is in preventing the occurrence of clinical relapses and radiologically visible disease activity. Yet, there is an urgent need for therapies targeting disease progression, as we still lack treatment options to modify the disease course in the primary or secondary progressive phases of MS. About 15-20 % of all MS patients have a course that is progressive from onset,[1,2] and of all patients with a relapsing onset around 80% exhibit secondary progressive disease after 20 years.[3] To facilitate trial design in (progressive) MS, predicting future clinically important disability by means of short-term changes on different clinical scales is crucial.[4]

Therefore, in this study we selected a group of patients with progressive MS, either primary or secondary, and studied the disease course as documented with clinical scales commonly used in MS, ranging from objective (neurological examination and walking distance: the Expanded Disability Status Scale [EDSS], leg function and arm dexterity: the Timed 25-Foot Walk [T25FW] and 9-Hole Peg Test [9HPT]) to more subjective (the Guy's Neurological Disability Scale [GNDS], a questionnaire driven by patient interview to assess disability) measures.[5-7]

The EDSS has well known psychometric disadvantages,[8] is quite time-consuming to perform and requires a great deal of experience in neurology and MS in particular. However, it is the widely known and accepted outcome measure for the rating of disability in MS. Therefore in this study the EDSS was used as long-term disability outcome. Apart from the EDSS, we explored short term changes on the other more easy to obtain assessments (T25FW, 9HPT and GNDS) as potential predictors of long-term disability.

This study aimed to explore which short-term changes on these clinical scales either alone or in combination were most predictive of the long-term outcome of disability as rated by the EDSS in progressive MS.

METHODS

Patients

From a large cohort of MS patients who were prospectively followed up by a series of test procedures as part of a health status program designed to improve individual patient care at the MS Center of the VU University Medical Center, we selected all progressive patients (primary progressive [PP], and secondary progressive [SP]).[9] Patients were denominated progressive by their treating neurologist and were diagnosed as having MS as ascertained by Poser et al.[10]

or revised McDonald criteria.[11] Once being denominated 'progressive' by the neurologist, we did not adjust the subtype of secondary progressive MS patients when they turned out to be clinically stable over the complete follow up duration according to the EDSS score. The study was performed with the approval of the medics ethics committee of the VU University Medical Center, and informed consent was obtained from all participants.

No criteria for age, gender or disability level were applied during selection of data for the present analyses. However, apart from having a progressive MS subtype, patients were selected on the basis of repeated EDSS, T25FW, 9HPT and GNDS examinations being available within a time interval of 1-2 years (short-term change, change during the first time interval, also mentioned as 'early change'). We invited all patients who fulfilled the selection criteria of two visits back for a third visit consisting of another series of clinical test procedures (EDSS, T25FW, 9HPT and GNDS) after a time interval from the second examination of at least three years (in order to look at the long-term change, change during the second time interval). For both intervals we used a time window of two months (the first interval duration was at least 10 months and maximal 26 months, and the second time interval was at least 34 months).

The complete cohort comprised 1278 MS patients, of whom 916 patients were excluded because of their subtype (relapsing remitting [RR], clinically isolated syndrome [CIS] or unknown). Of the remaining 362 progressive patients 76 patients were excluded because they had only one single visit. Two-hundred and eighty-six patients remained with two or more visits, of whom 214 patients had complete visits with all required examinations. Of those 214 patients, 181 patients could be contacted and were willing to come for another, third visit (13 patients had deceased, 9 were lost to follow-up and 11 were not willing to participate further). The disease characteristics of the progressive patients that were ultimately selected for this study (gender, mean age and mean disease duration at the first visit) did not differ from the characteristics of the progressive patients that were excluded.

Test procedures

All assessments of EDSS, T25FW, 9HPT and GNDS were obtained as part of routine outpatient care, during regular visits and under standardized conditions (by well trained examiners). In the SPMS patients, none of the assessments were obtained during clinical relapse. If patients were unable or not willing to come to the hospital for the third visit, assessment was performed by telephone for the EDSS[12] and the GNDS.[7] In these cases the T25FW and 9HPT were missing. If patients were unable to perform either the T25FW or 9HPT due to MS-related symptoms, the maximum allowed time for this test was assigned (180 seconds for T25FW and 300 seconds

for 9HPT). Each subcategory of the GNDS was scored separately ranging from 0 (normal) to 5 (maximum help required). If a patient refused to answer on the interview section of the subcategory sexual function, the round off mean score of the subcategories lower limb function, bladder function and bowel function was used as the sexual function score. The 12 subcategories of the GNDS were scored and summed to create the GNDS sum score, ranging from 0 to 60.[7]

Analyses and statistics

Differences (short-term and long-term) in all test scores were calculated and annualized to account for differences in duration of the intervals. EDSS scores at the third visit were converted into seven groups of roughly equal size. We used ordinal logistic regression with short-term changes as predictors and EDSS scores at the third visit (EDSS M3) as the outcome (dependent variable) to see which early changes were most predictive of the long-term EDSS. As long-term EDSS scores are evidently correlated with baseline EDSS level and duration of the investigated time interval, baseline EDSS (EDSS M1) and the duration of the second (long-term) time-interval (I 2) were included as variables in all models. As the short-time changes are annualized, there is no need to correct for the duration of the first time-interval. To point out which changes are most predictive, we compared the R^2, Wald χ^2 and p-values of different models with early changes on different scales as predictors. First, we compared different early changes as single predictors, to see which was the strongest single predictor. Then we tried to improve the prediction model with the strongest single predictor by combining it with early changes on other scales. Finally, we left out the strongest single predictor and combined two other predictors to see whether together they could replace the strongest single predictor.

All statistical analyses were performed using the Statistical Package for Social Sciences (SPSS) version 15.0 (SPSS, Inc., Chicago, IL, USA).

RESULTS

Descriptives of the study population

A total of 181 patients fulfilled the selection criteria, of which 85 (47%) were primary progressive and 96 (53%) secondary progressive. The mean durations of the first and second intervals was 1.3 and 6.2 years (mean total follow-up duration 7.5 years). At the time of the first visit the mean disease duration was 10.8 years and the mean age was 49.6 years. **Table 3.1.1** shows demographics and disease characteristics at baseline (first visit).

Table 3.1.1 Patient characteristics at baseline (n=181)

Type MS	
SP	96 (53%)
PP	85 (47%)
Gender	
Female	102 (56%)
Male	79 (44%)
Disease duration, mean (SD), y	10.8 (8.0)
Age, mean (SD), y	49.6 (9.3)

MS: multiple sclerosis, PP: primary progressive, SP: secondary progressive, y: years.

Of those 181 patients fulfilling the selection criteria, 120 were willing and able to come to the hospital for a (face to face) third visit. During the follow-up interval between the second and third visit (6 years on average), seven patients had died due to MS (EDSS 10), meaning that in this progressive MS cohort approximately 0.6% of the patients died due to MS each year. The average time to death from onset of MS among these patients was 19 years. The mean age of the patients at the time of death due to MS was 58 years. Another 54 patients were unable or not willing to come back to the hospital for a third visit, but they were assessed by telephone.

Missing data

Because patients were selected on the basis of two complete examinations of EDSS, T25FW, 9HPT and GNDS being available, during the first two visits there were no missing data. During the third visit, obviously, in the patients who had died (4%) the T25FW, 9HPT and GNDS were missing. In those in whom the visit took place as a telephone interview (30%), only the T25FW and 9HPT were missing; EDSS and GNDS were assessed by telephone. (Exception: when it was clear from the telephone interview that the patient could not at all walk anymore, for the T25FW the maximum allowed time was assigned just as at the face-to-face visit, so that this relevant score could be included in the analyses.)

Disease Modifying Therapy (DMT)

Only 16 patients (9%) were exposed to a Disease Modifying Therapy during the third visit: 13 patients were treated with interferon beta, two were treated with methotrexate and one with glatiramer acetate. Sixty-one patients (34%) had ever used DMT (interferon beta, glatiramer acetate or methotrexate). Given the lack of proven effective therapies for progressive MS and

the relatively small number of patients exposed to immunomodulatory treatment during the last visit, we did not explore the influence of treatment in this study.

Benign PPMS

Two out of 85 PPMS patients (2%) had an EDSS score of 3.0 or lower after a disease duration of at least 10 years and thus fulfilled the criteria for 'benign MS'.

Clinical scores at all visits

Table 3.1.2 shows mean scores on the various clinical scales during the three visits.

When the mean change scores on the various scales during follow-up are annualized, one can see that during the second interval mean changes are smaller than during the first interval, with further in the disease changes per time unit being smaller on all scales (**Table 3.1.3**). Comparing PP- and SPMS patients: On the EDSS and T25FW mean annualized change scores during both time intervals were quite similar (slightly higher for PPMS): on the EDSS annualized change during first interval was 0.3 for SPMS patients and 0.4 for PPMS patients, during second interval annualized change was 0.2 for both. On the T25FW annualized change during first interval

Table 3.1.2 Mean scores on the scales during the three visits (total n=181)

	M1		M2		M3	
EDSS						
median (IQR)	4.5 (4.0-6.0)	*n=181*	5.5 (4.0-6.5)	*n=181*	6.5 (6.0-7.5)	*n=181*
range	2.0-7.0		2.0-7.5		2.0-10.0	
T25FW						
mean (SD), s	10.7 (12.7)	*n=181*	26.4 (50.3)	*n=181*	84.1 (82.2)	*n=142*
median (IQR), s	7.1 (5.4-10.5)		8.3 (6.0-14.1)		24.8 (9.1-180.0)	
range, s	3.0-111.9		3.0-180.0		2.6-180.0	
9HPT						
mean (SD), s	29.6 (16.3)	*n=181*	35.3 (32.0)	*n=181*	51.8 (56.7)	*n=120*
median (IQR), s	24.9 (21.3-31.3)		25.4 (21.9-35.4)		27.9 (23.3-43.6)	
range, s	14.5-161.9		16.0-300.0		16.5-300.0	
GNDS						
mean (SD)	16.1 (6.7)	*n=181*	17.1 (6.4)	*n=181*	20.8 (8.0)	*n=174*
median (IQR)	16.0 (11.0-20.0)		17.0 (12.5-21.0)		20.0 (14.8-26.3)	
range	2-35		2-36		4-44	

EDSS: Expanded Disability Status Scale, GNDS: Guy's Neurological Disability Scale, 9HPT: 9-Hole Peg Test, M1: first visit, M2: second visit, M3: third visit, IQR: inter quartile range, s: seconds, T25FW: Timed 25-Foot Walk.

Table 3.1.3 Mean change scores during the first and second time interval

	I 1	I 1	I 2	I 2
Duration of interval, mean (SD), y	1.3 (0.4)	Annualized	6.2 (1.8)	Annualized
EDSS	0.5	0.4	1.2	0.2
T25FW, s	15.7	13.2	57.8	9.2
9HPT, s	5.6	4.8	21.5	3.7
GNDS	1.1	0.8	4.0	0.7

EDSS: Expanded Disability Status Scale, GNDS: Guy's Neurological Disability Scale, 9HPT: 9-Hole Peg test, I 1: first interval, I 2: second interval, s: seconds, T25FW: Timed 25-Foot Walk, y: years.

was 13.1 for SPMS patients and 13.4 for PPMS patients, during second interval annualized change was 9.0 for SPMS patients and 9.4 for PPMS patients. On the 9HPT progression rates were higher for SPMS than for PPMS patients (annualized change during first interval was 5.7 resp. 3.9 and during second interval 5.0 resp. 2.1) and during first interval on the GNDS progression rates were higher for SPMS than for PPMS (1.5 resp. -0.1; during second interval progression rates were equal). Of all scales, change on the T25FW was most distinct.

The EDSS scores during the third visit (EDSS M3) ranged from 2.0-10.0 (median 6.5, IQR 6.0-7.5). For further analyses (ordinal logistic regression), those EDSS scores were converted into seven groups of roughly equal size based on important functional milestones as is shown in **Table 3.1.4**.

Table 3.1.4 EDSS subgroups used with ordinal logistic regression

EDSS M3	n (%)	
EDSS < 4.5	22 (12)	ambulatory without aid or rest for ≥ 500 m
EDSS 4.5 + 5.0 + 5.5	20 (11)	walking distance without aid is limited, max. between 300-100 m
EDSS 6.0	32 (18)	unilateral assistance
EDSS 6.5	34 (19)	bilateral assistance
EDSS 7.0	25 (14)	in need of a wheelchair (unable to walk 5 m even with aid)
EDSS 7.5 + 8.0	31 (17)	restricted to wheelchair
EDSS > 8.0	17 (9)	restricted to bed

EDSS: Expanded Disability Status Scale, EDSS M3: EDSS at third visit.

Regression analyses results

As already mentioned in the *Methods* section, we used an ordinal logistic regression model with early changes on either EDSS, T25FW, 9HPT or GNDS as independent variables, baseline EDSS (EDSS M1) and duration of the second (long-term) time interval (I 2) always included in the model, and EDSS during third visit (EDSS M3) as the dependent variable. **Table 3.1.5** shows the results of the different models according to various early changes as predictors. The upper four rows show the four different early changes as predictors in a 'single predictor' model. However, the models always contained multiple variables as in all models baseline EDSS (EDSS M1) and duration of the second time interval (I 2) were included. The lower rows show combinations of two different early changes as predictors in a 'combined predictor' model. The pseudo R-square values refer to the complete models with early change, baseline EDSS and second time interval (df=3). The Wald χ^2 values and p-values refer to the parameter on the concerning line within the rows. They refer to the value of adding the early change parameter

Table 3.1.5 Different logistic regression models according to various early changes as predictors

Model	Pseudo R-square (R^2) Nagelkerke	χ^2 Wald*	Significance* (p)
early change EDSS	0.38	42.65	*<0.001*
early change T25FW	0.27	12.35	*<0.001*
early change 9HPT	0.22	2.73	0.099
early change GNDS	0.24	5.37	*0.020*
early change EDSS & early change T25FW	0.40	34.35 4.41	*<0.001* ***0.036***
early change EDSS & early change 9HPT	0.38	41.13 0.40	*<0.001* 0.529
early change EDSS & early change GNDS	0.39	39.78 2.45	*<0.001* 0.117
early change T25FW & early change 9HPT	0.28	11.33 1.70	*0.001* 0.192
early change T25FW & early change GNDS	0.29	10.92 4.01	*0.001* *0.045*
early change GNDS & early change 9HPT	0.25	5.03 2.29	*0.025* 0.131

Early change, change during the first interval. EDSS M1 (EDSS during first visit) and I 2 (second time interval) were included in all models.

* Wald and p-value refer to the value when adding the early change parameter to the model with only EDSS M1 and I 2.

to the 'fundamental' model with only baseline EDSS and second time interval and, in case of a 'combined predictor' model, the 'other' early change parameter (df=1).

When looking at the different prediction models presented, it is clear that early change on the EDSS is the strongest predictor (Wald χ^2=42.65, p<0.001) in a 'single predictor' model. Early T25FW change is second best, as it is quite a strong predictor as well (Wald χ^2=12.35, p<0.001) and forms a significant prediction model. Early 9HPT change does not form a significant prediction model in a 'single predictor' model (Wald χ^2=2.73, p=0.099), whereas early GNDS change does (Wald χ^2=5.37, p=0.020), but less significant than early T25FW change. When looking at the 'combined predictor' models, early T25FW change is the only predictor significantly improving prediction when added to early EDSS change (Wald χ^2=4.41, p=0.036). Early 9HPT change does not improve prediction when added to early EDSS change (Wald χ^2=0.40, p=0.529), nor does early GNDS change (at least not significantly; Wald χ^2=2.45, p=0.117). 'Combined predictor' models with early T25FW change and early 9HPT change (R^2=0.28) or with early T25FW change and early GNDS change (R^2=0.29) are not equivalent to the 'single predictor' model with early EDSS change only (R^2=0.38) and can therefore together not replace the strongest single predictor.

DISCUSSION

Our analyses clearly show that both early change on EDSS and T25FW are good predictors of long-term EDSS. Prediction based on early T25FW change is not identical to early EDSS change and therefore cannot replace the latter. Nor can the combinations of early T25FW change and early 9HPT change or early T25FW change and early GNDS change replace early EDSS change. However, early T25FW change adds significant independent information and improves the prediction model with EDSS change only. Early GNDS change is a significant predictor of long-term EDSS as well, but not as strong as early change on EDSS or T25FW. Early GNDS change does not add significant independent information to the prediction model with only early EDSS change. Early 9HPT change is not a significant predictor of long-term EDSS, nor does it improve the prediction model with only early EDSS change.

During the follow-up in our study, 4% of patients (seven patients) died due to MS. The mean time to death from onset of MS was 19 years, which is comparable to a natural history study of PPMS patients where the mean time to death due to MS was 22 years.[13] It can be stated that our population is quite similar to what has been described in earlier studies.

It is important to recognise some possible problems and limitations of our study. Most important, we were hampered by the EDSS being an ordinal scale, which makes it hard to use

as dependent variable. Because of that we had to use ordinal logistic regression and divide all EDSS scores during the third visit into several subgroups with a subsequent loss of information. Second, we cannot exclude that some learning curve in the Multiple Sclerosis Functional Composite (MSFC) might have played a role in our analyses, because not all patients underwent formal MSFC testing in advance of the scorings used for this study, i.e. for the 9HPT, as we left out the Paced Auditory Serial Addition Test (PASAT) already (and on the T25FW there is no known learning curve).[14,15] Third, the reported long-term changes were not confirmed by repeated measurements after e.g. 3 or 6 months.

In this study we compared different widely used measures to document progression in MS, in order to determine the characteristics of these rating scales, both individually and in combination, in terms of predicting long-term disease disability. This study is unique in that it provides longitudinal data of a variety of scales in 181 MS patients with progressive disease during 7.5 years of follow-up on average.

Despite the mentioned limitations our data convincingly show that short-term changes on EDSS and T25FW are good predictors of long-term EDSS. As long as the EDSS is accepted as the gold standard to rate clinical disability, it is therefore helpful to assess both EDSS and T25FW early in the progressive phase of the disease, as they both have predictive value. It would be important to replicate our results in independent cohorts. Another issue that remains to be studied is the way in which these results should be incorporated in trial design. How should EDSS and T25FW be combined and used in the definition of an endpoint in clinical studies? Are there specific optimal cut-off values to be defined or should continuous information be used? Further studies are needed to address these issues.

In conclusion based on our results, we recommend the use of the T25FW in addition to the traditionally used EDSS as outcome measure in clinical trials of progressive MS: it is not only easy to administer but it also adds independent prognostic information.

REFERENCES

1. Thompson AJ, Polman CH, Miller DH, et al. Primary progressive multiple sclerosis. Brain 1997; 120: 1085-1096.
2. Compston A, Coles A. Multiple sclerosis. Lancet 2002; 359: 1221-1231.
3. Kremenchutzky M, Rice GP, Baskerville J, Wingerchuk DM, Ebers GC. The natural history of multiple sclerosis: a geographically based study 9: observations on the progressive phase of the disease. Brain 2006; 129: 584-594.
4. Rudick RA, Kappos L. Measuring disability in relapsing-remitting MS. Neurology 2010; 75: 296-297.
5. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983; 33: 1444-1452.
6. 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.
7. Sharrack B, Hughes RA. The Guy's Neurological Disability Scale (GNDS): a new disability measure for multiple sclerosis. Mult Scler 1999; 5: 223-233.
8. Hobart J, Freeman J, Thompson A. Kurtzke scales revisited: the application of psychometric methods to clinical intuition. Brain 2000; 123: 1027-1040.
9. 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: 907-911.
10. Poser CM, Paty DW, Scheinberg L, et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. Ann Neurol 1983; 13: 227-231.
11. Polman CH, Reingold SC, Edan G, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". Ann Neurol 2005; 58: 840-846.
12. Lechner-Scott J, Kappos L, Hofman M, et al. Can the Expanded Disability Status Scale be assessed by telephone? Mult Scler 2003; 9: 154-159.
13. Cottrell DA, Kremenchutzky M, Rice GP, et al. The natural history of multiple sclerosis: a geographically based study. 5. The clinical features and natural history of primary progressive multiple sclerosis. Brain 1999; 122: 625-639.
14. Cohen JA, Cutter GR, Fischer JS, et al. Use of the multiple sclerosis functional composite as an outcome measure in a phase 3 clinical trial. Arch Neurol 2001; 58: 961-967.
15. Solari A, Radice D, Manneschi L, Motti L, Montanari E. The multiple sclerosis functional composite: different practice effects in the three test components. J Neurol Sci 2005; 228: 71-74.