2

# Measuring fatigue in patients with multiple sclerosis: reproducibility, responsiveness and concurrent validity of three Dutch self-report questionnaires

M.B. Rietberg
E.E.H. van Wegen
G. Kwakkel

Disabil Rehabil. 2010; 32(22): 1870-6.

## Abstract

**Purpose:** To determine the reproducibility, responsiveness and concurrent validity of Dutch versions of the Fatigue Severity Scale (FSS), Modified Fatigue Impact Scale (MFIS), and Checklist Individual Strength (CIS20R) in patients with Multiple Sclerosis (MS).

**Method:** Forthy-three ambulatory patients with MS (mean age 48.7 years; SD 7 years; 30 women; median EDSS score 3.5) completed the questionnaires twice within one week. The Intraclass Correlation Coefficients (ICCs), Bland and Altman analysis, the smallest detectable change (SDC) and the Minimal detectable change (MDC) were calculated. Concurrent validity was determined by Pearson's correlation coefficients.

**Results:** ICCs ranged from 0.76 (FSS), to 0.85 (MFIS) to 0.81 (CIS20R). Bland and Altman analysis showed no significant systematic differences between assessments. MDCs were 20.7% (FSS), 19.23% (MFIS), and 17.7% (CIS20R). Pearson correlation coefficients were r=0.66 (FSS-MFIS), r=0.54 (MFIS-CIS20R) and r=0.42 (CIS20R-FSS).

**Conclusion:** Despite good test-retest reliability of FSS, MFIS and the CIS20R, the present study shows that fatigue questionnaires are not very responsive for change in patients with MS. This finding suggests that future trials should monitor profiles of fatigue by repeated measurements rather than pre-post assessments alone. The moderate associations suggest that the three questionnaires largely measure different aspects of perceived fatigue.

## Introduction

Numerous studies report fatigue as the most common symptom in multiple sclerosis (MS).[1-6] Fatigue is reported by 65 to 95% of all MS patients, and between 15 and 60% of the patients report fatigue as the most disabling problem, severely limiting daily activities and having a major impact on quality of life.[1-4,7-10]

Although the exact etiology of fatigue in MS is unclear and consensus on defining fatigue is still lacking, proposed definitions[11,12] support the general clinical notion that MS-related fatigue is subjective and multidimensional in nature. The multidimensionality is believed to result from a complex interplay between underlying disease process,[12-14] psychological[15,16] and physical characteristics[14] as well as patients' environmental factors.[17,18] The multidimensionality of MS related fatigue is illustrated by the large number of questionnaires used in MS samples,[14] such as the Fatigue severity Scale (FSS),[19] the Fatigue Assessment Instrument (FAI),[20] the Fatigue Impact Scale (FIS),[21] the Checklist Individual Strength (CIS20R),[15] the Modified Fatigue Impact Scale (MFIS),[11] and the Fatigue Descriptive Scale (FDS).[22]

The multidimensionality of MS related fatigue is also manifested by the different conceptual approaches of measuring fatigue. For example, the FSS[19] assesses the severity of fatigue symptoms and its impact on an individual's daily functioning during the past week, whereas the MFIS assesses the perceived impact of fatigue on the domains physical, cognitive and psychosocial functioning during the past four weeks. The CIS20R[15] assesses four dimensions related to fatigue: subjective experience of fatigue; reduction in motivation; reduction in activity and reduction in concentration over the last two weeks.

Most approaches to fatigue assessment can be classified as either self-report scales or performance-based measures of motor or cognitive output.[6] The most commonly used method, and perchance the best way to quantify fatigue, in clinical practice and in research is the use of self-report questionnaires.[23,24] Of the above listed self-report instruments for assessing fatigue, the FSS is perhaps one of the most commonly used measures of fatigue severity in patients with MS.[23] The psychometric properties of the FSS have been evaluated in MS patients,[19] the FSS is easy to administer and has a high degree of validity and sensitivity to clinical changes.[19] The MFIS is recommended for clinical practice and research by the Multiple Sclerosis Council for Clinical guidelines.[11] Psychometrics have been evaluated in a Dutch version of the MFIS.[25] That study indicates that the Dutch version of the MFIS is a reliable, valid and responsive tool to assess the impact of MS-related fatigue on daily life.

The Checklist Individual Strength (CIS20R)[15] recognises the multidimensionality of fatigue in MS, but its use in MS research is limited until now. While norm scores for severe fatigue are available,[26] psychometric properties of the CIS20R, like reproducibility and concurrent validity with other commonly used scales in the MS population are lacking.

The aim of the present study was to determine the reproducibility, responsiveness and concurrent validity of the Dutch versions of the Fatigue Severity Scale (FSS), Modified Fatigue Impact Scale (MFIS) and Checklist Individual Strength (CIS20R) in patients with MS.

## Methods

### Subject selection

Patients suffering from MS were recruited from the MS center of the VU University Medical Center (VUmc), the Netherlands. Patients met the following inclusion criteria: (1) older than 18 years; (2) a definite diagnosis of MS;[27] (3) an Expanded Disability Status Scale (EDSS)[28] score below 6.5; (4) no co-morbidity that could influence fatigue; (5) written informed consent. All participants gave informed consent, in accordance with the ethical standards of the declaration of Helsinki. The local medical ethics committee approved the present study.

### Fatigue questionnaires

Three questionnaires, the FSS, the MFIS, and the CIS20R, were used to assess fatigue. The Fatigue Severity Scale (FSS)[19] is a nine-item self-report questionnaire to assess the severity of fatigue and its impact on an individual's daily functioning. Participants rate their agreement with a statement ranging from one point, reflecting 'strongly disagree' to seven points representing 'strongly agree', depending on how appropriate they feel the statement applies to them. A total sum score is calculated. Translation and back translation was performed by two independent linguists and evaluated by a panel of three clinical experts.

The Modified Fatigue Impact Scale (MFIS) is a shortened Dutch version of the 40–item Fatigue Impact Scale[3] and assesses the perceived impact of fatigue on the subscales physical, cognitive and psychosocial functioning during the past four weeks. Participants rate on a five point Likert scale, with 0 = 'Never' to 4 = 'Almost always', their agreement with 21

statements.[29] The items can be aggregated into a total MFIS score, as well as into a score for the three subscales.

The CIS20R[15] assesses fatigue during the past two weeks and consists of four dimensions: subjective experience of fatigue; reduction in motivation; reduction in activity and reduction in concentration. The CIS20R consists of twenty statements for which the participant has to indicate on a seven point scale ranging from 'Yes, that is true' to 'No, that is not true' to what extent the particular statement applies to him or her.[30] Subscores for the domains as well a total score are calculated.

## Design

Participants filled in the questionnaires twice with an interval of one week, during visits from an assessor in patients' own home. To prevent carry-over effects, three different questionnaire orders (MFIS/FSS/CIS20R, CIS20R/MFIS/FSS, and FSS/CIS20R/MFIS) were composed and given in a random order. In random order, half of the participants had both the test and retest assessments in the morning, the other half of the participants in the afternoon to control for influences of diurnal fluctuations in perception of fatigue. The participants were verbally instructed to read each statement carefully, and then circle the one number that best indicates their agreement. In case participants had difficulty with selecting an answer, they were told to choose the answer that comes closest to describing their perceived symptoms of fatigue. If the participant needed help in understanding words or phrases, or marking their responses, the assessor assisted.



## Statistical analysis

All data were analyzed with SPSS statistical package (version 15.0). First, descriptive statistics were used to determine mean age, gender, duration of disease and, type of MS. Next, associations of these characteristics with fatigue following the three questionnaires were explored, using Pearson correlation coefficients.

Reproducibility concerns the degree to which repeated measurements provide similar results.[30] Reproducibility was determined by calculating Intraclass Correlation Coefficients (ICCs) for test-retest reliability and by applying the Bland and Altman method for agreement between the two measurements.

## Test-retest reliability

Reliability was defined as how well the scores of the participants can be distinguished from each other on a fatigue questionnaire, despite existing measurement errors.[31] For the ICCs, a two-way random effects model was used assuming included patients and assessors are a random selection of both populations.

## Agreement

In addition we were interested in the absolute agreement between two consecutive assessments and therefore used the absolute agreement definition in the calculation of ICC.[32] In the present study, an ICC beyond 0.70 was defined as good reliability, an ICC between 0.40 and 0.70 as moderate reliability and an ICC below 0.40 as poor reliability. Agreement concerns the measurement error, and assesses how close derived scores on the fatigue questionnaires produces exactly the same outcome.[31] For this purpose, the Bland and Altman method was used by plotting the mean difference (Mean Δ) between the two consecutive measurements against the standard deviation (SD) of this difference.[33] The 'limits of agreement' were calculated as the mean difference ± 1.96 times the standard deviation of the differences.

## Responsiveness

Bland and Altman analyses indicated no large systematic differences with regard to the limits of agreement for the FSS, MFIS, and CIS20R, therefore we choose to calculate the Smallest Detectable Change (SDC) on the basis of the limits of agreement, which is based on the Standard Error of Measurement (SEM) consistency. Responsiveness is the ability of an instrument to measure real or important change over time, in the concept being measured.[34]

A distribution-based method was used to estimate the percentage change between the two assessments which should be exceeded to exclude measurement error, by determining the SDC.[35,36] The SDC was calculated by $1.96 \times \sqrt{2} \times SEM$ to indicate 95% confidence for real change between the two assessments scores.[37] The SEM was calculated by $SD \times \sqrt{(1-R)}$, with R=ICC and SD=√(total variance).[31] In order to allow comparison between the three questionnaires, the Minimal Detectable Change (MDC) was calculated by expressing the SDC as a percentage of the maximal feasible score for each questionnaire.[36]

### Concurrent validity

Since visual inspection of histograms of FSS, MFIS, and CIS20R scores for MS patients showed a normal distribution, concurrent validity was determined using Pearson's correlation coefficients. Strong association was defined if coefficients were beyond 0.70, whereas coefficients between 0.30 and 0.70 were classified as moderate to substantial and correlation coefficients less than 0.30 as a weak association.[38]

## Results

### Patient characteristics

Table 2.1 shows characteristics of the participants. Forty-three patients (mean age 48.7 years, median EDSS score 3.5) completed the three fatigue questionnaires. Of the 43 participants 13 (30%) patients were male. Participants had median scores of 52 on the FSS, 41 on the MFIS and 78.5 on the CIS20R. Age, gender, type of MS, duration of the disease, and EDSS

**Table 2.1 Participants characteristics (N=43)**

| Variable | Mean (SD) | Min–Max |
|---|---|---|
| Age (years) | 48.7 (7.0) | 38–64 |
| Gender; male/female | 13/30 | |
| Disease duration (years) | 14.3 (9.2) | 2–51 |
| Type MS; RR/SP/PP | 26/10/7 | |
| **Variable** | **Median (IQR)** | **Min–Max** |
| EDSS median | 3.5 | 1–6.5 |
| FSS | 52 (6) | 15–63 |
| MFIS | 41 (18) | 1–74.5 |
| Physical subscale | 21.5 (6) | 1–32 |
| Cognitive subscale | 17 (8) | 0–35.5 |
| Psychosocial subscale | 4 (2.5) | 0–8 |
| CIS20R | 78.5 (19) | 31.5–121.5 |
| Subjective feeling | 32.5 (13) | 9.5–56 |
| Concentration | 20.5 (7) | 5–31.5 |
| Motivation | 13.5 (8) | 4–25 |
| Physical activity | 12 (6) | 3–20.5 |

MS, Multiple Sclerosis; RR, Relapse Remitting; SP, Secondary Progressive; PP, Primary Progressive; EDSS, Expanded Disability Status Scale; FSS, Fatigue Severity Scale; MFIS, Modified Fatigue Impact Scale; CIS20R, Checklist Individual Strength; IQR, Inter Quartile Range; SD, Standard Deviation; Min, minimum; Max, maximum.

score were not significantly correlated with the FSS, MFIS, and CIS20R. All assessments were applied with a mean measurement interval of 7 days, according to the measurement protocol.

## Test-retest reliability

Table 2.2 shows the test-retest reliability of the FSS, MFIS and CIS20R for MS patients. Briefly, the ICCs for the FSS, MFIS and CIS20R were good (0.76, 0.85 and 0.81 respectively). ICCs for the MFIS domains were good, ranging from 0.73 to 0.88 and the ICCs for the CIS20R domains were also good, ranging from 0.77 to 0.84.

## Agreement

Figure 2.1 displays Bland and Altman plots for the total scores of the three fatigue questionnaires of the patients with MS. No systematic differences were observed between the first and second assessments of the various questionnaires.

## Responsiveness

Table 2.2 shows the SDCs and MDCs for the fatigue questionnaires in patients with MS. Responsiveness expressed by the SDC was 13.1 for the FSS, 16.2 for the MFIS and 24.8 for the CIS20R, resulting in a MDC of 20.7% for the FSS, 19.2% for the MFIS and 17.7% for the CIS20R, respectively.

**Table 2.2 Test retest reliability and responsiveness**

| | ICC (95% CI) | SDC | MDC (%) |
|---|---|---|---|
| FSS | 0.76 (.60 to .86) | 13.1 | 20.8 |
| MFIS | 0.85 (.74 to .92) | 16.2 | 19.3 |
| Physical subscale | 0.73 (.55 to .84) | 8.9 | 24.7 |
| Cognitive subscale | 0.88 (.79 to .93) | 8.0 | 20 |
| Psychosocial subscale | 0.81 (.68 to .89) | 2.3 | 28.8 |
| CIS20R | 0.81 (.67 to .89) | 24.8 | 17.7 |
| Subjective feeling | 0.84 (.72 to .91) | 11.8 | 21.1 |
| Concentration | 0.77 (.62 to .87) | 9.7 | 27.7 |
| Motivation | 0.81 (.67 to .89) | 6.6 | 23.6 |
| Physical activity | 0.84 (.54 to .84) | 6.9 | 32.9 |

FSS, Fatigue Severity Scale; MFIS, Modified Fatigue Impact Scale; CIS20R, Checklist Individual Strength; ICC, Intraclass Correlation Coefficients; SDC, Smallest Detectable Change; MDC%, Minimal Detectable Change; All ICCs $p<0.001$.

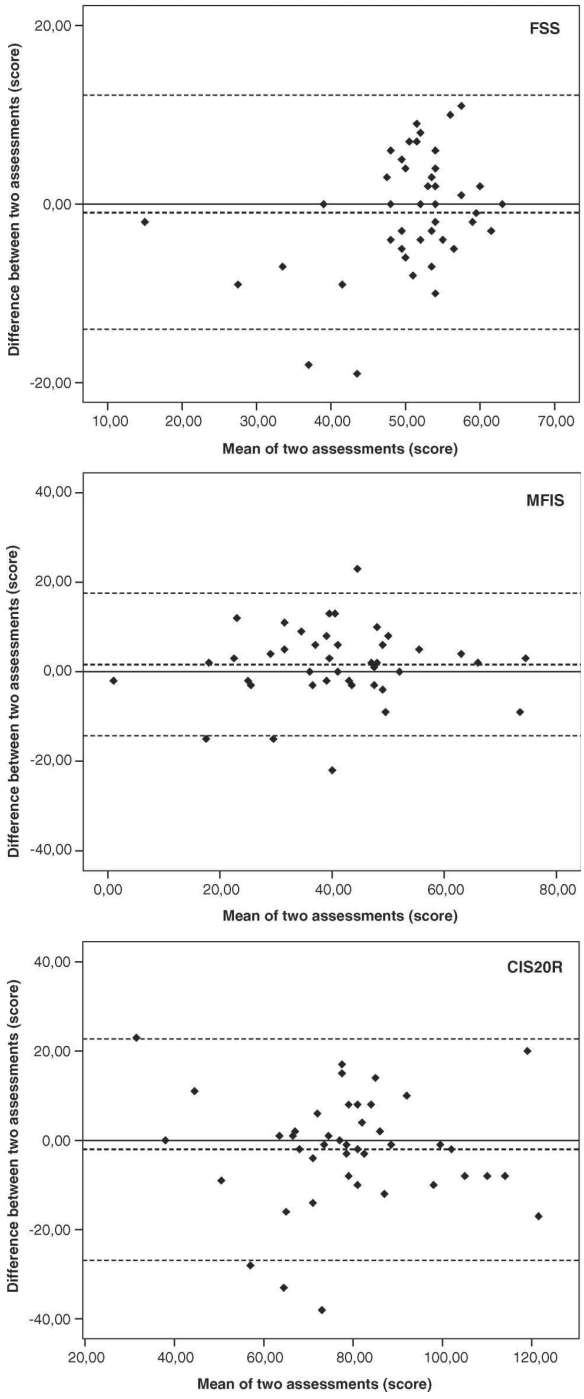


**Figure 2.1 Graphic representation according to the Bland and Altman technique.**
The **-----** bold lines represent the mean difference score, ------ lines represents the limits of agreement, defined as the mean ± 1.96 the standard deviation of the difference score.

### Concurrent validity

Moderate, but significant correlation coefficients were found for FSS versus MFIS (r=0.66; p<0.001), MFIS versus CIS20R (r=0.54; p<0.001) and CIS20R versus FSS (r=0.42; p=0.005).

## Discussion

The aim of the present study was to determine reproducibility, responsiveness and concurrent validity of the Dutch versions of the FSS, the MFIS and the CIS20R in patients with MS. All three self-report questionnaires showed good test-retest reliability for the total score and for the domains scores of the MFIS and CIS20R. Satisfactory agreement for the three questionnaires was found for the two consecutive assessments, using the Bland and Altman method.

Despite good test-retest reliability of FSS, MFIS and the CIS20R, the present study shows that fatigue questionnaires are not very responsive for change in patients with MS. The present study shows relatively large MDCs for all three measurement instruments. To our knowledge, no previous data on distribution based responsiveness of the FSS, the MFIS and the CIS20R have been published in patients with MS, making it difficult to interpret and compare the calculated minimal detectable changes. The relative large minimal detectable changes, ranging from 17 to 20% are likely to reflect true time-dependent fluctuations of perceived fatigue between two consecutive assessments. Contrary to conventional pre/post assessments, future studies should consider applying longitudinal designs with multiple repeated measures in which patients are monitored more frequently during the study.[39,40] The advantages of such designs are that the patterns of change over time can be analysed more precisely as a result of multiple repeated measurements within subjects, obtaining a more precise estimate of outcome measures and reducing the risk of type II error.

The rather modest correlation coefficients found between the three self-report questionnaires reflect a shared variance that ranged from 18% to 44% ($R^2$ 0.18–0.44). The relatively small overlap suggests that the FSS, MFIS and CIS20R measure for a large part different aspects of fatigue. Whereas the MFIS focuses on perceived impact of fatigue on physical, cognitive and psychosocial functioning, the FSS assesses the severity, frequency, modality and impact of fatigue on daily functioning.[25,29,41] The moderate association of FSS versus MFIS in MS populations was also reported in previous studies.[25,29,41] However, the found moderate associations for the CIS20R versus the FSS and MFIS in subjects with MS are new to the

literature, clearly illustrating that the various scales have different properties and cover different dimensions of the fatigue spectrum. In their attempt to provide a multidimensional characterization of fatigue in MS patients, Vercoulen et al.[15] already recognized that each dimension provides a unique contribution to the description of the patients' fatigue, and that use of a one-dimensional assessment would be a shortcoming in comprehensive fatigue assessment.

In addition to this multidimensionality, periodic fluctuations of fatigue severity under psychological and physiological stimuli[12] and the absence of an objective test means that quantification of MS-related fatigue remains a challenge. In an attempt to come to a comprehensive evaluation of fatigue, self report of fatigue should probably be accompanied by assessment of fatigue-related aspects such as mood, medication intake, sleep disturbances and infection that are likely to confound assessment of MS related fatigue.[11] The present study has some limitations. First, a relatively small sample of only ambulatory MS patients (i.e., EDDS score below 6.5) were included which limits the generalization of the present findings to the population of MS patients in general. Second, we used the distribution-based method to gain some insight in the percentage change between the two assessments which should be exceeded to exclude measurement error in our sample, but it is not necessarily informative as to what extent this is clinically meaningful. For determining clinically meaningful information about the relation between the found change and its importance a patient's or clinician's perspective is needed. Where a distribution based method informs about observed change in the sample, anchor-based methods estimate minimal important change directly.[42] Third, due to the relatively short interval between test and retest we can cannot completely rule out some recall bias.



In conclusion, the present study shows that the FSS, the MFIS and the CIS20R are highly reproducible self-report questionnaires to measure perceived fatigue in ambulatory patients with MS. However, the low responsiveness for change and limited overlap between these three questionnaires suggest that fatigue should be measured more frequently within subjects by using different questionnaires simultaneously. In addition, the found low responsiveness for change imply that the scales are less appropriate for use in individual assessment.

## Acknowledgements

This study was supported by 'Stichting MS Research' (project number 04-553 MS).

## References


1. Freal JE, Kraft GH, Coryell JK. Symptomatic fatigue in multiple sclerosis. Arch Phys Med Rehabil. 1984 Mar; 65(3): 135-8.
2. Murray TJ. Amantadine therapy for fatigue in multiple sclerosis. Can J Neurol Sci. 1985 Aug; 12(3): 251-4.
3. Fisk JD, Pontefract A, Ritvo PG, Archibald CJ, Murray TJ. The impact of fatigue on patients with multiple sclerosis. Can J Neurol Sci. 1994; 21: 9-14.
4. Bergamaschi R, Romani A, Versino M, Poli R, Cosi V. Clinical aspects of fatigue in multiple sclerosis. Funct Neurol. 1997 Sep-Oct; 12(5): 247-51.
5. Bakshi R, Shaikh ZA, Miletich RS, Czarnecki D, Dmochowski J, Henschel K,Janardhan V, Dubey N, Kinkel PR. Fatigue in multiple sclerosis and its relationship to depression and neurologic disability. Mult Scler. 2000 Jun; 6(3): 181-5.
6. Krupp LB, Christodoulou C. Fatigue in multiple sclerosis. Curr Neurol Neurosci Rep. 2001 May; 1(3): 294-8. Review.
7. Krupp LB, Alvarez LA, LaRocca NG, Scheinberg LC. Fatigue in multiple sclerosis. Arch Neurol. 1988 Apr; 45(4): 435-7.
8. Bakshi R, Miletich RS, Henschel K, Shaikh ZA, Janardhan V, Wasay M, Stengel LM, Ekes R, Kinkel PR. Fatigue in multiple sclerosis: cross sectional correlation with brain MRI findings in 71 patients. Neurology. 1999; 53: 1151-3.
9. Brañas P, Jordan R, Fry-Smith A, Burls A, Hyde C. Treatments for fatigue in multiple sclerosis: a rapid and systematic review. Health Technol Assess. 2000; 4(27): 1-61.
10. Rosenberg JH, Shafor R. Fatigue in multiple sclerosis: a rational approach to evaluation and treatment. Curr Neurol Neurosci Rep 2005; 5: 140-6.
11. Multiple Sclerosis Council for Clinical Guidelines. Fatigue and Multiple Sclerosis: Evidence Based Management Strategies for Fatigue in Multiple Sclerosis. Washington, DC: Paralyzed Veterans of America; 1998.
12. Chaudhuri A, Behan PO. Fatigue in neurological disorders. Lancet. 2004 Mar 20; 363(9413): 978-88.
13. Bakshi R. Fatigue associated with multiple sclerosis: diagnosis, impact and management. Mult Scler. 2003 Jun; 9(3): 219-27. Review.
14. Kos D, Kerckhofs E, Nagels G, D'hooghe MB, Ilsbroukx S. Origin of fatigue in multiple sclerosis: review of the literature. Neurorehabil Neural Repair. 2008 Jan-Feb; 22(1): 91-100. Epub 2007 Apr 4. Review.
15. Vercoulen JH, Hommes OR, Swanink CM, Jongen PJ, Fennis JF, Galama JM, van der Meer JW, Bleijenberg G. The measurement of fatigue in patients with multiple sclerosis. A multidimensional comparison with patients with chronic fatigue syndrome and healthy subjects. Arch Neurol. 1996 Jul; 53(7): 642-9.

16. Bol Y, Duits AA, Hupperts RM, Vlaeyen JW, Verhey FR. The psychology of fatigue in patients with multiple sclerosis: a review. J Psychosom Res. 2009 Jan; 66(1): 3-11. Epub 2008 Sep 24. Review.

17. Baker DG. Multiple sclerosis and thermoregulatory dysfunction. J Appl Physiol. 2002 May; 92(5): 1779-80.

18. Johansson S, Ytterberg C, Hillert J, Widén Holmqvist L, von Koch L. A longitudinal study of variations in and predictors of fatigue in multiple sclerosis. J Neurol Neurosurg Psychiatry. 2008 Apr; 79(4): 454-7.

19. Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. Arch Neurol 1989; 46: 1121-2.

20. Schwartz JE, Jandorf L, Krupp LB. The measurement of fatigue: a new instrument. J Psychosom Res. 1993 Oct; 37(7): 753-62.

21. Fisk JD, Ritvo PG, Ross L, Haase DA, Marrie TJ, Schlech WF. Measuring the functional impact of fatigue: initial validation of the fatigue impact scale. Clin Infect Dis. 1994 Jan; 18 Suppl 1: S79-83.

22. Iriarte J, Katsamakis G, de Castro P. The Fatigue Descriptive Scale (FDS): a useful tool to evaluate fatigue in multiple sclerosis. Mult Scler. 1999 Feb; 5(1): 10-6.

23. Schwid SR, Covington M, Segal BM, Goodman AD. Fatigue in Multiple Sclerosis: Current understandings and future directions. J Rehab Res Dev. 2002; 29 (2): 211-24.

24. Zwarts MJ, Bleijenberg G, van Engelen BG. Clinical neurophysiology of fatigue. Clin Neurophysiol. 2008 Jan; 119(1): 2-10. Epub 2007 Nov 26. Review.

25. Kos D, Kerckhofs E, Nagels G, D'Hooghe BD, Duquet W, Duportail M, Ketelaer P. Assessing fatigue in multiple sclerosis: Dutch modified fatigue impact scale. Acta Neurol Belg. 2003 Dec; 103(4): 185-91.

26. Vercoulen JHMM, Alberts M, Bleijenberg G. De Checklist Individual Strength (CIS). Gedragstherapie. 1999; 32: 131-6.

27. McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD, McFarland HF, Paty DW, Polman CH, Reingold SC, Sandberg-Wollheim M, Sibley W, Thompson A, van den Noort S, Weinshenker BY, Wolinsky JS. Recommended Diagnostic Criteria for Multiple Sclerosis: Guidelines from the International Panel on the Diagnosis of Multiple Sclerosis. Ann Neurol. 2001; 50: 121-7.

28. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology (Cleveland) 1983; 33: 1444-52.

29. Kos D, Kerckhofs E, Carrea I, Verza R, Ramos M, Jansa J. Evaluation of the Modified Fatigue Impact Scale in four different European countries. Multiple Sclerosis. 2005; 11: 76-80.

30. Beurskens AJ, Bültmann U, Kant I, Vercoulen JH, Bleijenberg G, Swaen GM. Fatigue among working people: validity of a questionnaire measure. Occup Environ Med. 2000 May; 57(5): 353-7.

31. De Vet HCW, Terwee CB, Knol DL, Bouter LM: When to use agreement versus reliability measures. J Clin Epidemiol. 2006; 59: 1033-9.

32. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. Clin Rehabil. 1998; 12(3): 187-99.

33. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986; 1(8476): 307-10.

34. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. Qual Life Res. 2003 Jun; 12(4): 349-6.

35. De Boer MR, de Vet HC, Terwee CB, Moll AC, Völker-Dieben HJ, van Rens GH. Changes to the subscales of two vision-related quality of life questionnaires are proposed. J Clin Epidemiol. 2005 Dec; 58(12): 1260-8. Epub 2005 Sep 12.

36. Uhlig T, Kvien TK, Pincus T. Test-retest reliability of disease activity core set measures and indices in rheumatoid arthritis. Ann Rheum Dis. 2009 Jun; 68(6): 972-5. Epub 2008 Oct 28.

37. Roebroeck ME, Harlaar J, Lankhorst GJ. The application of generalizability theory to reliability assessment: an illustration using isometric force measurements. Phys Ther. 1993; 73: 386-95.

38. McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care. 1993 Mar; 31(3): 247-63.

39. Goldstein H, Healy MJ, Rasbash J. Multilevel time series models with applications to repeated measures data. Stat Med. 1994 Aug 30; 13(16): 1643-55.

40. Twisk JWR. Applied longitudinal data analysis for epidemiology. Cambridge University Press. 2003.

41. Flachenecker P, Kümpfel T, Kallmann B, Gottschalk M, Grauer O, Rieckmann P,Trenkwalder C, Toyka KV. Fatigue in multiple sclerosis: a comparison of different rating scales and correlation to clinical parameters. Mult Scler. 2002 Dec; 8(6): 523-6.

42. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008 Feb; 61(2): 102-9. Epub 2007 Aug 3.