

# Assessment set for evaluation of clinical outcomes in multiple sclerosis: psychometric properties

Kamila Rasova<sup>1</sup>  
Patricia Martinkova<sup>2</sup>  
Jana Vyskotova<sup>3</sup>  
Michaela Sedova<sup>2</sup>

<sup>1</sup>Department of Rehabilitation, <sup>3<sup>rd</sup></sup> Faculty of Medicine, Charles University in Prague and Faculty Hospital Královské Vinohrady, Prague, Czech Republic; <sup>2</sup>Center of Biomedical Informatics and Department of Medical Informatics and Biostatistics, Institute of Computer Science, AS CR, Prague, Czech Republic; <sup>3</sup>Faculty of Medicine, Ostrava University, Ostrava, Czech Republic

**Purpose:** Multiple sclerosis (MS) manifests itself in a wide range of symptoms. Physiotherapy plays an important role in the treatment of those symptoms connected with mobility. For this therapy to be at its most effective it should be based on a systematic examination that is able to describe and classify damaged clinical functions meaningfully. The purpose of this study was to develop and validate a battery of tests and composite tests that can be used to systematically evaluate clinical features of MS treatable by physiotherapy.

**Methods:** The authors assembled a proposed battery of tests comprising known, standard, and validated assessments (low-contrast letter acuity testing; the Motricity Index; the Modified Ashworth Scale; the Berg Balance Scale; scales of postural reactions, tremor, dysdiadochokinesia, and dysmetria; the Nine-Hole Peg Test; the Timed 25-Foot Walk; and the 3-minute version of the Paced Auditory Serial Addition Test) and one test (knee hyperextension) of the authors' own. Normalization was calculated and six composite assessments were measured. Seventeen ambulatory subjects with MS were tested twice with the assessment set before undergoing physiotherapy, and 12 were also tested with the assessment set after the physiotherapy. The test–retest reliability, stability, internal consistency of composite measurements, sensitivity to changes after therapy, and correlation between measurements and the Kurtzke Expanded Disability Status Scale score were evaluated for all tests in the assessment set.

**Results:** A good internal consistency was confirmed for all tests in the proposed battery, and most of the tests also showed good test–retest reliability. While no significant changes occurred without treatment, significant posttreatment improvement was proved in all tests except for low-contrast letter acuity testing, where only a trend to improvement was proved.

**Conclusion:** The proposed assessment set is a good tool for the evaluation of clinical features of MS treatable by physiotherapy. This battery of tests is applicable in both clinical practice and research.

**Keywords:** outcome assessment, reproducibility of results, psychometric properties, test–retest reliability, internal consistency

## Introduction

Multiple sclerosis (MS) is a chronic autoimmune disease pathologically characterized by the presence of areas of demyelination and T-cell perivascular inflammation in the brain white matter, as well as by axonal degeneration. It clinically manifests itself by neurological abnormalities such as fatigue, numbness, paresthesia, muscular weakness and spasticity, double vision, optic neuritis, ataxia, bladder control problems, dysphagia, dysarthria, and cognitive dysfunction.<sup>1,2</sup> Physiotherapy plays an important role in the maintenance and improvement of damaged clinical func-

Correspondence: Kamila Rasova  
Department of Rehabilitation,  
Third Faculty of Medicine, Charles  
University in Prague, Ruská 87,  
100 00 Prague 10, Czech Republic  
Tel +42 0604 15 23 58 (mobile)  
Email kamila.rasova@centrum.cz

tions.<sup>3</sup> However, there is no consensus on what may be the most effective approach to achieve the best possible functionality, given the individual limitations. Contemporary rehabilitation research in MS lacks strict adherence to rigorous methodology and consistent use of a range of clinically appropriate and scientifically sound outcome measures.<sup>4,5</sup>

Haigh et al<sup>6</sup> conducted a survey on instruments commonly used in Europe to measure outcomes for MS patients. A questionnaire was sent to facilities providing rehabilitation (acute settings and rehabilitation units, both publicly and privately funded). Just over 100 outcome measures were reported as being used to assess patients with MS, although the majority of these measures were only used in a small number of centers. (A large number of measures – including the Environmental Status Scale, the Medical Outcomes Study Short-Form General Health Survey, or the Assessment of Motor and Process Skills – were being used in only one location, or a small number of locations, and with relatively few patients.) The Kurtzke Incapacity Status Scale, the Berg Balance Scale (BBS), and the Rivermead Mobility Index were the only measures that were used in more than five centers. The measures used most widely with MS patients were the Kurtzke Expanded Disability Status Scale (EDSS), the Functional Independence Measure, and the Ashworth Scale.

In a review by Khan et al<sup>3</sup> of multidisciplinary rehabilitation for MS patients, eight trials fulfilled the selection criteria, and a total of 42 outcome measures were used in these trials.

Based on the examples given, it is clear that the study and assessment of rehabilitation in MS has sparked the development of numerous outcome measures applicable to one or more of the disease's many dimensions. Outcomes research requires a systematic approach to describe and classify the outcomes meaningfully. The purpose of this study was to develop and validate a battery of tests and composite tests that can be used to systematically evaluate clinical features of MS treatable by physiotherapy.

Aims of this study were:

- to prepare standard tests for use in the Czech Republic (translation of standard tests and their validation);
- to validate standard tests for MS (those tests not yet validated for MS);
- to prepare a battery of tests and composite tests that is systematic, reliable, practical, acceptable to patients, capable of demonstrating rehabilitation effect, and predictive of clinically meaningful change.

## Methods

### Design of the study

An assessment set comprising 12 tests and six composite tests for the evaluation of clinical outcomes in MS was prepared. Seventeen patients with MS who met the inclusion criteria were selected. An independent neurologist determined the EDSS<sup>7</sup> score and duration of the disease. The assessment set was performed twice within 3–5 weeks by an independent physiotherapist. The patients did not change their habits during this time. After the second examination with assessment set, a physiotherapy program, consisting of two 2-hour sessions each week for 2 months, was offered to the patients. Twelve patients finished the physiotherapy program, and these patients were also examined at the end of the program.

### Selection and characteristics of the subjects

Seventeen outpatients with the diagnosis of MS according to the criteria of McDonald et al<sup>8</sup> (either gender; age range, 30–57 years; suffering from relapsing-remitting, primary progressive, or secondary progressive MS; stability of clinical status in the preceding 3 months; prevailing motor impairment; able to move independently; able to walk at least 200 m with two canes [EDSS score  $\leq$  5]; able to undergo ambulatory treatment; and right-handed)<sup>9</sup> were chosen randomly from MS centers in the Czech Republic. Persons with cognitive impairment that could hinder understanding of the tasks to be accomplished were excluded from enrollment. All patients were required to sign an informed consent document before inclusion in this study. Table 1 outlines the characteristics of the patients.

### Preparation of assessment set and procedure

The assessment set was prepared from well-known, standard, and validated tests and one test of the authors' own. The selection of tests was based on team experience and literature review.

The authors included the tests used most frequently in clinical trials of MS: low-contrast letter acuity (L-CLA) testing, the Nine-Hole Peg Test (NHPT), the Timed 25-Foot Walk (T25FW), and the 3-minute version of the Paced Auditory Serial Addition Test (PASAT 3). The authors also included frequently used tests that evaluate the leading problems (eg, spastic paresis, cerebellar symptoms) that therapists deal with in MS patients: the Motricity Index (MI),

**Table 1** Characteristics of patients in study

| Characteristic                           | Patients   |
|--|------------|
| Sex                                      |            |
| Female [n (%)]                           | 10 (59)    |
| Male [n (%)]                             | 7 (41)     |
| Age (years)                              |            |
| Mean (SD)                                | 43.3 (9.0) |
| Range                                    | 30–57      |
| Type of MS                               |            |
| Primary progressive [n (%)]              | 1 (6)      |
| Relapsing remitting [n (%)]              | 11 (65)    |
| Secondary progressive [n (%)]            | 5 (29)     |
| Disease duration since diagnosis (years) |            |
| Mean (SD)                                | 10.1 (5.8) |
| Range                                    | 3–23       |
| EDSS score                               |            |
| Mean (SD)                                | 3.7 (1.0)  |
| Range                                    | 1.5–5.0    |
| 0.0–2.0 [n (%)]                          | 2 (12)     |
| 2.5–4.0 [n (%)]                          | 10 (59)    |
| 4.5–6.5 [n (%)]                          | 5 (29)     |

**Abbreviations:** SD, standard deviation; MS, multiple sclerosis; EDSS, Expanded Disability Status Scale.

the Modified Ashworth Scale (MAS), the BBS, and tests for tremor (T), dysdiadochokinesia (DD), and dysmetria (DM). Finally, the authors included tests that evaluate clinical features of MS that, in the authors' opinion, best react to physical therapy: scales of postural reactions (PRs) and the authors' own test – knee hyperextension (KH).

The back-translation method<sup>10</sup> was used for the translation of each test.

A trained physiotherapist experienced in performing ambulatory examination administered the scoring of patients. The amount of time required to complete the whole battery of tests was about 1 hour, and the whole assessment was videotaped.

The same examiner performed the consecutive assessments at the same time of day, and preferably at the same day of the week, with the measures administered in the same order each time (all subtests first in lying, then in sitting, then in standing, and, finally, in walking). The examiner used a detailed protocol with precise and standardized instructions. Participants received refreshments.

## Tests included in assessment set

To evaluate visual function, a L-CLA score<sup>11</sup> was used to measure the total number of letters read correctly at three contrast levels (100%, 2.5%, and 1.25%). Visual function was determined as an average of the three contrast levels (each giving a minimum of 0 and a maximum of 60 correct answers).

To evaluate muscle power function (strength), the MI<sup>12</sup> was used. The MI value for each extremity was determined – the extremity MI includes three actions, each scored between 0 and 33 (where 0 indicates worst muscle power function), which are added together to make a total possible score of 99 plus 1, giving a scale of 1–100.<sup>13</sup> The total MI includes 12 items (three items for each extremity), which, added together plus 4, give a scale of 4–400. The three actions for the left and right upper extremities are pinch grip, elbow flexion, and shoulder abduction; the three actions for the left and right lower extremities are ankle dorsiflexion, knee extension, and hip flexion.

To evaluate muscle tone function (spasticity), the MAS<sup>14,15</sup> was used. The MAS is an 18-item five-point rating scale, each item ranging from 0 to 4 (where 0 indicates no increased tone and 4 indicates limb rigid in flexion or extension). The amount of tone felt as a limb was moved passively through its arc of motion was measured. The MAS score for each extremity was determined: the MAS for upper extremities covers elbow flexors, elbow pronators, elbow supinators, wrist flexors, and digital flexors; the MAS for lower extremities covers hip adductors, knee extensors, knee flexors, and plantar flexors.

To evaluate changing and maintaining a position (balance), the BBS<sup>16,17</sup> was used. The BBS is a 14-item five-point rating scale, each item ranging from 0 to 4 (where 0 indicates the lowest level of function), that assesses the performance of functional tasks. Further, PRs (righting, equilibrium, and protective reactions)<sup>18</sup> were evaluated from videotape, using a rating scale from 0 to 3 (where 0 indicates only head righting reactions noted and 3 indicates normal reactions – all equilibrium and protective reactions are present),<sup>19</sup> in 12 actions (being drawn left and right by another person in a sitting position on a stationary supporting surface; tipped backwards, forwards, left, and right in standing; and steps to save forwards, backwards, left, and right).<sup>20</sup>

Rest, postural, and intention tremor (T) on the upper and lower extremities was evaluated using a procedure described by Fahn et al.<sup>21</sup> This procedure comprises 12 items (three for each extremity) rated on a five-point scale (where 0 indicates none and 4 indicates severe amplitude).

To evaluate DD, a five-point rating scale (where 0 indicates no problem and 4 indicates the subject is unable to perform a repetitive sequential movement) described by Alusi et al<sup>22</sup> was used for each extremity. The scores for each extremity were added together for the total DD score on a scale of 0–16.

To evaluate DM, a five-point rating scale (where 0 indicates no impairment and 4 indicates the subject cannot use hands/legs) also described by Alusi et al<sup>22</sup> was used for each extremity. The scores for each extremity were added together for the total DM score on a scale of 0–16.

To evaluate the stability of joint function, the authors' own scale was used. This scale rates genu recurvatum (KH test) ranging from 0 to 6 (where 0 indicates there is no hyperextension in the knee either in standing or in quick walking; 1 indicates there is hyperextension in the knee only during quick walking, and it is voluntarily influenced; 2 indicates there is hyperextension in the knee only during quick walking, but it cannot be influenced voluntarily; 3 indicates there is hyperextension in the knee also during slow walking, and it is voluntarily influenced; 4 indicates there is hyperextension in the knee also during slow walking, but it cannot be influenced voluntarily; 5 indicates there is hyperextension in the knee even in standing, and it is voluntarily influenced; and 6 indicates there is hyperextension in the knee in standing, but it is not influenced voluntarily). The function of the knee was evaluated for both left and right lower extremities and the total KH score was calculated as their average value.

To evaluate fine motor skills, the NHPT,<sup>23</sup> a quantitative measure of upper extremities (arm and hand), was used. The NHPT measures the time interval (in seconds) during which a patient places nine pegs into holes in a testing board as fast as possible and then picks them up with one hand, one peg after another, and puts them into a bowl. The duration of the test was limited to 60 seconds. The NHPT was performed twice for each upper extremity and then averaged – this average was calculated as the total NHPT score.

To evaluate walking, the T25FW test,<sup>23</sup> which measures maximal walking speed over a distance of 25 feet or 7.6 m from a standing start, was used. The duration of the test was limited to 20 seconds. The score was calculated as the average from two consecutive measurements.

To evaluate mental function, the PASAT 3<sup>23</sup> was used. It consists of 60 true/false items, where a total of 0 indicates the worst function.

## Data preparation and normalization

### Assessment set

The data were recorded on a Microsoft Excel® spreadsheet (Microsoft Corporation, Redmond, WA) by an independent person with MS and were controlled by a second independent person with MS (paid by a project of European Social Fund Involving Training Workplaces for Disabled People).

Total scores obtained for tests in the assessment set were normalized to a scale from 0 to 1 (where 0 indicates

the worst function and 1 indicates the best function). Normalization provides better orientation in scales, allows better comparison, and allows calculation of totals for all four extremity functions and total index of clinical functions (TICF). To calculate normalization, the minimum (min) possible value was subtracted from the measurement and this difference was then divided by the difference between the maximum (max) and min possible values:

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

If necessary, this was subtracted from 1 in the case of opposite scoring – that is, if 0 stands for the best function, which is the case for the MAS, T, DD, DM, KH, the NHPT, and the T25FW.

The logarithms of time measurements (NHPT and T25FW) were used for normalization. Minimum and maximum values were set to be 10 and 60 seconds for NHPT and 3 and 20 seconds for T25FW, respectively.

### Six composite assessments

As well as the normalization of total scores, normalization of total scores for the extremities was calculated and averaged into the total extremity function. For normalized left (NLUEF) and right upper extremity function (NRUEF), the following normalized extremity total scores were averaged: normalized Modified Ashworth Scale (NMAS), normalized Motricity Index (NMI), normalized tremor, (NT), normalized dysdiadochokinesia (NDD), normalized dysmetria (NDM), and normalized Nine-Hole Peg Test (NNHPT). For normalized left (NLLEF) and right lower extremity function (NRLEF), the following normalized extremity total scores were averaged: NMAS, NMI, NT, NDD, NDM, and normalized knee hyperextension (NKH). The balance index (BI) was calculated as an average of normalized BBS and normalized PR scores. For TICF, all normalized measurements (normalized low-contrast letter acuity [NL-CLA], NMI, NMAS, normalized Berg Balance Scale [NBBS], normalized postural reactions [NPRs], NT, NDD, NDM, NKH, NNHPT, normalized Timed 25-Foot Walk [NT25FW], and normalized 3-minute version of the Paced Auditory Serial Addition Test [NPASAT 3]) of one patient were averaged.

### Statistical analysis

The test–retest reliability was evaluated by intraclass correlation coefficient (ICC) (3,1), consistency version.<sup>24</sup> Stability of measurements (changes without treatment) and improvement after treatment were tested by paired *t*-test;

*P*-values were corrected for multiple comparisons using false discovery rate correction.<sup>25</sup> Internal consistency of composite assessments (L-CLA, MI, MAS, BBS, T, DD, DM, PRs, KH, NHPT, and T25FW) was evaluated by Cronbach's alpha, estimated from the second examination. Pearson correlations and a dendrogram of cluster analysis were used to assess connections between measurements. Spearman correlations were used to assess connections of clinical measures and EDSS scores. Statistical analyses were processed using software R<sup>26</sup> and its library psych.<sup>27</sup>

## Results

Seventeen patients were enrolled in the study. Descriptive statistics for all three measurements (Examinations 1, 2, and 3) are shown in Table 2. Descriptive statistics for normalized first measurements (Examination 1) are shown in Table 3. These statistics shows that in many assessments, MS patients reach only a narrow band of possible values – MI, MAS, BBS, T, and DM scores are generally higher than 60% (no patient had low values in these functions). On the other hand, the function of the KH test was lower than 40% in all the patients.

All of the composite tests showed a good internal consistency (>0.75) (see Table 4).

There were no significant changes without treatment in any of the tests or composite tests (see Table 5). Good test-retest reliability (>0.75) was obtained in seven of 12 tests (L-CLA, BBS, PRs, KH, NHPT, T25FW, and PASAT 3) and four composite tests (LLEF, RLEF, BI, TICF). The lowest ICC (0.39) was obtained for left upper extremity function.

All of the tests in the assessment set were sensitive to posttreatment changes: significant posttreatment improvement was proved in all tests in the battery except for L-CLA testing, where a trend to improvement was proved.

Correlations between normalized EDSS score and normalized clinical assessments are shown in Table 6. A greater number of patients would be needed to prove significance or to fit an optimal model of prediction of EDSS score (measured by a neurologist) by clinical assessment (measured by a therapist). The highest correlations were reached between normalized EDSS score and MI (0.60), NBBS (0.47), NHPT (0.46), and DD scores (0.44).

Connections between assessments were estimated by Pearson correlations (see Figure 1). High correlations were found between NBBS and NMI scores (0.80) and NBBS and NT25FW scores (0.75), indicating the possible reduction of the battery. Generally, the correlation matrix and dendrogram of cluster analysis (see Figure 2) suggest that the proposed battery of tests is multidimensional and that it provides complex information on a patient's clinical condition.

## Discussion

Seventeen patients were enrolled at the beginning of the study. As the study took a relatively long period of time, only 12 subjects completed the study, with the rest dropping out for personal or/and health reasons. The sample size was relatively small but comparable with other studies assessing psychometric properties of clinical tests mentioned in the literature.<sup>13,15,28-31</sup> Despite the small sample size, the results look consistent. The gender distribution, type of MS, spectrum of disease duration, age, and range of EDSS score represent ambulatory MS patients in general.

## The assessment set

The battery of tests was prepared with the aim of evaluating clinical functions connected with motor deficit in patients with MS and to sensibly detect the types of changes connected

**Table 2** Assessment set: descriptive statistics

| Measure | Examination 1 |       |              | Examination 2 |       |              | Examination 3 |       |              |
|---------|---------------|-------|--------------|---------------|-------|--------------|---------------|-------|--------------|
|         | Mean          | SD    | Patients (n) | Mean          | SD    | Patients (n) | Mean          | SD    | Patients (n) |
| L-CLA   | 35.31         | 7.80  | 15           | 32.10         | 10.35 | 17           | 36.64         | 8.77  | 12           |
| MI      | 316.53        | 29.03 | 15           | 313.06        | 29.96 | 17           | 342.33        | 24.36 | 12           |
| MAS     | 24.43         | 4.07  | 15           | 22.47         | 4.33  | 17           | 11.08         | 3.78  | 12           |
| BBS     | 50.27         | 4.48  | 15           | 49.09         | 5.36  | 17           | 52.17         | 2.30  | 12           |
| T       | 9.23          | 3.34  | 15           | 9.24          | 2.82  | 17           | 6.17          | 2.09  | 12           |
| DD      | 6.12          | 1.42  | 15           | 5.65          | 1.67  | 17           | 4.12          | 1.36  | 12           |
| DM      | 3.63          | 0.88  | 15           | 4.18          | 1.25  | 17           | 2.75          | 0.92  | 12           |
| PRs     | 27.33         | 5.83  | 15           | 27.57         | 4.96  | 17           | 31.02         | 3.67  | 12           |
| KH      | 5.33          | 0.74  | 15           | 5.29          | 0.75  | 17           | 4.62          | 1.17  | 12           |
| NHPT    | 24.93         | 4.66  | 15           | 24.02         | 4.16  | 16           | 22.37         | 3.80  | 12           |
| T25FW   | 5.87          | 1.61  | 13           | 5.99          | 1.74  | 14           | 4.71          | 0.49  | 12           |
| PASAT 3 | 41.40         | 15.25 | 15           | 45.50         | 11.78 | 16           | 49.25         | 9.21  | 12           |

**Abbreviations:** SD, standard deviation; L-CLA, low-contrast letter acuity; MI, Motricity Index; MAS, Modified Ashworth Scale; BBS, Berg Balance Scale; T, tremor; DD, dysdiadochokinesia; DM, dysmetria; PRs, postural reactions (righting, equilibrium, and protective reactions); KH, knee hyperextension; NHPT, Nine-Hole Peg Test; T25FW, Timed 25-Foot Walk; PASAT 3, 3-minute version of the Paced Auditory Serial Addition Test.

**Table 3** Descriptive statistics of normalized measurements (Examination I)

| Measure  | Examination I |      |      |      |          |          |
|----------|---------------|------|------|------|----------|----------|
|          | Mean          | SD   | Min  | Max  | Skewness | Kurtosis |
| NL-CLA   | 0.59          | 0.13 | 0.31 | 0.76 | -0.48    | -0.58    |
| NMI      | 0.79          | 0.07 | 0.65 | 0.93 | 0.13     | -0.50    |
| NMAS     | 0.65          | 0.06 | 0.55 | 0.75 | -0.13    | -1.06    |
| NBBS     | 0.90          | 0.08 | 0.70 | 1.00 | -0.93    | 0.13     |
| NT       | 0.81          | 0.07 | 0.69 | 0.96 | 0.36     | -0.56    |
| NDD      | 0.62          | 0.09 | 0.47 | 0.73 | -0.25    | -1.29    |
| NDM      | 0.77          | 0.05 | 0.69 | 0.88 | 0.53     | -0.79    |
| NPRs     | 0.65          | 0.14 | 0.30 | 0.83 | -1.00    | 0.30     |
| NKH      | 0.11          | 0.12 | 0.00 | 0.33 | 0.70     | -1.16    |
| NNHPT    | 0.78          | 0.05 | 0.71 | 0.85 | -0.08    | -1.30    |
| NT25FW   | 0.79          | 0.08 | 0.60 | 0.90 | -0.88    | -0.21    |
| NPASAT 3 | 0.69          | 0.25 | 0.10 | 0.98 | -0.62    | -0.52    |
| NLUEF    | 0.76          | 0.04 | 0.70 | 0.83 | 0.26     | -1.01    |
| NRUEF    | 0.77          | 0.07 | 0.64 | 0.89 | -0.26    | -0.70    |
| NLLEF    | 0.57          | 0.06 | 0.47 | 0.70 | 0.17     | -0.71    |
| NRLEF    | 0.60          | 0.06 | 0.49 | 0.74 | 0.64     | -0.20    |
| BI       | 0.77          | 0.07 | 0.63 | 0.87 | -0.14    | -1.06    |
| TICF     | 0.69          | 0.04 | 0.62 | 0.77 | 0.43     | -1.12    |

**Abbreviations:** SD, standard deviation; Min, minimum; Max, maximum; NL-CLA, normalized low-contrast letter acuity; NMI, normalized Motricity Index; NMAS, normalized Modified Ashworth Scale; NBBS, Normalized Berg Balance Scale; NT, normalized tremor; NDD, normalized dysdiadochokinesia; NDM, normalized dysmetria; NPRs, normalized postural reactions (righting, equilibrium, and protective reactions); NKH, normalized knee hyperextension; NNHPT, normalized Nine-Hole Peg Test; NT25FW, normalized Timed 25-Foot Walk; NPASAT 3, normalized 3-minute version of the Paced Auditory Serial Addition Test; NLUEF, normalized left upper extremity function; NRUEF, normalized right upper extremity function; NLLEF, normalized left lower extremity function; NRLEF, normalized right lower extremity function; BI, balance index; TICF, total index of clinical functions.

with physiotherapy. Results of this study show that the chosen tests are sensitive to posttreatment changes. The authors are convinced that the assessment set may also be useful for detecting differences between therapies and their effects (Rasova K, unpublished data, 2012).

Recently it has been recommended that clinical practice in MS, including rehabilitation, should be based on the International Classification of Functioning, Disability and Health – a globally agreed upon framework and system for classifying the typical spectrum of problems in the functioning of people, given the environmental context in which they live.<sup>32</sup> Based on this model, there are many different domains that have to be measured and treated, and hence the authors assembled the proposed battery with 12 tests. Similarly, Paltamaa et al,<sup>33</sup> who also used the International Classification of Functioning, Disability and Health model, assembled a proposed battery with 12 tests, but of these 12 tests only the MAS and the BBS were the same as the tests selected by the present authors.

The most appropriate (standardized, quantitative, with minimal costs and special equipment, applicable in

ambulatory practice, safe and feasible) generic and/or disease-specific measures were selected for inclusion in the assessment set from different domains of body function based on information available in the literature. The assessment set is multidimensional, in order to reflect the principal way in which MS affects clinical functions, and it provides mainly interval data. A skilled physiotherapist is able to perform the assessment set within 1 hour, which is the standard length of a physiotherapeutic examination paid for by health insurance. Patients were familiar with most of the tests other than the PASAT. No negative events such as muscle pain or tiredness were increased in connection with the tests. The assessment set is no doubt demanding in its requirements for organization and time. In usual clinical practice the domains and measures are chosen according to what is considered important for the MS subject or what effect the target of therapy wants to achieve (two to three primary and eventually two to three secondary outcomes are measured). On the other hand, the proposed assessment set provides objective, systematic, and multidimensional information required about clinical functions for efficient physiotherapy.

### Tests in the battery: how they were chosen and their psychometric properties

Among clinical measures evaluating visual functions, contrast letter acuity (Sloan charts) and contrast sensitivity (Pelli–Robson chart) demonstrate the greatest capacity to identify binocular visual dysfunction in MS. Sloan chart testing also captures unique aspects of neurologic dysfunction not captured by current EDSS or Multiple Sclerosis Functional Composite (MSFC) components.<sup>34</sup> For this reason, the authors selected L-CLA testing for inclusion in the assessment set. The authors conclude that the L-CLA score demonstrates good test–retest reliability (ICC: 0.82) and good internal consistency (Cronbach’s alpha: 0.90). However, from the proposed battery of tests, L-CLA testing was the assessment least sensitive to posttreatment changes. Baier et al<sup>11</sup> confirmed a very good concurrent and predictive validity in patients with relapsing–remitting and secondary progressive MS (correlated with the EDSS and the MSFC) that provides additional information relevant to the MS disease process.

Several tests have been developed to evaluate motor impairment: the Motor Club Assessment, the Northwick Park Motor Assessment, the Rivermead Motor Assessment, the Medical Research Council Scale, and the Motricity index.<sup>13</sup> The authors selected the MI for inclusion in the assessment

**Table 4** Assessment set: test–retest reliability, stability (changes without treatment), and changes after therapy

| Measure  | Test–retest reliability |       |       | Stability/changes without treatment (E2–E1) |      |              |         | Improvement after treatment (E3–E2) |      |              |         |
|----------|-------------------------|-------|-------|---|------|--------------|---------|-------------------------------------|------|--------------|---------|
|          | ICC                     | LCL95 | UCL95 | Mean  | SD   | Patients (n) | P-value | Mean                                | SD   | Patients (n) | P-value |
| NL-CLA   | 0.82                    | 0.54  | 0.93  | −0.03                                       | 0.09 | 15           | 0.579   | 0.04                                | 0.10 | 12           | 0.081   |
| NMI      | 0.56                    | 0.09  | 0.83  | −0.02                                       | 0.07 | 15           | 0.579   | 0.05                                | 0.07 | 12           | 0.015   |
| NMAS     | 0.49                    | −0.01 | 0.79  | 0.04  | 0.06 | 15           | 0.139   | 0.16                                | 0.08 | 12           | <0.001  |
| NBBS     | 0.78                    | 0.47  | 0.92  | −0.02                                       | 0.06 | 15           | 0.579   | 0.04                                | 0.07 | 12           | 0.036   |
| NT       | 0.52                    | 0.03  | 0.81  | 0.01  | 0.06 | 15           | 0.709   | 0.06                                | 0.05 | 12           | 0.001   |
| NDD      | 0.40                    | −0.13 | 0.75  | 0.02  | 0.09 | 15           | 0.709   | 0.09                                | 0.13 | 12           | 0.020   |
| NDM      | 0.47                    | −0.03 | 0.79  | −0.02                                       | 0.05 | 15           | 0.579   | 0.10                                | 0.09 | 12           | 0.003   |
| NPRs     | 0.96                    | 0.88  | 0.99  | 0.00  | 0.04 | 15           | 0.897   | 0.11                                | 0.11 | 12           | 0.005   |
| NKH      | 0.98                    | 0.93  | 0.99  | 0.00  | 0.03 | 15           | >0.999  | 0.15                                | 0.24 | 12           | 0.029   |
| NNHPT    | 0.88                    | 0.69  | 0.96  | 0.01  | 0.02 | 15           | 0.579   | 0.01                                | 0.01 | 11           | 0.005   |
| NT25FW   | 0.95                    | 0.84  | 0.99  | −0.01                                       | 0.03 | 11           | 0.709   | 0.04                                | 0.04 | 10           | 0.005   |
| NPASAT 3 | 0.92                    | 0.78  | 0.97  | 0.07  | 0.09 | 15           | 0.137   | 0.06                                | 0.10 | 11           | 0.039   |
| NLUEF    | 0.39                    | −0.13 | 0.74  | 0.01  | 0.04 | 15           | 0.593   | 0.07                                | 0.05 | 11           | 0.001   |
| NRUEF    | 0.68                    | 0.27  | 0.88  | 0.00  | 0.05 | 15           | 0.992   | 0.07                                | 0.04 | 11           | <0.001  |
| NLLEF    | 0.84                    | 0.60  | 0.95  | 0.02  | 0.03 | 15           | 0.257   | 0.10                                | 0.05 | 12           | <0.001  |
| NRLEF    | 0.75                    | 0.41  | 0.91  | 0.00  | 0.04 | 15           | 0.992   | 0.11                                | 0.07 | 12           | 0.001   |
| BI       | 0.88                    | 0.68  | 0.96  | −0.01                                       | 0.04 | 15           | 0.579   | 0.07                                | 0.06 | 12           | 0.003   |
| TICF     | 0.77                    | 0.35  | 0.93  | 0.01  | 0.03 | 11           | 0.709   | 0.06                                | 0.02 | 10           | <0.001  |

**Abbreviations:** E, examination; ICC, intraclass correlation coefficient; LCL95, lower limit of 95% two-sided confidence interval; UCL95, upper limit of 95% two-sided confidence interval; SD, standard deviation; NL-CLA, normalized low-contrast letter acuity; NMI, normalized Motricity Index; NMAS, normalized Modified Ashworth Scale; NBBS, normalized Berg Balance Scale; NT, normalized tremor; NDD, normalized dysdiadochokinesia; NDM, normalized dysmetria; NPRs, normalized postural reactions (righting, equilibrium, and protective reactions); NKH, normalized knee hyperextension; NNHPT, normalized Nine-Hole Peg Test; NT25FW, normalized Timed 25-Foot Walk; NPASAT 3, normalized 3-minute version of the Paced Auditory Serial Addition Test; NLUEF, normalized left upper extremity function; NRUEF, normalized right upper extremity function; NLLEF, normalized left lower extremity function; NRLEF, normalized right lower extremity function; BI, balance index; TICF, total index of clinical functions.

set because it is a simple and quick measure of the loss of voluntary motor power (general strength of movement at each joint of upper and lower extremities) that can also inform about general motor impairment of the extremities. For psychometric properties, a good to excellent criterion validity of lower extremities,<sup>35</sup> good upper extremity Pearson

correlations with a handheld dynamometer, a good construct validity of upper extremities,<sup>36</sup> and a good interrater reliability and validity has been confirmed, although only in stroke patients.<sup>13</sup> In the present study, the MI demonstrated moderate test–retest reliability (ICC: 0.56) and good internal consistency (Cronbach’s alpha: 0.87).

**Table 5** Internal consistency of composite assessments

| Measure | Patients (n) | Items (n) | Cronbach’s alpha | LCL95 | UCL95 |
|---------|--------------|-----------|------------------|-------|-------|
| L-CLA   | 17           | 3         | 0.90             | 0.76  | 0.96  |
| MI      | 17           | 12        | 0.87             | 0.75  | 0.94  |
| MAS     | 17           | 18        | 0.78             | 0.59  | 0.91  |
| BBS     | 17           | 20        | 0.94             | 0.89  | 0.97  |
| T       | 17           | 12        | 0.76             | 0.55  | 0.90  |
| DD      | 17           | 8         | 0.92             | 0.84  | 0.97  |
| DM      | 17           | 4         | 0.82             | 0.62  | 0.93  |
| PRs     | 17           | 14        | 0.92             | 0.85  | 0.97  |
| KH      | 17           | 2         | 0.85             | 0.57  | 0.94  |
| NHPT    | 16           | 4         | 0.93             | 0.84  | 0.97  |
| T25FW   | 14           | 2         | 0.96             | 0.87  | 0.99  |

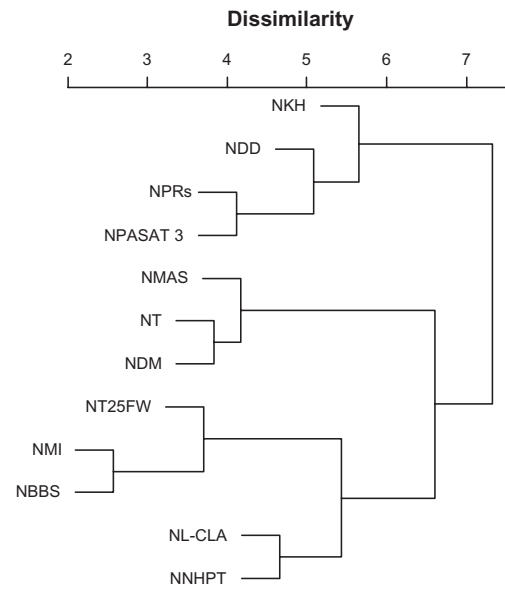
**Abbreviations:** LCL95, lower limit of 95% two-sided confidence interval; UCL95, upper limit of 95% two-sided confidence interval; L-CLA, low-contrast letter acuity; MI, Motricity Index; MAS, Modified Ashworth Scale; BBS, Berg Balance Scale; T, tremor; DD, dysdiadochokinesia; DM, dysmetria; PRs, postural reactions (righting, equilibrium, and protective reactions); KH, knee hyperextension; NHPT, Nine-Hole Peg Test; T25FW, Timed 25-Foot Walk.

Multiple biomechanical and electrophysiological methods for measuring muscle tone function have been developed (H-reflex testing, quantification of deep tendon reflexes and clonus, resonant frequency test, pendulum test, instrumented torque measurements during passive motion at present velocity, isokinetic dynamometry, and electromyography). Unfortunately, these methods have many limits – mainly that they need special equipment, differ in methodology, and are not accessible and administrable by clinicians.<sup>37</sup> It seems that for spasticity evaluation, clinical scales could be more useful. Twenty-four clinical scales that assess spasticity and/or related phenomena as well as ten scales for “active function” and three scales for “passive function” having an association with spasticity could be identified. For many scales, reliability data is missing.<sup>38</sup> However, the evaluation of spasticity is usually performed using the MAS, and this was the main reason why the authors selected this scale for

**Table 6** Spearman correlations between normalized Expanded Disability Status Scale score and clinical assessments

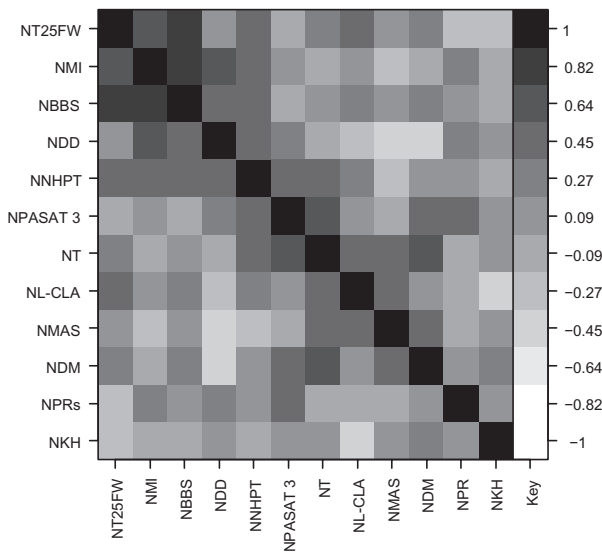
| Measure  | Correlation coefficient (r) | P-value |
|----------|-----------------------------|---------|
| NL-CLA   | 0.02                        | 0.972   |
| NMI      | 0.60                        | 0.197   |
| NMAS     | 0.05                        | 0.958   |
| NBBS     | 0.47                        | 0.197   |
| NT       | 0.17                        | 0.673   |
| NDD      | 0.44                        | 0.197   |
| NDM      | 0.26                        | 0.528   |
| NPRs     | 0.37                        | 0.253   |
| NKH      | 0.01                        | 0.972   |
| NNHPT    | 0.46                        | 0.197   |
| NT25FW   | 0.23                        | 0.645   |
| NPASAT 3 | 0.15                        | 0.689   |
| NLUEF    | 0.45                        | 0.197   |
| NRUEF    | 0.17                        | 0.673   |
| NLLEF    | 0.46                        | 0.197   |
| NRLEF    | 0.43                        | 0.197   |
| BI       | 0.50                        | 0.197   |
| TICF     | 0.43                        | 0.244   |

**Abbreviations:** NL-CLA, normalized low-contrast letter acuity; NMI, normalized Motricity Index; NMAS, normalized Modified Ashworth Scale; NBBS, normalized Berg Balance Scale; NT, normalized tremor; NDD, normalized dysdiadochokinesia; NDM, normalized dysmetria; NPRs, normalized postural reactions (righting, equilibrium, and protective reactions); NKH, normalized knee hyperextension; NNHPT, normalized Nine-Hole Peg Test; NT25FW, normalized Timed 25-Foot Walk; NPASAT 3, normalized 3-minute version of the Paced Auditory Serial Addition Test; NLUEF, normalized left upper extremity function; NRUEF, normalized right upper extremity function; NLLEF, normalized left lower extremity function; NRLEF, normalized right lower extremity function; BI, balance index; TICF, total index of clinical functions.



**Figure 2** Dissimilarities between assessments (dendrogram of cluster analysis). **Abbreviations:** NL-CLA, normalized low-contrast letter acuity; NMI, normalized Motricity Index; NMAS, normalized Modified Ashworth Scale; NBBS, normalized Berg Balance Scale; NT, normalized tremor; NDD, normalized dysdiadochokinesia; NDM, normalized dysmetria; NPRs, normalized postural reactions (righting, equilibrium, and protective reactions); NKH, normalized knee hyperextension; NNHPT, Normalized Nine-Hole Peg Test; NT25FW, normalized Timed 25-Foot Walk; NPASAT 3, normalized 3-minute version of the Paced Auditory Serial Addition Test.

**Correlation plot**



**Figure 1** Pearson correlations between assessments. **Abbreviations:** NL-CLA, normalized low-contrast letter acuity; NMI, normalized Motricity Index; NMAS, normalized Modified Ashworth Scale; NBBS, normalized Berg Balance Scale; NT, normalized tremor; NDD, normalized dysdiadochokinesia; NDM, normalized dysmetria; NPRs, normalized postural reactions (righting, equilibrium, and protective reactions); NKH, normalized knee hyperextension; NNHPT, normalized Nine-Hole Peg Test; NT25FW, normalized Timed 25-Foot Walk; NPASAT 3, normalized 3-minute version of the Paced Auditory Serial Addition Test.

inclusion in the assessment set. Nevertheless, there is not yet general accordance on the validity of this scale.<sup>39</sup> Some studies have reported the MAS to have a moderate to good interrater reliability,<sup>15,40</sup> but most studies have reported poor reliability.<sup>39,40</sup> Furthermore, poor intrarater agreement of the MAS has been confirmed.<sup>33</sup> In other studies, the intrarater reliability of the MAS was found to be either moderate<sup>39</sup> or good.<sup>40</sup> In the present study, the MAS demonstrated good internal consistency (Cronbach’s alpha: 0.78) but poor test–retest reliability (ICC: 0.49). Nevertheless, it is very sensitive to posttreatment changes (corrected *P*-value of <0.001).

Using the Ashworth Scale to evaluate spasticity is controversial because of its weak psychometric properties, as the relationship between spasticity and motor performance has not yet been confirmed. Furthermore, it is an ordinal scale that lacks sensitivity for detection of changes, and it uses constant speed for evaluation of spasticity; however, spasticity was defined as a velocity-dependent response to stretch. The AS not only evaluates spasticity but also passive resistance – the intrinsic properties of muscle, tendon, and connective tissue too.<sup>37</sup> Finally, the AS is not sensitive enough to detect changes in quality of life or functional outcomes.<sup>38,40</sup>

A variety of laboratory techniques and clinical scales have been proposed to evaluate balance,<sup>16</sup> but the instruments most commonly used in the clinical setting are clinical

scales. Clinical scales provide insight for the planning of rehabilitation, are less expensive than laboratory techniques, do not require specific training of raters, and are easily applicable in the clinical setting. Of the clinical scales, the BBS, the Dynamic Gait Index, the Dizziness Handicap Inventory, the Timed Up and Go Test, the Ambulation Index, the Activities-Specific Balance Confidence Scale, the Functional Reach Test, and the Postural Stability Test have gained popularity within the clinical and scientific community for MS.<sup>33,42</sup> The authors selected the BBS, which is the most frequently used test, for inclusion in the assessment set. The BBS was developed to measure balance among older people with impairment in balance function by assessing the performance of functional tasks.<sup>16</sup> The BBS was found to be a valid and reliable instrument in the elderly and post-stroke.<sup>41</sup> Psychometric properties of the BBS have also been evaluated for MS. The BBS shows a good concurrent validity (high specificity), bad discrimination validity (low sensitivity) that does not distinguish well between fallers and nonfallers,<sup>42,43</sup> and good interrater (ICC: 0.96) and test–retest reliability (ICC: 0.96).<sup>44</sup> Results of the present study did confirm good test–retest reliability (ICC: 0.78), very good internal consistency (Cronbach's alpha: 0.94) was also demonstrated.

For evaluation of righting, equilibrium, and protective reactions, the scale for evaluation of PR described by Corriveau et al<sup>20</sup> has been used previously. The present authors selected this evaluation for inclusion in the proposed battery, although this protocol was specially prepared to evaluate therapeutic modality developed by Bobath<sup>44</sup> and quantifiable patient progress in connection with this concept. The present authors are convinced that this protocol is well prepared to evaluate PRs (righting, equilibrium, and protective reactions). This evaluation was validated with the Brunnstrom Scale, the Fugl-Meyer Test, the Upper Extremity Functional Test, and the present pain intensity scale of the McGill Pain Questionnaire. The protocol is sensitive to motor recovery over time. Results of the present study confirmed very good psychometric properties: test–retest reliability (ICC: 0.96) and internal consistency (Cronbach's alpha: 0.92).

To evaluate tremor, accelerometer has been used as the objective method of measurement, and clinical rating systems and patient self-assessments have been used as the subjective methods of measurement.<sup>45</sup> In MS, the Fahn's Tremor Rating Scale<sup>46,47</sup> and the Tremor Rating Scale are most frequently used.<sup>48,49</sup> The Fahn's Tremor Rating Scale was used for evaluation of upper and lower extremities in the present study, as it has good psychometric properties: high interrater reli-

ability for intention tremor (kappa: 0.65–0.74)<sup>45,48</sup> and very good intrarater reliability.<sup>49,50</sup> Results confirm only moderate test–retest reliability (ICC: 0.61) and internal consistency (Cronbach's alpha: 0.74).

Alusi et al<sup>22</sup> described fair to moderate psychometric properties in assessing dysmetria (intrarater reliability: kappa, 0.35–0.45; interrater reliability: 0.40–0.59) and dysdiadochokinesia (intrarater reliability: kappa, 0.47–0.59; interrater reliability: 0.33–0.58). Results of the present study showed poor test–retest reliability (ICC: 0.40) but very good internal consistency of DD (Cronbach's alpha: 0.92). The results also demonstrated poor test–retest reliability (ICC: 0.47) but very good internal consistency of DM (Cronbach's alpha: 0.82).

Genu recurvatum (knee extension greater than 5 degrees) is a common entity found in clinical practice. It is a consequence of poor control over the knee joint due to muscle weakness, impaired tonus, and deficit in joint proprioception. Uncontrolled locking of the knee during ambulation causes recurrent microtrauma, which leads to degenerative changes and instability.<sup>51</sup> However, this is a problem of neurological diseases in general, and research has predominantly involved stroke patients.<sup>52,53</sup> Knee extension can be evaluated using different kinds of goniometers – handheld goniometer, electrogoniometer,<sup>54</sup> gravity-based goniometer,<sup>55</sup> fluid-based inclinometer<sup>56</sup> three-dimensional motion analysis system,<sup>57</sup> or goniometer based on gait analysis.<sup>57</sup> The authors did not have a validated electrogoniometer or a three-dimensional motion analysis system (which would be able to accurately locate the center of knee joint rotation), but a handheld goniometer was available. This is why the authors instead created a KH test that is easy, quick, and targeted to the knee function (standing and walking) – the function that the treatment targets. Results confirmed very good test–retest reliability (ICC: 0.98) and good internal consistency (Cronbach's alpha: 0.85).

To evaluate fine motor skills in MS, the NHPT, the Box and Blocks Test, and the Purdue Pegboard Test are used. The NHPT is the most frequently used, mainly as part of the MSFC. For this reason, the authors selected the NHPT for inclusion in the assessment set. The interrater reliability of the NHPT is high (ICC: 0.84–0.96) and so is its intrarater reliability (ICC: 0.91–0.99).<sup>58</sup> Cutter et al<sup>59</sup> described modest correlation between the 1-year change in the NHPT results and change in the EDSS score ( $r = 0.27$ ). Also, in the present study, results for the NHPT demonstrated very high test–retest reliability (ICC: 0.88) and very high internal consistency (Cronbach's alpha: 0.93).

Many tests that evaluate walking can be found in the literature. Some of these tests are aimed at the measurement

of velocity (10-Meter Walk Test, T25FW test, and Timed Tandem Gait), some of them at walking distance (2- or 6-Minute Walk Test), and some of them at the quality of walking – these tests assess walking as part of complex movement with the aim to change body position (Timed Up and Go Test, Functional Gait Assessment, Dynamic Gait Index, Ambulation Index, Tinetti Assessment Tool – Gait, and Kela Coordination Test).<sup>33</sup> The authors selected the T25FW test for inclusion in the assessment set because it is the most frequently used of the tests in MS, mainly as part of the MSFC. Its psychometric properties are also very good, having high inter- and intrarater reliability.<sup>60,61</sup> The change in the timed walk and the change in EDSS score showed a correlation of  $r = 0.41$ .<sup>60</sup> In the present study, the results indicated very high test–retest reliability (ICC: 0.95) and very high internal consistency (Cronbach’s alpha: 0.96).

To evaluate mental function, the PASAT (2- and 3-minute versions), the Symbol Digit Modalities Test, the Controlled Oral Word Association Test, and the Mental Fatigue Scale are used in MS. The most frequently used is the PASAT 3, as part of the MSFC. This is why the authors selected this test for the assessment set. Solari et al<sup>58</sup> reported high inter- (ICC: 0.9–0.97) and intrarater reliability (ICC: 0.94–0.98) for the PASAT. Similarly, Rosti-Otajärvi et al<sup>60</sup> confirmed very good intra- (0.75–0.96) and interrater reliability (0.68–0.95). The internal consistency of the PASAT is excellent (split-half reliability: 0.96).<sup>5</sup> Also, the results of the present study showed very high test–retest reliability (ICC: 0.92). Besides significant improvement of NPASAT after treatment ( $P = 0.04$ ), there is also some improvement without therapy (mean of 0.07), but this is not significant after correction for multiple comparisons. It is likely that the improvement is the result of the practice effect of patients (increasing familiarity with the test). Cutter et al<sup>59</sup> also described this practice effect.

## Six composite assessments

Many neurological rating scales have been suggested to assess the impact of MS on patients, but none has been universally accepted. The EDSS is based on neurological examination of eight functional systems, usually performed by a neurologist. While problems of standardization, sensitivity (mainly to arm and cognitive changes), reliability, and rater-to-rater variability have been documented, the EDSS remains a useful tool for classifying MS patients by disease severity and has been used extensively to assess disability and its changes in MS.<sup>60</sup>

Whitaker et al<sup>61</sup> emphasized the necessity of developing a new clinical rating scale that would be multidimensional,

to reflect the varied clinical expression of MS across patients and over time, and would be able to register changes over time. Based on analyses of pooled data from natural history studies and from placebo groups in clinical trials, the National Multiple Sclerosis Society’s Clinical Outcomes Assessment Task Force has recently proposed a new multidimensional clinical outcome measure, the MSFC.<sup>62</sup> The MSFC comprises the T25FW test, the NHPT, and the PASAT 3 as a multidimensional test. Scores on component measures are converted to standard scores ( $z$ -scores), which are averaged to form a single MSFC score. The MSFC ( $z$ -score) shows excellent intra- (0.97, 0.97, and 0.99 for the T25FW test, NHPT, and PASAT 3, respectively) and interrater (0.95, 0.96, and 1.0 for the T25FW test, NHPT, and PASAT 3, respectively) reliability,<sup>58,60,62</sup> and it also shows strong evidence of face validity as well as convergent and divergent validity with the EDSS. Further, changes in the MSFC correlate with change in the EDSS (the MSFC change predicted subsequent change in the EDSS).<sup>58,62</sup> Even with the increased variability in the early testing sessions due to the practice effect, the MSFC demonstrated excellent reliability.<sup>62</sup> The MSFC is a very good composite for MS, but it is not optimal – for example, when there are too many variables of which only a few exhibit change, the average shows little change.<sup>59,63,64</sup>

In this study, the authors prepared six composite tests that characterize clinical functions that are important in physiotherapy: normalized left (NLUEF) and right upper extremity function (NRUEF), normalized left (NLLEF) and right lower extremity function (NRLEF), BI, and TICF. The authors found weak test–retest reliability (ICC: 0.39) in NLUEF and moderate test–retest reliability (ICC: 0.68) in NRUEF. In other composite measures, the authors found good test–retest reliability. These composite tests evaluate clinical functions in a complex way. These indexes document well the function of each extremity (muscle tone, strength, coordination, functional ability) and balance (proactive and reactive balance reactions); the TICF is a mathematical expression of the actual status of the MS patient from the therapist’s point of view.

The EDSS is based on neurological examination of eight functional systems, usually performed by a neurologist. The proposed assessment set was created for the clinical practice of a physiotherapist. The power of this assessment set to predict EDSS score should be verified in further study on a larger sample of patients.

## Conclusion

In this study, the following achievements were made:

- Standard outcome measures were prepared for use and validated in the Czech Republic. Sensitivity to posttreatment changes, good test–retest reliability and internal consistency were confirmed.
- The normalization of standard outcome measures was introduced and their importance for orientation in examination results was shown.
- A proposed battery of tests was designed comprising standard outcome measures and one test of the authors' own that objectively and systematically evaluate clinical features of MS treatable by physiotherapy.
- Six composite tests that evaluate function of left and right upper and lower extremities (NLUEF, NRUEF, NLLEF, and NRLEF), balance (BI), and total function (TICF) were introduced.

Based on experience from clinical practice and research, the authors can conclude that this battery of tests and six composite tests is practical to use, is acceptable to patients, is capable of demonstrating effects of rehabilitation, and can be used with confidence to evaluate effects of physiotherapy in MS.

## Acknowledgments

The authors would like to thank the Ministry of Health of the Czech Republic (1A/8628-5), the Ministry of Education, Youth and Sports of the Czech Republic (project 1M06014), RVO:67985807, a European Social Fund project (Involving Training Workplaces for Disabled People), the state budget of the Czech Republic, and the budget of the city of Prague for supporting this study.

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. Thompson AJ. Symptomatic management and rehabilitation in multiple sclerosis. *J Neurol Neurosurg Psychiatry*. 2001;71 Suppl 2:ii22–ii27.
2. Henze T, Rieckmann P, Toyka KV; for Multiple Sclerosis Therapy Consensus Group of the German Multiple Sclerosis Society. Symptomatic treatment of multiple sclerosis. *Eur Neurol*. 2006;56(2):78–105.
3. Khan F, Turner-Stokes L, Ng L, Kilpatrick T. Multidisciplinary rehabilitation for adults with multiple sclerosis. *Cochrane Database Syst Rev*. 2007;2:CD006036.
4. Thompson AJ. Neurorehabilitation in multiple sclerosis: foundations, facts and fiction. *Curr Opin Neurol*. 2005;18(3):267–271.
5. Thompson AJ. The effectiveness of neurological rehabilitation in multiple sclerosis. *J Rehabil Res Dev*. 2000;37(4):455–461.
6. Haigh R, Tennant A, Biering-Sørensen F, et al. The use of outcome measures in physical medicine and rehabilitation within Europe. *J Rehabil Med*. 2001;33(6):273–278.
7. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*. 1983;33(11):1444–1452.
8. McDonald WI, Compston A, Edan G, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the Diagnosis of Multiple Sclerosis. *Ann Neurol*. 2001;50(1):121–127.
9. Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*. 1971;9(1):97–113.
10. Management of substance abuse: process of translation and adaptation of instruments [web page on the Internet]. Geneva, Switzerland: World Health Organization; 2012. Available from: [http://www.who.int/substance\\_abuse/research\\_tools/translation/en/index.html](http://www.who.int/substance_abuse/research_tools/translation/en/index.html). Accessed March 15, 2011.
11. Baier ML, Cutter GR, Rudick RA, et al. Low-contrast letter acuity testing captures visual dysfunction in patients with multiple sclerosis. *Neurology*. 2005;64(6):992–995.
12. Demeurisse G, Demol O, Robaye E. Motor evaluation in vascular hemiplegia. *Eur Neurol*. 1980;19(6):382–389.
13. Collin C, Wade D. Assessing motor impairment after stroke: a pilot reliability study. *J Neurol Neurosurg Psychiatry*. 1990;53(7):576–579.
14. Elovic E, Baerga E. Associated topics in physical medicine and rehabilitation: spasticity. In: Cuccurullo SJ, editor. *Physical Medicine and Rehabilitation Board Review*. New York: Demos Medical Publishing; 2004:743–750.
15. Bohannon RW, Smith MB. Interrater reliability of a modified Ashworth Scale of muscle spasticity. *Phys Ther*. 1987;67(2):206–207.
16. Berg K, Wood-Dauphinee S, Williams JL. The Balance Scale: reliability assessment with elderly residents and patients with an acute stroke. *Scand J Rehab Med*. 1995;27(1):27–36.
17. *Berg Balance Scale*. American Academy of Health and Fitness. Available from: [http://www.aahf.info/pdf/Berg\\_Balance\\_Scale.pdf](http://www.aahf.info/pdf/Berg_Balance_Scale.pdf). Accessed February 19, 2009.
18. Guarna F, Corriveau H, Chamberland J, Arsenault AB, Dutil E, Drouin G. An evaluation of the hemiplegic subject based on Bobath approach: Part I. The model. *Scand J Rehab Med*. 1988;20(1):1–4.
19. Davies PM. *Steps to Follow: A Guide to the Treatment of Adult Hemiplegia; Based on the Concept of K. and B. Bobath*. Berlin: Springer-Verlag; 1993.
20. Corriveau H, Guarna F, Dutil E, Riley E, Arsenault AB, Drouin G. An evaluation of the hemiplegic subject based on the Bobath approach: Part II. The evaluation protocol. *Scand J Rehabil Med*. 1988;20(1):5–11.
21. Fahn S, Tolosa E, Marin C. Clinical rating scale for tremor. In: Jankovic J, Tolosa E, editors. *Parkinson's Disease and Movement Disorders*. 2nd ed. Baltimore (MD): Williams and Wilkins; 1993:271–280.
22. Alusi SH, Worthington J, Glickman S, Findley LJ, Bain PG. Evaluation of three different ways of assessing tremor in multiple sclerosis. *J Neurol, Neurosurg Psychiatry*. 2000;68(6):756–760.
23. Coulthard-Morris L. Clinical and rehabilitation outcome measures. In: Burks JS, Johnson KP, editors. *Multiple Sclerosis: Diagnosis, Medical Management, and Rehabilitation*. New York: Demos Medical Publishing; 2000:236–290.
24. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420–428.
25. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B*. 1995;57(1):289–300.
26. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2012. Version 2.14.2 (2012-02-29), Available from: <http://www.R-project.org/>. Accessed March 4, 2012.
27. Revelle W. *Package Psych: Procedures for Personality and Psychological Research*; 1.0–92 ed. Evanston: Northwestern University; 2010. Available from: <http://www.personality-project.org/r/book/psych.pdf>. Accessed August 9, 2012.
28. Cameron D, Bohannon RW. Criterion validity of lower extremity Motricity Index scores. *Clin Rehabil*. 2000;14(2):208–211.
29. Gregson JM, Leathley MJ, Moore AP, Smith TL, Sharma AK, Watkins CL. Reliability of measurements of muscle tone and muscle power in stroke patients. *Age Ageing*. 2000;29(3):223–228.

30. Nuyens G, De Weerd W, Ketelaer P, et al. Inter-rater reliability of the Ashworth Scale in multiple sclerosis. *Clin Rehabil*. 1994;8(4):286–292.
31. Blum L, Korner-Bitensky N. Usefulness of the Berg Balance Scale in stroke rehabilitation: a systematic review. *Phys Ther*. 2008;88(5):559–566.
32. World Health Organization (WHO). *International Classification of Functioning, Disability and Health: ICF*. Geneva, Switzerland: WHO; 2001.
33. Paltamaa J, West H, Sarasoja T, Wikström J, Mälkiä E. Reliability of physical functioning measures in ambulatory subjects with MS. *Physiother Res Int*. 2005;10(2):93–109.
34. Balcer LJ, Baier ML, Pelak VS, et al. New low-contrast vision charts: reliability and test characteristics in patients with multiple sclerosis. *Mult Scler*. 2000;6(3):163–171.
35. Bohannon RW, Andrews AW. Standards for judgments of unilateral impairments in muscle strength. *Percept Mot Skills*. 1999;89(3 Pt 1): 878–880.
36. Damiano DL, Quinlivan JM, Owen BF, Payne P, Nelson KC, Abel MF. What does the Ashworth Scale really measure and are instrumented measures more valid and precise? *Dev Med Child Neurol*. 2002;44(2): 112–118.
37. Platz T, Eickhof C, Nuyens G, Vuadens P. Clinical scales for the assessment of spasticity, associated phenomena, and function: a systematic review of the literature. *Disabil Rehabil*. 2005;27(1–2):7–18.
38. Dario A, Tomei G. Management of spasticity in multiple sclerosis by intrathecal baclofen. *Acta Neurochir Suppl*. 2007;97(Pt 1):189–192.
39. Ansari NN, Naghdi S, Moammeri H, Jalaie S. Ashworth scales are unreliable for the assessment of muscle spasticity. *Physiother Theory Pract*. 2006;22(3):119–125.
40. Ansari NN, Naghdi S, Arab TK, Jalaie S. The interrater and intrarater reliability of the Modified Ashworth Scale in the assessment of muscle spasticity: limb and muscle group effect. *Neuro Rehabilitation*. 2008;23(3):231–237.
41. Hobart JC, Riazi A, Thompson AJ, et al. Getting the measure of spasticity in multiple sclerosis: the Multiple Sclerosis Spasticity Scale (MSSS-88). *Brain*. 2006;129(Pt 1):224–234.
42. Cattaneo D, Jonsdottir J, Repetti S. Reliability of four scales on balance disorders in persons with multiple sclerosis. *Disabil Rehabil*. 2007;29(24):1920–1925.
43. Cattaneo D, Regola A, Meotti M. Validity of six balance disorders scales in persons with multiple sclerosis. *Disabil Rehabil*. 2006;28(12):789–795.
44. Arsenault AB, Dutil E, Lambert J, Corriveau H, Guarna F, Drouin G. An evaluation of the hemiplegic subject based on the Bobath approach: Part III. A validation study. *Scand J Rehabil Med*. 1988;20(1):13–16.
45. Bain PG, Findley LJ, Atchison P, et al. Assessing tremor severity. *J Neurol Neurosurg Psychiatry*. 1993;56(8):868–873.
46. Feys P, D'hooghe M, Nagels G, Helsen WF. The effect of levetiracetam on tremor severity and functionality in patients with multiple sclerosis. *Mult Scler*. 2009;15(3):371–378.
47. Feys PG, Davies-Smith A, Jones R, et al. Intention tremor rated according to different finger-to-nose test protocols: a survey. *Arch Phys Med Rehabil*. 2003;84(1):79–82.
48. Plaha P, Khan S, Gill SS. Bilateral stimulation of the caudal zona incerta nucleus for tremor control. *J Neurol Neurosurg Psychiatry*. 2008; 79(5):504–513.
49. Bryant JA, De Salles A, Cabatan C, Frysinger R, Behnke E, Bronstein J. The impact of thalamic stimulation on activities of daily living for essential tremor. *Surg Neurol*. 2003;59(6):479–484.
50. Hooper J, Taylor R, Pentland B, Whittle IR. Rater reliability of Fahn's tremor rating scale in patients with multiple sclerosis. *Arch Phys Med Rehabil*. 1998;79(9):1076–1079.
51. Loudon JK, Goist HL, Loudon KL. Genu recurvatum syndrome. *J Orthop Sports Phys Ther*. 1998;27(5):361–367.
52. Basaglia N, Mazzini N, Boldrini P, Bacciglieri P, Contenti E, Ferraresi G. Biofeedback treatment of genu-recurvatum using an electrogoniometric device with an acoustic signal: one-year follow-up. *Scand J Rehabil Med*. 1989;21(3):125–130.
53. Trueblood PR, Walker JM, Perry J, Gronley JK. Pelvic exercise and gait in hemiplegia. *Phys Ther*. 1989;69(1):18–26.
54. Chao EY, Laughman RK, Schneider SS, Stauffer RN. Normative data of knee joint motion and ground reaction forces in adult level walking. *J Biomech*. 1983;16(3):219–233.
55. Ekstrand J, Wiktorsson M, Oberg B, Gillquist J. Lower extremity goniometric measurements: a study to determine their reliability. *Arch Phys Med Rehabil*. 1982;63(4):171–175.
56. Rheault W, Miller M, Nothnagel P, Straessle J, Urban D. Intertester reliability and concurrent validity of fluid-based and universal goniometers for active knee flexion. *Phys Ther*. 1988;68(11):1676–1678.
57. Pomeroy VM, Evans E, Richards JD. Agreement between an electrogoniometer and motion analysis system measuring angular velocity of the knee during walking after stroke. *Physiotherapy*. 2006; 92(3):159–165.
58. 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(1):71–74.
59. Cutter GR, Baier ML, Rudick RA, et al. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain*. 1999;122(Pt 5):871–882.
60. Rosti-Otajärvi E, Hämäläinen P, Koivisto K, Hokkanen L. The reliability of the MSFC and its components. *Acta Neurol Scand*. 2008;117(6): 421–427.
61. Whitaker JN, McFarland HF, Rudge P, Reingold SC. Outcomes assessment in multiple sclerosis clinical trials: a critical analysis. *Mult Scler*. 1995;1(1):37–47.
62. Rudick RA, Cutter G, Baier M, et al. Use of the Multiple Sclerosis Functional Composite to predict disability in relapsing MS. *Neurology*. 2001;56(10):1324–1330.
63. Cohen JA, Fischer JS, Bolibrush DM, et al. Intrarater and interrater reliability of the MS functional composite outcome measure. *Neurology*. 2000;54(4):802–806.
64. Syndulko K, Ke D, Ellison GW, Baumhefner RW, Myers LW, Tourtellotte WW; for Multiple Sclerosis Study Group. Comparative evaluations of neuroperformance and clinical outcome assessments in chronic progressive multiple sclerosis: I. Reliability, validity and sensitivity to disease progression. *Mult Scler*. 1996;2(3):142–156.

## Patient Related Outcome Measures

### Publish your work in this journal

Patient Related Outcome Measures is an international, peer-reviewed, open access journal focusing on treatment outcomes specifically relevant to patients. All aspects of patient care are addressed within the journal and practitioners from all disciplines are invited to submit their work as well as healthcare researchers and patient support groups. The manuscript

Submit your manuscript here: <http://www.dovepress.com/patient-related-outcome-measures-journal>

Dovepress

management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.