Validity of six balance disorders scales in persons with multiple sclerosis

DAVIDE CATTANEKO, ALBERTO REGOLA, & MATTEO MEOTTI

LaRiCE: Gait and Balance Disorders Laboratory, Servizio riabilitazione neurologica adulti (Int. 282); Don Gnocchi Foundation, Milano, Italy

Accepted October 2005

Abstract

Purpose. The aim of this study was to test concurrent and discriminant validity of several tests of static and dynamic balance in a sample of subjects suffering from MS.

Method. A group of 51 patients were enrolled in the study. The following tests were administered: Berg Balance Scale (BBS), Timed Up and Go Test (TUG), Dynamic Gait Index (DGI), Hauser Deambulation Index (DI), Dizziness Handicap Inventory (DHI), and Activities-specific Balance Confidence (ABC). The scales used in this study were initially translated into Italian.

Results. The sample of subjects reported a mean of 0.98 (1.8 SD) falls in the month prior to evaluation. The tests demonstrated good concurrent validity: Higher correlation coefficients among tests tapping the same aspect and lower correlation coefficients among tests tapping slightly different aspects. ABC and DHI tests discriminated better than the others between fallers and non-fallers and appeared the best predictors of fall status. BBS and DGI were not as efficient in discriminating between groups. Conversely, all tests showed good difference validity in the prediction of patients who used an assistive device.

Conclusion. BBS, TUG, DI, DGI, ABC, DHI have acceptable concurrent validity. The scales have poor performance in discriminating between faller and non-faller.

Keywords: Balance, validity, scales, multiple sclerosis

Introduction

Multiple sclerosis (MS) is the commonest progressive neurologic disease in young adults [1]. In Italy MS has a prevalence of 30 cases per 100,000 inhabitants and an incidence of 2.2–3.7 new cases per year [2].

Abnormalities in balance control are common findings in subjects suffering from multiple sclerosis [3–6]. This abnormality, along with other risk factors, increases risk of falls [7] preventing the patients from performing their daily life activities. In the past decade much attention has been directed toward the rehabilitation of balance in elderly subjects. However, the frequency of falls in this population was found to be 1.5 falls per hospital bed per year [8], while another study looking at a mixed population of MS inpatients and outpatients showed a mean of 9 falls per year [7]. Assessment of different aspects of motor impairment and accurate determination of factors contributing to falls are necessary for the development of a program for falls prevention [9–11]. Since falls among patients with MS have a multifactorial etiology, multiple assessments are necessary to explore the balance system in order to classify patients according to their degree of impairment, disability, handicap or quality of life.

Although several scales have been developed to assess functional level of patients with MS [12,13] the possibility of using frequently administered scales also in this group of patients allows comparison of the results with other studies and with other pathologies. Moreover, the use of frequently administered scales does not require specific training of raters. Over the past decade, balance-related instruments such as Berg Balance Scale, Dynamic Gait Index and Timed Up & Go Test have become
increasingly popular in clinical studies. Despite of the high incidence of falls in the MS population no study reports the validity of those instruments for this population of patients.

The aim of this study was to test concurrent and discriminant validity of several tests on static and dynamic balance in a sample of subjects suffering from MS.

Method

Subjects

A total of 63 patients suffering from MS attending the Department of Multiple Sclerosis of Don Gnocchi Foundation were assessed. To be eligible for the study, the patients had to meet the following inclusion criteria: Clinically or laboratory definite relapsing-remitting or secondary progressive multiple sclerosis, ability to stand independently in upright position more than 3 seconds, ability to walk for 6 m even with an assistive device.

We excluded patients with cognitive impairments that might hinder understanding of the tasks to be accomplished. Patients with visual problem and impairment of VIII cranial nerve were not excluded.

A sub-group of 51 patients met the inclusion-exclusion criteria and were enrolled in the study. The group consisted of 16 males and 35 females, mean age 45.3 years (standard deviation, SD, 18.1 years). The onset of pathology was 15.6 years (7.6 years SD) before the beginning of the study. Fifteen patients used a walking aid in their daily activities. After informed consent was obtained, patients completed a questionnaire providing information about their age, the onset of pathology, and the number of falls one month before the assessment procedure. A fall was defined as any event that led to an unexpected contact with a support surface. Due to the natural course of the pathology we restricted the falls reported by patient to one month prior to evaluation in order to obtain a reliable picture of patients’ characteristics. Patients were tested wearing their normal shoes. The assessment was carried out in one session, the testing protocol took 30–40 min to be administered, patients were allowed to rest during testing if necessary. The assessment protocol consisted of 1 test on static balance, 3 tests on dynamic balance and 2 self-reported scales on balance confidence and level of handicap.

Instruments

Berg Balance Scale (BBS). The scale rates performance from 0 (cannot perform) to 4 (normal performance) on 14 items. The items explore the ability to sit, stand, lean, turn and maintain the upright position on one leg [14]. The psychometric properties of the scale have been assessed on populations of elderly subjects. In those groups of subjects the scale proved to be a valid and reliable instrument [15–17]. The intrarater and interrater reliability of BBS were very high, the ICC ranged from 0.98–0.99 for intrarater reliability and 0.98 for interrater reliability [16,18]. Moreover the BBS was found to be a good predictor of falls in a cohort of elderly subjects [19,20]. A cut-off score of 45 is an established criterion to identify elderly subjects with high risk of fall [21].

Timed Up and Go Test (TUG). The test is a measure of dynamic balance. It requires the subjects to stand up from a chair, walk 3 m, turn around and be seated. The subject is timed from the moment he lifts the pelvis from the chair until he returns with the pelvis in the chair [22]. The validity and reliability of the instrument was tested on a cohort of elderly subjects and patients with unilateral lower limb amputation. The results showed good correlation between TUG and BBS (r = 0.81) and moderate correlation between TUG and Groningen Activity Restriction Scale (r = 0.39 GARS) [22]. The tests showed also good interrater (0.96) and intrarater (0.93) reliability. Times of 13.5 seconds or greater have been related to increased risk of falling in older adults [22].

Dynamic Gait Index (DGI). The scale measures the mobility function and the dynamic balance. The eight tasks of this scale include walking, walking with head turns, pivoting, walking over objects, walking around objects and going up stairs. The performance is rated on a 4-point scale. Whitney found a correlation between the scores of the Dynamic Gait Index and the Berg Balance Scale of 0.71 in persons with vestibular disorders [23]. In a similar group of patient a score of 19 or less has been shown to be related to self-reported number of falls in persons with vestibular disorders [24]. Data on reliability have already been published in a population of MS patients and showed both good inter and intrarater reliability [25].

Hauser Deambulation Index (DI). The index rates gait performance. It has been used in several studies focusing on patients suffering from MS [26,27]. It has 10 grades ranging from 0–9; a grade of 0 means “no gait impairment”; a grade of 10 means “Restricted to wheelchair”.

Dizziness Handicap Inventory (DHI). It is a multidimensional self-assessment scale that quantifies the level of disability and handicap in three subscales: Physical, emotional and functional. It is possible to use both the sum score and the scores of the three
subscales separately. Scores range from 0–100; where 100 means high level of disability and handicap. DHI demonstrated good internal consistency (0.91) and test retest reliability (0.97) [28].

**Activities-specific Balance Confidence (ABC).** It is a scale in which the patient rates his perceived level of confidence while performing 16 daily living activities [29]. The Test-retest reliability of the ABC Scale among people who have a lower-limb amputation was 0.91; moreover, the associations with the 2 Meters Walking Test and TUG test were observed with correlations of 0.72 and −0.70 [30]. A study aimed to assess convergent validity between ABC Scale and Falls Efficacy Scale (FES) in a population of elderly people indicated the ABC and FES were highly correlated [31].

The scales used in this study were initially translated by two physical therapists proficient in English whose native language was Italian. The two translations were compared and when differences were identified, the texts were modified to obtain consensus between the two translations. Problematic items and response choices were clarified contacting the authors of the scales. The final version was also given to four physical therapists, so that they could apply the scale to patients in order to test the intelligibility of the items. The Italian versions of the scales are available from the authors on request.

**Data analysis**

Descriptive statistics were used to describe the study population in terms of demography and disease characteristics and number of falls.

To study the ceiling effect, the percentages of the sample scoring the maximum possible scores were computed. Those percentages reflect the extent that scores cluster at the top of the scale range. Ceiling effects >20% was considered to be significant.

Group differences validity was assessed by testing the extent to which the scores of these scales differentiated between patients using or not an assistive device to walk and between fallers and non-fallers. Mann-Whitney U Test was used to determine group differences. Logistic regression was carried out to determine the cut-off value that predicted a probability to be classified as a faller of 0.5 or greater.

Sensitivity and specificity in predicting falls status were calculated. The sensitivity was defined as the percentage of the fallers who were correctly identified. Alternately, specificity was defined as the percentage of non-fallers who were correctly identified.

Concurrent validity was tested by assessing the degree to which each scale correlated with the other scales, by Spearman correlation coefficient. Correlation coefficients of 0.35–0.49 were interpreted as weak, those of 0.5–0.79 as moderate and those of 0.8 or greater as strong.

**Results**

The whole sample of subjects reported a mean of 0.98 (1.8 SD) falls in the month prior to evaluation. The faller group consisted of 20 subjects that accounted for 39% of the sample with a mean of 2.34 (2.1 SD) falls/month.

Ceiling effects occurred respectively 4 times in the DI and DGI tests (7.8%), 3 times in the BBS (5.8%), 1 time for the DHI (1.9%), no ceiling effect was observed for the ABC test.

Table I shows the test scores for the whole group and those for the faller and non-faller groups.

The whole group of subjects reported moderate impairment in static balance (BBS scale) and more pronounced impairment in dynamic balance (DGI scale). The self-administered tests showed marked low confidence of patients in their balance skills. With respect to the three domains of DHI the emotional domain was the less impaired, with a score reaching 41% of the maximum possible score. Conversely, physical and functional domains were more impaired with scores reaching respectively 50.8% and 51.6% of the maximum possible score.

| Mean and SD scores of the tests for the whole group of patients and the subgroup of fallers and non-fallers. |
|---|---|---|---|---|
| **Whole group** | **Non-fallers** | **Fallers** | **Fall vs. non-fall** |
| **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** | **p** |
| BBS | 47.5 | 6.0 | 48.9 | 5.6 | 45.3 | 6.2 | 0.042 |
| DGI | 15.5 | 5.3 | 16.9 | 5.0 | 13.3 | 5.2 | 0.025 |
| TUG (s) | 2.6 | 0.5 | 2.5 | 0.5 | 2.7 | 0.6 | 0.650 |
| DI | 2.6 | 1.5 | 2.4 | 1.5 | 3.0 | 1.5 | 0.228 |
| ABC | 51.4 | 25.8 | 61.1 | 25.3 | 36.9 | 19.2 | 0.001 |
| DHI | 45.5 | 22.8 | 38.5 | 22.8 | 56.0 | 18.8 | 0.009 |

BBS, Berg Balance Scale; DGI, Dynamic Gait Index; TUG, Timed Up & GO Test (s, seconds); DI, Deambulation Index; ABC, Activities-specific Balance Confidence; DHI, Dizziness Handicap Inventory; p column reports the p level of the differences between fallers and non-fallers (U test).
A statistically significant difference of three points in BBS and DGI were observed between fallers and non-fallers. DI did not show a statistically significant difference between groups. TUG showed no statistical or clinical significant difference between groups. Marked statistically significant differences in ABC and DHI tests were observed between fallers and non-fallers.

Figure 1 shows the means for each item of the BBS and the DGI tests. Since the inclusion-exclusion criteria Item 3 of BBS (Sitting unsupported) was always scored as 4, also Item 1 (sitting to standing), 2 (standing unsupported) and 5 (transfer) showed poor variability among subjects and were almost always scored as 4. Items 6–10 showed more variability among subjects. The scores were lowest for the last 4 items of the scale. Those items tap the ability to keep the balance on a narrower base of support and during weight shifting.

DGI showed greater impairment in dynamic balance with respect to static balance. None of the items of the scale reached the maximum score; moreover the variability of the items among patients was higher with respect to BBS. The scores of the items varied around a mean value of 1.9 points with the exception of item number 7 (walking around objects) and 3 (walking with head turning). The former was the easiest while the latter was the most difficult (1.5 points).

Group differences between fallers and non-fallers are shown in Table I. ABC and DHI tests discriminated better than BBS and DGI tests between groups. Hauser Ambulation Index (DI) and Timed Up and Go Test did not discriminate between the two groups.

Logistic regression was used to determine the cut-off value sensitivity and specificity for the tests discriminating between groups (Table II). Generally sensitivity was poor, while the ABC test showed the best psychometric characteristics. Specificity was good especially for the Berg Balance Scale.

All tests were able to discriminate between subjects who used an assistive device and subjects who did not use it (p values were always less than 0.001).

Concurrent validity among tests is depicted in Table III. Berg Balance Scale showed moderate

![Figure 1. Mean scores of the Items of the Berg Balance Scale (BBS) and the Dynamic Gait Index (DGI). Means and standard deviations (T bars) of the Items of Balance Scale (BBS) and the Dynamic Gait Index (DGI) ordered by Item score.](image-url)
Table II. Cut-off scores, sensitivity and specificity of tests.

<table>
<thead>
<tr>
<th>Test</th>
<th>Cut-off</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBS</td>
<td>&gt;44</td>
<td>40</td>
<td>90</td>
</tr>
<tr>
<td>DGI</td>
<td>&gt;12</td>
<td>45</td>
<td>80</td>
</tr>
<tr>
<td>ABC</td>
<td>&gt;40</td>
<td>65</td>
<td>77</td>
</tr>
<tr>
<td>DHI</td>
<td>&lt;59</td>
<td>50</td>
<td>74</td>
</tr>
</tbody>
</table>

BBS, Berg Balance Scale; DGI, Dynamic Gait Index; ABC, Activities-specific Balance Confidence; DHI, Dizziness Handicap Inventory.

Table III. Spearman correlation coefficients among tests.

<table>
<thead>
<tr>
<th></th>
<th>BBS</th>
<th>DGI</th>
<th>TUG</th>
<th>DI</th>
<th>ABC</th>
<th>DHI</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBS</td>
<td>1.00</td>
<td>-0.72</td>
<td>-0.70</td>
<td>0.48</td>
<td>-0.32</td>
<td></td>
</tr>
<tr>
<td>DGI</td>
<td>1.00</td>
<td>-0.72</td>
<td>-0.80</td>
<td>0.54</td>
<td>-0.39</td>
<td></td>
</tr>
<tr>
<td>TUG</td>
<td>1.00</td>
<td>0.74</td>
<td></td>
<td>-0.38</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>DI</td>
<td>1.00</td>
<td>0.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABC</td>
<td>1.00</td>
<td>0.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DHI</td>
<td>1.00</td>
<td>0.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

BBS, Berg Balance Scale; DGI, Dynamic Gait Index; TUG, Timed Up & Go Test (s, seconds); DI, Deambulation Index; ABC, Activities-specific Balance Confidence; DHI, Dizziness Handicap Inventory; All correlation coefficients were statistically significant (p < 0.001).

correlation with dynamic tests especially with the Dynamic Gait Index. Low correlation between Berg Balance Scale and self-assessment scales was observed. The dynamic balance tests showed moderate to strong correlations among them. ABC and DHI test correlated moderately (−0.7).

Discussion

The aim of this study was to assess the psychometric properties of well-known scales on balance to promote functional evaluation of this construct also in patients suffering from MS.

Our results highlight the incidence of balance disorders in patients suffering from MS. Our data are in agreement with those of other studies that emphasize balance disorder in this pathology [3,5]. Those disorders in the MS population lead to an increase in the number of falls far beyond those which are reported for elderly patients [8].

In this study, the frequency of falls was slightly higher with respect to our previous findings [7]; this difference could be due to differences in the sample characteristics.

Falls are due to the sum of multiple impairment; this must be appreciated when evaluating a patient with balance disorder. With respect to static balance patients showed poorer performance in the items requiring a small base of support and weight shifting and in tasks involving head or whole body rotation.

Some items seem to have poor discriminating properties because too easy for this population of patients.

Similarly to static balance, dynamic balance was impaired especially during head rotation in the sagittal and horizontal planes. Those findings, combined with the ones obtained with measurement of head motion during walking [32], suggest impairment in trunk-head control and/or difficulty to maintain the balance when visuo-vestibular information is challenged by head movements.

Although the four tests discriminate between fallers and non-fallers, the ability to categorize the subjects in two groups was poor. The tests showed low sensitivity. ABC and DHI discriminated better than the other tests between fallers and non-fallers and appear the best predictors of fall status. The scores of those tests were substantially worse for the faller group than for the non-faller group. ABC scored 65% higher and DHI 31% lower in the faller group than in the non-faller group.

Surprisingly, BBS and DGI were not very efficient in discriminating between groups compared with other pathologies. Even taking into account the difference in the population, the Equiscale Test showed better performance in discriminating between fallers and non-fallers. BBS showed the highest specificity but the lowest sensitivity. This finding correlated with that of Thorbahn and Newton [20] which found a sensitivity of 53% and a specificity of 96%. Conversely, Shumway-Cook et al. [19] in a prospective study found values of 77% for sensitivity and 86% for specificity.

The cut-off score was lower (44 points) than that reported in the literature (45 points), however cut-off points varied on the basis of different parameters. The cut-off point for BBS in an elderly population varied from 47–38 depending on whether the patients were categorized as fallers with 1 fall or more than one falls within 6 months prior to the study [33]. DGI showed performance similar to BBS.

In contrast with other studies on elderly subjects [33,34] and patients with unilateral lower limb amputation [22], the TUG test was not able to discriminate between fallers and non-fallers in this population of patients. The mean difference between the two groups of patients was just 0.2s. The difference in the discriminating power between DGI and TUG may be due to the DGI having multiple items on a wider spectrum of activities; moreover the DGI also rates the performance of the movements.

No total score ceiling effect was observed on any of the scales used in the study.

Low sensitivity ceiling effect was observed on any of the scales used in the study.

Validit of balance scales in the MS population
group may be due to different factors not assessed by the test. In order to predict falls in this population of patients other factors have to be considered prior to the planning of a prospective study.

Conversely all tests showed good difference validity in the prediction of patients using an assistive device. This confirms the validity of the tests as a measure of physical mobility in patients with MS.

The tests demonstrated a good concurrent validity and the pattern of correlations among tests was consistent with what expected: Higher correlation coefficients between tests tapping the same aspect (e.g. dynamic balance) and lower correlation coefficients among tests tapping slightly different aspects (e.g. static balance test vs. test inferring balance confidence and psychological concerns).

In our study we targeted subjects who had moderate impairment in balance skills.

Although a number of variables distinguished between fallers and non-fallers and between patients who required an assistive device to walk, the relationships may not be causal.

Conclusion

BBS, TUG, DI, DGI, ABC, DHI have acceptable concurrent validity. Tasks requiring head or body rotation appeared to be the most difficult tasks for patients suffering from MS. Caution should be used in discriminating between fallers and non-fallers on the basis of test scores.

References


