J Rehabil Med 2011; 43: 445–453

ORIGINAL REPORT

UNIFIED BALANCE SCALE: CLASSIC PSYCHOMETRIC AND CLINICAL PROPERTIES

Fabio La Porta, MD[1], Marco Franceschini, MD[2], Serena Caselli, PT[1], Sonia Susassi, PT[1], Paola Cavallini, PT[1] and Alan Tennant, PhD[3]

From the [1]Rehabilitation Medicine Unit, AUSL Modena, Modena, [2]Neuro-Rehabilitation Units IRCCS San Raffaele Pisana, Roma, Italy and [3]Department of Rehabilitation Medicine, Faculty of Medicine and Health, University of Leeds, UK

***Objective:* To evaluate the classic psychometric and clinical profile of the Unified Balance Scale, a novel Rasch-based measure of balance.**

***Methods:* The Unified Balance Scale was administered to 219 neurological patients (providing 302 observations) admitted to rehabilitation, together with: Timed Up & Go, 10-meters walking test, Functional Ambulation Classification (FAC), Walking Handicap Scale, FIM™, Trunk Control Test, Motricity Index, and posturographic indexes. Analyses included: concurrent validity, external construct validity (convergent, divergent and discriminant validity), responsiveness, interpretability, predictive validity and usability.**

***Results:* External construct validity (e.g. correlation with FAC: rho = 0.80; with the motor FIM™: rho = 0.55), adequate responsiveness (effect size 1.13), interpretability (the relationship of Unified Balance Scale scores with those of the originating scales and, indirectly, with the risk of falling), and, finally, predictive validity (e.g. relative risk of nursing home admission: 4.33 (95% confidence interval 2.43–7.73) for Unified Balance Scale scores ≤ 2 on admission) were demonstrated for the Unified Balance Scale. Analysis of usability suggested a mean administration time of 20–30 min.**

***Conclusion:* Although further studies are needed to generalize these results to different samples, to confirm its fall risk estimation capabilities and to improve its usability, the Unified Balance Scale presents itself as a psychometrically sound outcome measurement tool to evaluate the effectiveness both of fall reduction plans and of rehabilitation interventions aimed at improving balance.**

Key words: postural balance; accidental falls; outcome measures; rehabilitation; neurological disorders; psychometrics.

J Rehabil Med 2011; 43: 445–453

Correspondence address: Fabio La Porta, Medicina Riabilitativa, Nuovo Ospedale Civile "S. Agostino-Estense", Via Giardini 1355, IT-41126, Modena, Italy. E-mail address: fabiolaporta@mail.com

Submitted September 30, 2010; accepted February 4, 2011

INTRODUCTION

As measurement scales are increasingly used today as outcome measures both in rehabilitation and in allied disciplines (1, 2) it is necessary to follow a systematic approach to ensure that the numbers produced by these tools are valid measures of the variables they claim to quantify, if clinical and research inferences are to be made from these tools (2–4).

In an earlier paper we presented the Unified Balance Scale (UBS), a new activity-based, bed-to-community and aetiology-independent measurement scale for balance within the neuro-rehabilitation setting (5). We constructed UBS by merging items from 3 existing scales of balance (the Berg Balance Scale (BBS), the Performance Oriented Mobility Assessment (POMA) and the Fullerton Advanced Balance Scale (FAB)), and its validity, reliability and measuring properties were demonstrated under the framework of Rasch analysis, which provided support for the internal construct validity of the instrument. Once the latter has been established, it becomes important to evaluate other properties, such as external validity, responsiveness, interpretability of the measures and usability (3, 4) for a safe and feasible implementation of the new measuring tool, both in the clinical and the research setting.

Thus, this paper reports on the evaluation of the psychometric and clinical profile (external validity, responsiveness, interpretability and usability) of the UBS.

METHODS

Patients and setting

Full details of the patients and setting are given elsewhere (5). In brief, data were collected within a Rehabilitation Unit in an Italian general hospital from April 2007 to June 2009. All patients with a neurological disease requiring rehabilitation admitted to the unit as in- or out-patients were included in the study. For inclusion, patients needed to be able to sit unsupported for 30 s without using their upper limbs or to participate, even minimally, in transfers (5). Where possible, inpatients were assessed twice (on admission or as soon as the inclusion criteria were satisfied and, respectively, upon discharge from the unit), for responsiveness evaluation purposes. All observations were collected by 4 raters on a convenience sample of 217 patients. For 85 inpatients pre-treatment and post-treatment observations were available, making a total sample of 302 observations available for analysis.

All patients gave their informed consent to take part in the study that was undertaken in compliance with the ethical principles set out in the Declaration of Helsinki (6).

Tools administered and administration guidelines

The BBS, the Tinetti Balance (TB) and Tinetti Walking (TW) scales (subscales of the POMA) and the FAB were administered to all patients. The following additional instruments were also administered for external validity purposes: the Timed Up & Go test (TUG); the 10-mt walking test (10mtWT); the Functional Ambulation Classification (FAC); the Walk-

 doi: 10.2340/16501977-0800

 ISSN 1650-1977

ing Handicap Scale (WHS); both the motor and the cognitive subscales of the FIM™; the Trunk Control Test (TCT); the Motricity Index (MI). Furthermore, as more able patients underwent balance training with the Biodex Balance System SD™ (Biodex Medical Systems Inc., New York, USA), patients' performances on this device's static (Balance Biodex™ postural stability test (BBPS)) and dynamic (Balance Biodext™ fall risk test (BBFR)) posturographic tests were measured within 72 h of the administration of the other tools. Additional variables recorded included both length of stay (LOS) and discharge destination for inpatients.

Details of the raters' training and the administration protocol are given elsewhere (5). The latter, and the scoring guidelines, are available on request from the corresponding author.

Exploration of the psychometric and clinical profile of Unified Balance Scale

After data collection, the UBS was calibrated, merging items from BBS, POMA and FAB, as described previously (5). Assumptions of normality were tested for UBS measures, TUG, 10mtWT and the BBPS and BBFR tests by using the Kolgomorov-Smirnov statistic. If the latter was not significant (normality assumption met), the appropriate parametric statistics were employed. In all other cases, or if the scale was ordinal, the appropriate non-parametric statistics were employed.

The following classical psychometric and clinical properties for UBS were tested:

- *Concurrent criterion validity* (4). In this context, it was tested by examining the correlation of UBS total scores with the total scores of the originating balance scales (BBS, TB, FAB), for which "very strong" correlations were expected.
- *External construct validity* (4). Three types of external construct validity were examined:
  - *Convergent validity* (7) was tested by examining the correlation of UBS with other scales and indicators related to walking (TW; FAC; WHS; item walking of motor FIM™; 10mtWT), functional mobility-balance (TUG), and motor independence (motor FIM™). All these correlations were expected to be from "strong" to "moderate".
  - *Divergent validity* (7) was tested by examining the correlation of UBS with indicators of trunk control (TCT), cognitive independence (cognitive FIM™), limbs motricity (MI) and balance at the impairment level, represented by BBFR and BBPS. These correlations were expected to be from "moderate" to "weak".
  - *Group differences or discriminating validity,* aiming at demonstrating that UBS was able to detect differences in groups known to differ in the quantity of balance. In particular, it was hypothesized that if UBS were a measure of balance, significantly higher UBS measures should be found in a variety of conditions, such as: full trunk control (TCT > 100); modified independence or above in motor activities (motor FIM™ ≥ 78); being able to transfer independently (FIM™ transfer ≥ 6) and to walk independently (FIM™ walking ≥ 6); being discharged in comparison with being just admitted to rehabilitation; being discharged home; and, finally, being discharged within 2 weeks of admission.
- *Responsiveness or sensitivity to change* (2). The relative responsiveness of UBS and originating scales was evaluated at the group-level by calculating an effect-size statistic (difference between the means at discharge and at admission divided by the standard deviation on admission) (8). In addition, responsiveness at the individual level was evaluated, using the individual significance of change method recently proposed by Hobart et al. (9). In particular, the significance of the individual changes (sig-change) was obtained by dividing the change score for each individual person (difference between discharge and admission score) by the square root of the sum of the squared standard error values at admission and discharge (9).
- *Interpretability* (4). Three goals were established for the analysis of interpretability of UBS:
  - *To equate on a common metric the UBS' raw scores and the corresponding BBS, POMA, TB, TW, FAB and TUG scores.* The discriminatory accuracy of UBS scores towards each score point of each originating scale and of the TUG was determined using receiving operator characteristic (ROC) curves. The Youden's Index (sensitivity + specificity – 1) was then used to identify the optimal UBS cut-off maximizing both sensitivity and specificity (10). For each selected score, several diagnostic performance indicators (sensitivity, specificity and diagnostic accuracy) were calculated and reported.
  - *To provide an estimate of the risk of falling associated with each UBS stratum.* In an earlier paper we had demonstrated that UBS was able to identify in this sample up to 9 strata (5), i.e. the statistically distinct levels of person ability that one scale can reliably distinguish (11). As the originating scales had been equated to the UBS ruler (5), it was then possible to link to various UBS scores the estimates of fall risk of the corresponding scores of the originating scales. Finally, a mean estimate of fall risk per stratum was calculated by simply calculating the mean of all the estimates of fall risk for the UBS score points falling in that stratum.
  - *To evaluate the level of balance ability at which various functional mobility goals were achieved in this sample.* The optimal UBS score points maximizing sensitivity and specificity were identified using Youden's Index for the following outcomes: need for supervision, modified independence or independence in activities known to be at risk for falling, such as, respectively, transfers and walking (12, 13) as expressed by the scores of 5, 6 and, respectively, 7 of the corresponding FIM™ items.
- *Predictive validity* (14). The predictive validity of UBS scores was evaluated on admission to rehabilitation in terms of relative risk of discharge to nursing home and relative risk of length of stay > 6 weeks. We also evaluated the relative risk of subsequent discharge to nursing home for patients at the end of the planned rehabilitation treatment.
- *Usability* (4). Administrative burden was evaluated by estimating a theoretical administration time for UBS considering the mean administration time of the originating scales, whereas respondent burden was checked by evaluating the responses to a satisfaction survey administered to the patients after the administration of the instrument. In order to improve usability, we established the minimum patient's ability criteria for the administration of UBS.

Statistical analysis

For all statistics, significance levels were set at 0.01. To facilitate the interpretation of the absolute values of correlation coefficients, a modified version of the cut-off criteria provided by Pallant (15) was adopted: *negligible*: 0–0.09; *weak*: 0.10–0.29; *moderate*: 0.30–0.49; *strong*: 0.50–0.79; *very strong*: ≥ 0.80. Effect sizes were interpreted according to the criteria provided by Cohen's (16): *small* (0.20–0.49); *medium* (0.50–0.79); *large* (≥ 0.80). The sig-change was interpreted using the criteria provided by Hobart et al. (9): *significant improvement* = Sig Change ≥ +1.96; *non-significant improvement* = 0 < Sig Change ≤ +1.95; *no change* = Sig Change = 0; *non-significant worsening* = –1.95 ≤ Sig Change < 0; *significant worsening* = Sig Change ≤ –1.96. Diagnostic performance indexes (sensitivity, specificity and diagnostic accuracy) were expressed as percentages and interpreted according to the criteria suggested by Cicchetti (17); *poor* (< 70%); *fair* (70–79%); *good* (80–89%); and *excellent* (90–100%).

All analyses were undertaken using SPSS (SPSS, Version 13 for Windows).

RESULTS

Patients recruited

Sample descriptive statistics have been reported elsewhere (5).

Exploration of the psychometric and clinical profile of Unified Balance Scale

The Kolgomorov-Smirnov (KS) test was not significant at the 0.01 level both for the UBS (n = 302; Z = 1.39; p = 0.410) and the BBFR (n = 115; Z = 0.70; p = 0.597), thus supporting the assumption of normality for these two measures. On the other hand, the latter could not be confirmed for the other interval-level indicators, as all KS tests resulted significant (TUG: $n=127$, $Z=2.35$, $p<0.000$; 10mtWT: $n=124$, $Z=2.03$, $p=0.001$; BBPS: $n=38$, $Z=2.18$, $p<0.000$). Therefore, for UBS and BBFR parametric statistics could be employed where possible, whereas in all other situations and for all ordinal indicators, only non-parametric statistics were used.

Concurrent criterion validity. As expected (Fig. 1), UBS correlated "very strongly" with BBS (Spearman's rho = 0.99, $p<0.000$, $n=288$), POMA (rho = 0.97, $p<0.000$, $n=293$), TB (rho = 0.97, $p<0.000$, $n=295$), and FAB (rho = 0.95, $p<0.000$, $n=249$).

External construct validity. Analysis of *convergent validity* (Fig. 1) showed that UBS correlated "very strongly" with TW (rho = 0.90, $p<0.000$, $n=298$) and FAC (rho = 0.80, $p<0.000$, $n=262$). As expected, correlations were "strong" with the indicators of walking independence, such as the WHS (rho = 0.75, $p<0.000$, $n=261$) and the 10mtWT (rho = 0.60, $p<0.000$, $n=124$). Correlations were also "strong" with TUG, an indicator of functional mobility and balance (rho=–0.65, $p<0.000$, $n=127$) and with motor FIM™, a scale of general motor functioning at the activity level (rho = 0.55, $p<0.000$, $n=214$).

Analysis of *divergent validity* (Fig. 1) showed that UBS correlated only "moderately" with TCT (rho = 0.48, $p<0.000$, $n=302$), with an indicator of limb motricity, such as the MI (rho = 0.47, $p<0.000$, $n=236$), and with an indicator of cognitive functioning at the activity level such as the cognitive FIM™ (rho = 0.37, $p<0.000$, $n=224$). Correlations were also "moderate" with indicators of balance at the impairment level, such as the BBPS test (rho = –0.43, $p<0.000$, $n=115$) and the BBFR test (rho = –0.39, $p<0.000$, $n=38$).

Analysis of *discriminant validity* confirmed the validity of all the hypothesis generated as shown in Table I. Effect sizes were all large (> 0.80) with the exception of the comparison between persons discharged within 2 weeks and those with a longer LOS, for which the effect size was just moderate (0.67).

Responsiveness. As shown in Table II, the effect size for all measures (UBS, BBS, TB, TW, POMA and FAB) were all > 1.0 and, therefore, were regarded as "large" according to Cohen's classification (16). UBS and BBS both had an effect size of 1.13. TB and TW had, respectively, the lowest (1.05) and the highest (1.27) effect sizes. However, analysis of the individual-level responsiveness showed that UBS was the best indicator at showing improvement (94.1%), followed by BBS (92.3%), whereas all other indicators showed a percentage of improvement ranging from 74.4% (TW) to 85.1% (POMA).

Interpretability. To equate on a common metric the UBS' raw scores and their corresponding BBS, POMA, TB, TW, FAB and TUG scores. For each point of the original scale, the equivalent UBS score was identified using ROC curves. Thus, for the BBS (score range: 0–56), the corresponding 57 UBS scores with the highest Youden's Index were identified, and the same was done also for FAB (score range: 0–40), POMA (score range: 0–28), TB (score range: 0–16) and TW (score range: 0–12). As a next step, these results were pooled together and ordered by increasing UBS score, as shown in Table III where, as an example, are represented only the UBS scores ranging from 58 to 62. As shown in Table III, scores from different scales were linked to a specific UBS score. For instance, BBS 54 (sensitivity: 93.9%; specificity: 100%), TB 15 (sensitivity 89.2%; specificity 96.3%), and TB 16 (sensitivity: 100%; specificity: 96.7%) were linked to a UBS score of 60.

Considering each original scale score separately (Table IV), the UBS yielded, on average, "excellent" diagnostic accuracy

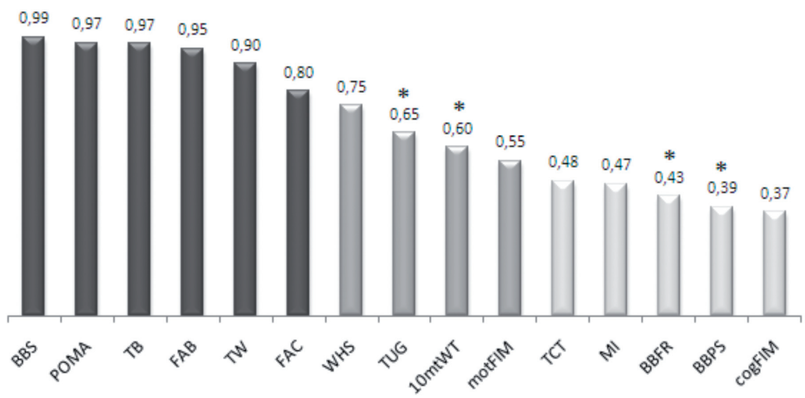

Fig. 1. Concurrent, convergent and divergent validity of the Unified Balance Scale (UBS). BBS: Berg Balance Scale; POMA: Performance-Oriented Mobility Assessment; TB: Tinetti balance; FAB: Fullerton Advanced Balance Scale; TW: Tinetti Walking; FAC: Functional Ambulation Classification; WHS: Walking Handicap Scale; TUG: Timed Up & Go test; 10mtWT: 10 m walking test; motFIM: motor subscore for FIM™; TCT: Trunk Control Test; MI: Motricity Index; BBFR: Balance Biodex™ fall risk test (degrees); BBPS: Balance Biodex™ postural stability test (degrees); cogFIM: cognitive subscore for FIM™. The values on top of the bars are the Spearmans' rho correlation coefficients. The dark grey bars represent very strong correlations (rho ≥ 0.80, whereas grey and light grey represent, respectively, strong correlations (rho: 0.50–0.79) and moderate correlations (rho: 0.30–0.49). In order to allow an easy comparison amongst different indicators, the negative correlations for TUG, 10mt WT, BBFR and BBPS (indicated by an asterisk) were presented using their absolute values.

Table I. *Discriminant validity for Unified Balance Scale (UBS)*

Groups	*n*	Mean UBS measure (logits)	F	DF	*p*	Effect size
TCT < 100	120	–2.99	78.98	301	< 0.000	0.98
TCT = 100	182	0.54				
mFIM < 78	162	–1.39	53.92	213	< 0.000	0.97
mFIM ≥ 78	52	2.60				
FIM transfers ≤ 5	118	–2.07	60.21	216	< 0.000	0.97
FIM transfers ≥ 6	99	1.49				
FIM walking ≤ 5	153	–1.43	43.71	215	< 0.000	0.95
FIM walking ≥ 6	63	2.00				
Admission	85	–2.22	61.64	169	< 0.000	0.98
Discharge	85	1.50				
Discharged to nursing home	32	–3.78	32.53	187	< 0.000	0.92
Discharged home	234	–0.55				
LOS > 2 weeks	159	–0.95	12.26	188	0.001	0.67
LOS ≤ 2 weeks	30	1.67				

F: F statistics; DF: degrees of freedom; TCT: Trunk Control Test; mFIM: motor FIM™; LOS: length of stay.

and specificity (≥ 90%) towards the individual originating scales. Sensitivity was "excellent" for FAB, BBS, and TB (ranging from 97.2 ± 0.6%) and "good" for TW and POMA (respectively 89.8 ± 2.3% and 87.8 ± 2.7%). Also the TUG scores could be linked to specific UBS scores (Table IV) with an average "good" sensitivity, specificity and diagnostic accuracy (respectively, 81.4 ± 3.8%, 82.2 ± 4.3% and 82.3 ± 2.3%).

Further to this, the scores of the originating scales and of the TUG were plotted against the UBS ruler (5), as shown in Fig. 2. In this way, those scores could be equated on a common metric of balance represented by the UBS ruler, thus allowing their conversion into UBS linear measures of balance and vice versa.

- *To provide an estimate of the risk of falling associated with each UBS stratum.* For both BBS (18, 19) and FAB (20) fall risk estimations for the whole scale were available, whereas for POMA (21), TW (21) and TB (21, 22) only estimation of fall risk for single scores could be identified. Further to this, all fall risk estimations were linked to the pertaining UBS cut-off scores as established in the previous section. As shown in Table III, all fall risk estimations falling within the same UBS stratum were then pooled together and, hence, it was possible to estimate a mean risk of falling for that stratum. The stratum fall risks so estimated are presented in Table V and Fig. 3. As shown in Fig. 3, the estimated fall risk was very high for the first 4 strata (75–78%), followed by an initial decrease to a 63% fall risk (although not statistically significant) in the 5th stratum. A very statistically significant further decrease to a mean 27% fall risk was observed on the 6th stratum and, following this, there was a further 13% decrease on the 7th stratum, followed by minor further decreases in the fall risk in the subsequent strata.
- *To evaluate the level of balance ability at which various functional mobility goals were achieved in this sample.* As shown in Fig. 2, the optimal UBS' scores were also determined for the "supervision" and "independence" scores for FIM™ transfers and FIM™ walking using Youden's Index. In particular, supervision and independence in transfers were linked to UBS scores of 20 (sensitivity: 70.0%, specificity: 81.8%, diagnostic accuracy: 77.4%) and, respectively, of 36 (sensitivity ranging from to 78.4% to 90.5%, specificity ranging from 55.4% to 73.4%, diagnostic accuracy ranging from 74.2% to 76.0%). Supervision, modified independence and complete independence in walking were linked, respectively, to an UBS scores of 42 (sensitivity: 79.9%, specificity: 64.4%, diagnostic accuracy: 75.5%), of 52 (sensitivity: 81.7%, specificity: 66.7%, diagnostic accuracy:

Table II. *Analysis of group-level and individual-level responsiveness of Unified Balance Scale (UBS) and of its originating scales*

	UBS	BBS	TB	TW	POMA	FAB
Admission-discharge completed records, *n*	85	78	81	82	80	56
Group level responsiveness						
Effect size	1.13	1.13	1.05	1.27	1.23	1.18
95% CI	0.87–1.53	0.87–1.56	0.84–1.50	0.98–1.64	0.96–1.63	0.70–1.34
Individual level responsiveness, %						
Sig-change: significantly worsened (a)	3.5	3.8	3.7	1.2	2.5	1.8
Sig-change: non-significantly worsened (b)	0.0	1.3	0.0	1.2	1.3	7.1
Sig-change: no change	2.4	2.6	14.8	23.2	11.3	14.3
Sig-change: non-significantly improved (c)	10.6	11.5	11.1	8.5	11.3	7.1
Sig-change: significantly improved (d)	83.5	80.8	70.4	65.9	73.8	69.9
Sig-change: total worsened (a+b)	3.5	5.1	3.7	2.4	3.8	8.9
Sig-change: total improved (c+d)	94.1	92.3	81.5	74.4	85.1	77.0

BBS: Berg Balance Scale; TB: Tinetti balance; TW: Tinetti Walking; POMA: Performance-Oriented Mobility Assessment; FAB: Fullerton Advanced Balance Scale; CI: confidence interval.

Table III. *Diagnostic performance indexes and associated fall risk for the Unified Balance Scale (UBS) cut-off scores falling into the 7th UBS' stratum*

Test		Originating scales' score			Diagnostic test performance		
UBS score	Stratum	Scale	Score	Associated fall risk, %	Sensitivity; %	Specificity; %	Diagnostic accuracy, %
58	7	POMA	25	–	88.9	97.5	90.1
58	7	TB	14	–	82.7	100	85.1
59	7	FAB	26	24.0[c]	94.1	100	94.8
59	7	FAB	27	22.0[c]	95.0	96.6	95.2
60	7	BBS	54	18[a], 4.2[b]	93.9	100	94.4
60	7	TB	15	–	89.2	96.3	89.8
60	7	TB	16	–	100	96.7	95.8
61	7	FAB	28	20.0[c]	91.8	100	92.4
61	7	FAB	29	18.5[c]	95.3	100	95.6
61	7	FAB	30	17.0[c]	96.6	100	96.8
61	7	FAB	31	15.0[c]	97.9	100	98.0
61	7	FAB	32	14.0[c]	98.7	100	98.8
61	7	FAB	33	13.0[c]	99.1	100	99.2
61	7	POMA	26	–	86.5	100	87.4
61	7	TW	11	–	82.9	100	83.9
61	7	TW	12	–	89.3	100	89.9
62	7	BBS	55	13.0[a]; 4.2[b]	93.5	100	93.8
62	7	BBS	56	11.0[a]; 4.20[b]	96.7	100	96.9
62	7	POMA	27	–	90.7	100	91.1
62	7	POMA	28	–	95.7	100	95.9
Stratum mean ± 95% CI				14.2 ± 4.6	92.6 ± 2.2	99.5 ± 0.2	93.3 ± 1.9

Only the diagnostic performance of the originating scales' scores corresponding to the UBS linked scores falling within the 7th stratum (UBS scores from 58 to 62) were displayed for space constraints. Data were ordered by increasing UBS scores. Within this stratum, the stratum fall risk (14.2 ± 4.6%) was calculated as a mean and 95% CI of all the fall risk estimates linked to the UBS scores falling in this stratum (displayed in the 5th column).
[a]Shumway-Cook et al. (18); [b]Bogle Thorbahn and Newton (19); [c]Hernandez and Rose (20).
POMA: Performance-Oriented Mobility Assessment; TB: Tinetti balance; FAB: Fullerton Advanced Balance Scale; BBS: Berg Balance Scale; TW: Tinetti Walking; CI: confidence interval.

75.5%) and of 56 (sensitivity: 86.5%, specificity: 44.4%, diagnostic accuracy: 77.8%). Although in this sample the diagnostic performance for the FIM™ items was, on average, in the "fair" to "good" range, it appeared that most patients reached: (*i*) supervision in transfers in the 4th stratum (fall risk: 69–87%); (*ii*) independence in transfers and supervision in walking in the 5th stratum (fall risk: 54–72%); (*iii*) independence in walking in the 6th stratum (fall risk 19–37%).

Predictive validity. As shown in Table VI, UBS scores ≤ 47 (upper limit of the 5th stratum) on admission carried a significant relative risk for both nursing home admission (NHA) and LOS > 6 weeks. For both outcomes, the relative risks were maxima for the first stratum and then progressively decreased across the 4 subsequent strata and became not significant in the last 4 strata.

On discharge from rehabilitation, an UBS total score ≤ 4 carried the highest relative risk for subsequent NHA (RR = 26,00, 95% CI from 2.73 to 247.25). For UBS total score ≥ 5, the relative risks for NHA progressively decreased and became non-significant for an UBS total score ≥ 40 (RR = 1.86; 95% CI from 0.98 to 3.52).

Usability. Considering the published administration times of the selected scales (respectively, BBS: 15–20 min; POMA: 10–15 min; FAB: 5–10 min) (23, 24), we calculated an estimated administration time for the total 40-item set ranging

Table IV. *Analysis of the mean diagnostic performance of Unified Balance Scale (UBS) towards the originating balance scales*

Gold standard	Score points (range)	Sensitivity		Specificity		Diagnostic accuracy	
		Mean	95% CI	Mean	95% CI	Mean	95% CI
BBS	57 (0–56)	93.4	4.4	98.3	0.4	97.5	0.4
FAB	41 (0–40)	97.2	0.6	98.6	0.5	97.4	0.5
TB	17 (0–16)	92.9	6.1	90.5	11.1	94.6	1.6
TW	13 (0–12)	89.8	2.3	95.2	2.8	91.7	2.0
POMA	29 (0–28)	87.8	2.7	90.0	3.2	91.5	1.3
TUG	–	81.4	3.8	82.2	4.3	82.3	2.3

The diagnostic performance of the optimal UBS scores identified for all scores of BBS, FAB, TB, TW and POMA is summarized as mean sensitivity, specificity and diagnostic accuracy for each originating balance scale. The diagnostic performance of TUG is also reported.
BBS: Berg Balance Scale; FAB: Fullerton Advanced Balance Scale; TB: Tinetti balance; TW: Tinetti Walking; POMA: Performance-Oriented Mobility Assessment; TUG: Timed Up & Go test; CI: confidence interval.

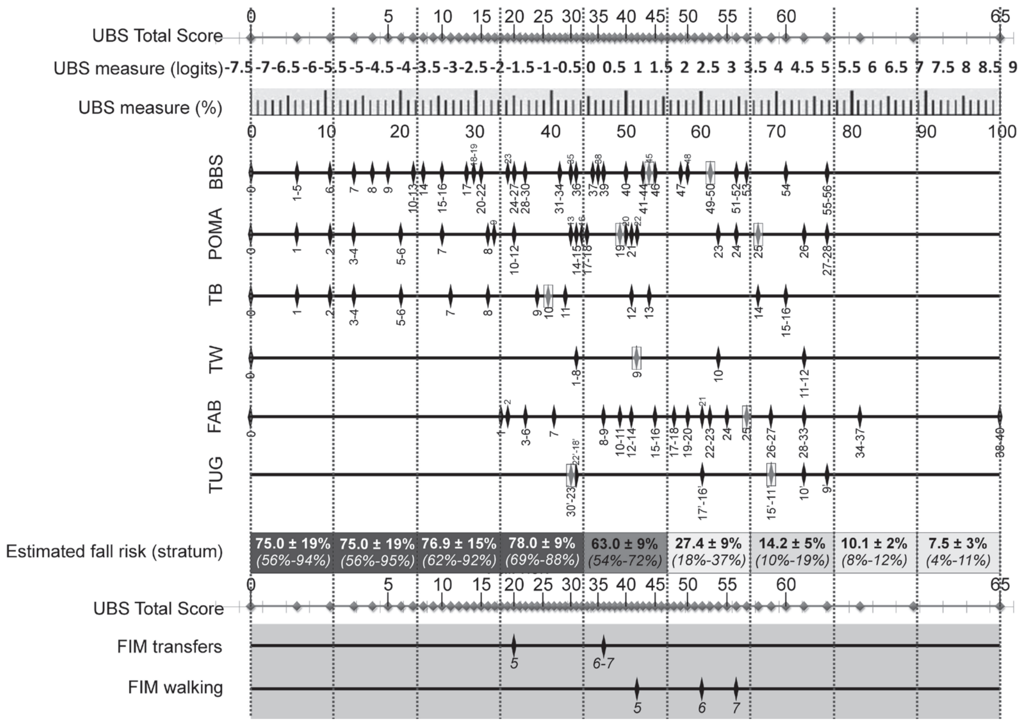

Fig. 2. Unified Balance Scale (UBS) measurement ruler with other balance indicators' cut-off scores and estimated risk of falling for each UBS strata. This figure contains 4 parts. The first part (above) is the UBS ruler. Below it, the rulers displaying the optimal cut-offs for the transformation of the originating scale's scores and Timed Up & Go test (TUG) into UBS score points and measures are shown. The small rectangular squares rounding some cut-offs (for instance, Berg Balance Scale (BBS 45)) indicated that these are published cut-offs relevant for fall risk estimation. The third part displays the estimated fall risk for each UBS stratum. Finally, the lower part (grey area) shows the diagnostic performance of the following external outcomes related to functional mobility: need for supervision, modified independence or independence in activities known to be at risk for falling, such as transfers and walking as expressed, respectively, by the scores of 5, 6, and 7 of the corresponding FIM™ items. POMA: Performance Oriented Mobility Assessment; TB: Tinetti balance; TW: Tinetti Walking; FAB: Fullerton Advanced Balance Scale.

from 30 to 45 mins, giving a theoretical administration time of 20–30 min for the 27 items of UBS.

The respondent burden of the whole assessment protocol was regarded as "acceptable" or "minimal" in terms of time requested, perceived difficulty and distress provoked by, respectively, 95.9%, 92.4%, and 93.4% of the 196 interviewed patients.

In order to establish the minimum patient's ability criteria for the administration of UBS and thus improve usability, all response patterns from patients with a total score of 1 on UBS ($n = 21$) were observed. From this, it appeared that such a score was determined in 71.4% of cases by a score of 1 in the BBS5 item ("transfers") and in the remaining 28.6% of cases by a score of 1 in the BBS4 item ("from standing to sitting"). As a consequence, the following minimal requirements for the administration of UBS in this sample were suggested: *either the ability to perform a transfer from bed to chair and vice versa with the assistance of maximum one person or without help, or the ability to sit without help from the standing position.*

Table V. *Estimated fall risk for each Unified Balance Scale (UBS) stratum*

UBS			Estimated fall risk		
Stratum	Score range	Measure range, %	Mean ± 95% CI	Lower 95% CI	Upper 95% CI
1	0–2	0–11.1	75.0 ± 18.5	56.5	93.5
2	3–7	11.1–22.2	75.0 ± 18.5	56.5	93.5
3	8–17	22.2–33.3	76.9 ± 14–7	62.1	91.6
4	18–32	33.3–44.4	78.0 ± 8.6	69.4	86.6
5	33–47	44.4–55.5	63.0 ± 8.8	54.1	71.8
6	48–57	55.5–66.6	27.4 ± 9.4	18.0	36.8
7	58–62	66.6–77.8	14.2 ± 4.6	9.5	18.8
8	63–64	77.8–88.9	10.1 ± 2.3	7.8	12.4
9	65	88.9–100	7.5 ± 3.3	4.2	10.8

CI: confidence interval.

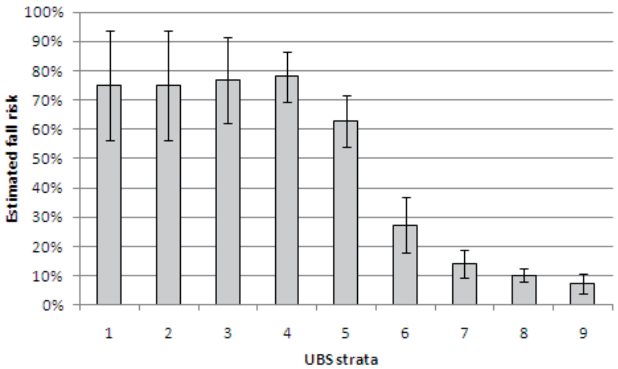

Fig. 3. Estimated fall risk for each Unified Balance Scale (UBS) stratum.

DISCUSSION

In this study, we examined the classical psychometric and clinical profile of UBS, demonstrating its external validity, predictive validity and responsiveness as a measure of balance in the calibrating sample. Furthermore, the score interpretability of UBS was ensured by equating the originating scales on a common metric of balance (the UBS ruler), that, in turn, allowed an indirect estimation of the mean fall risk for each UBS stratum. Also, specific UBS scores could be linked to specific rehabilitation outcomes in activities related to mobility that carry a high risk of falls. Analysis of usability suggested a mean administration time ranging from 20 to 30 min and good acceptability among patients.

Analysis of criterion and external construct validity confirmed the validity of UBS as a measure of balance at the activity level in the calibrating sample. Indeed, the very high correlations of UBS measures with BBS, TB, FAB and POMA were entirely expected, considering that UBS originates from items of these scales (5). We observed lower, but still strong, correlations with measures and indicators of walking and, to lesser extent, with motor FIM™. This is not surprising considering that balance is a prerequisite to perform effectively basic activities of daily living (25) and it is fundamental for effective stability during gait (25, 26). The high correlation of UBS with indicators of the latter could also be explained given that the UBS specifically incorporated various items assessing activities requiring balance during walking (such as tandem walk, walking with head turns, initiation of gait, etc.). We found only moderate correlations between UBS and laboratory tests of balance (Balance Biodex™ tests); in particular, we observed a slightly better correlation with the dynamic test (fall risk test) rather than with the static one (postural stability test). All these results were predicted, considering that UBS incorporated mostly dynamic activities, that these tools assess balance at the impairment level (by measuring postural sway) and, finally, that the correlations we found are similar to those reported by Berg et al. (27) for BBS. It should be noted, however, that as these tests require a fairly high level of functioning (28), they could be administered only to a minority of patients in the sample and, thus, this may have influenced the results.

In addition, analysis of discriminant validity confirmed all the hypotheses related to the construct in the examined sample. Quite interestingly, for these analyses we were entitled to make use of parametric statistics (i.e. analysis of variance) because UBS satisfied the requirements of interval-level measurement and the assumption of normality in this sample (29). The possibility of making full use of parametric statistics may provide further opportunities for application of UBS in the research setting, for example in path analysis to determine important moderating or mediating effects for the impact of, say, balance upon participation.

Good sensitivity to change or responsiveness to therapeutic interventions is a mandatory property if a scale is to be used as an outcome measure in routine clinical setting and clinical trials (9). The comparative analysis of group-level responsiveness (effect sizes) showed excellent responsiveness for UBS and of all the comparator measures in this sample, although FAB, POMA and TW showed better effect sizes and, hence, apparently better responsiveness. However, analysis of individual level responsiveness gave a different picture, indicating a better capacity of UBS at showing improvement, whereas FAB, POMA and TW were much less responsive in this sample. This is in line with the recent suggestions by Hobart et al. (9) that effect sizes are limited as indicator of responsiveness and may be misleading. Furthermore, it should be noted that effect sizes and the sig-change method are based on parametric statistics (e.g. means and standard deviations), methods that are not recommended for ordinal scales. Hence, it is possible that the relative responsiveness of the comparator indicators was inflated in view of their ordinal nature as opposed to the interval scaling properties of UBS, as suggested by a recent study (30).

Overall, from the analyses of interpretability, a key requirement for patient-based outcome measures (4, 31), several issues that may be relevant for the care of individual patients emerged.

Table VI. *Predictive validity of Unified Balance Scale (UBS) scores on admission to rehabilitation*

	UBS strata								
	1	2	3	4	5	6	7	8	9
UBS raw score threshold	≤2	≤7	≤17	≤32	≤47	≤57	≤62	≤64	≤65
Nursing home admission									
Relative risk	4.33	3.12	2.05	1.34	1.16	1.04	1.01	1.01	1.00
95% CI	2.43–7.73	2.26–4.31	1.63–2.58	1.18–1.53	1.06–1.27	n.s.	n.s.	n.s	n.s
LOS > 6 weeks									
Relative risk	2.69	2.00	1.67	1.27	1.19	1.04	1.01	1.01	1.0
95% CI	1.32–1.53	1.19–3.67	1.16–2.44	1.09–1.63	1.08–1.32	n.s.	n.s.	n.s.	n.s.

CI: confidence interval; LOS: length of stay in rehabilitation. n.s.: not significant as the CI includes the value of 1.

Firstly, it appeared that there was considerable variability in the originating scales regarding the cut-off estimations for the risk of a fall, as these cut-offs were dispersed across 4 strata, although most of these cut-offs were concentrated in the 5th and the 6th strata that, thus, appeared to be the most relevant ones for the assessment of fall risk. Secondly, the BBS 45 cut-off, generally regarded as a low-fall risk threshold (19, 27), appeared to lie rather in the 5th stratum that, according to our results, still carries a high risk of fall (54–72%). In fact, this finding is in line with literature data reporting fall risk estimations for this cut-off ranging from 50% (19) to 76% (18). Furthermore, Riddle & Stratford (32) showed a low sensitivity for this cut-off (leading to 33% missed fallers) and it has been suggested that higher BBS cut-offs (49, 50 or even 55) may in fact be a safer option (18, 32). Our results support these suggestions, as the latter BBS cut-offs lie in or beyond the 6th stratum which is characterized by a dramatic reduction in the fall risk in comparison with the previous strata. Thirdly, the interpretability data may assist clinicians in planning rehabilitation goals for individual patients. For instance, to be discharged home once they have reached the 6th ability stratum (27.4 ± 9% risk) may be considered a safe and acceptable rehabilitation goal for most young individuals without cognitive and bone impairments. On the other hand, the same goal may be too risky for an elderly patient with osteoporosis and/or cognitive deficits and/or taking poly-therapy or anticoagulants. All these risk factors for falls may, in fact, make the clinician aim, as a discharge goal, for a higher level of balance ability and/or to prescribe the most appropriate alternative (i.e. walking aids and/or constant supervision) to lower the fall risk further.

The interpretability results may also carry further advantages both for researchers and clinical policymakers. Firstly, it is possible to perform an "instant translation" of scores from one scale to another. For instance, a BBS score of 45 is equal, in terms of "quantity of balance", to a TB score of 13 and to an UBS score of 44 that, in turn, equates to a 58% measure of balance and a 63 ± 9% fall risk on the UBS ruler. This may facilitate the interpretation of the UBS scores in light of the more familiar originating scales' scores and cut-offs. Secondly, the score translation will allow the retrospective comparison of balance scores from different patients and/or from different centres using different balance instruments. Furthermore, comparisons will also be possible prospectively, even in centres using the originating scales alone (BBS, POMA or FAB). The possibility of such comparisons may be relevant in the context of multicentre clinical trials. Finally, the relationship between UBS measures and activities carrying a high risk of falls, such as transfers and walking, uncovered by the analysis of diagnostic performance of the corresponding FIM™ items, showed that, in our sample, both the wheelchair and the walking aids were abandoned when the fall risk may have been too high for some patients. Hence, similar kinds of knowledge may contribute to the issuing and/or improvement of local policies and practices in terms of fall prevention. For instance, specific rehabilitation goals and/or outcomes (e.g. walking with supervision, walking with aids and independent walking) may be operationalized by linking those to specific UBS scores in order to minimize the risk of falling.

Although the general validity of the interpretability results should be considered only within the context of our sample, it may be useful to consider two methodological issues. Firstly, the risk of introducing biases due to subjectivity in choosing the optimal UBS linking scores, just on the basis of the specific clinical circumstances in which the diagnostic test was applied (33), was largely reduced by the adoption of a universal criterion of diagnostic performance represented by the Youden's Index (10). The latter was preferred to other indicators (e.g. diagnostic odds ratio, etc.) in view of its simplicity and because it does not require sample-dependent information such as prevalence rates (10), which is one of the advantages that makes it the diagnostic index of choice in meta-analyses (34). Secondly, the methodology adopted to establish the strata estimates of fall risk can be regarded itself as similar to a meta-analysis (35), as we combined published fall risk estimates for different scales from different studies after co-calibrating those scales upon the same underlying metric. Hence, in our opinion, these two methodological aspects, coupled with the sample-independency characteristics of the UBS measures, specific to Rasch-calibrated scales, may enhance the generalizabilty of these results.

The analysis of predictive validity added a further small piece of validity evidence for UBS as a measure of balance in this sample. Our results suggested that the lower the balance on admission to rehabilitation, the higher the likelihood of remaining severely dependent and, hence, of being discharged to a nursing home, as well as the higher the risk of prolonged stay as inpatient. These results are in line with literature data suggesting that BBS is a significant predictor of discharge destination and LOS in stroke patients (36). The size of the relative risks for the latter was lower probably because of the influence of other independent variables, such as medical complications and/or social issues (36).

This study did not directly address reliability, as this has been reported at 0.975 in this sample within the framework of Rasch Analysis in a previous paper (5). Another aspect of reliability that was addressed previously is stability of the UBS over time, demonstrated by the absence of differential item functioning for each individual item between pre- and post-treatment assessments (5).

One possible weakness of the UBS was demonstrated in terms of usability, where an administration time longer than the recommended maximum of 15 min for a feasible instrument was estimated (3). This may prevent the routine usage of UBS in view of the long-standing tension between the need for comprehensive measurement of relevant outcomes and the demand for tools that can be administered quickly and easily in busy clinical settings (1). This is a relevant problem that has been only minimally addressed in this paper by providing clear minimum ability criteria for the administration of UBS, thus preventing its inappropriate administration. However, as the consequence of a fall has significant implications for the patient and their future care needs, we believe that an accurate assessment of this risk is vital, irrespective of the length of time taken to assess this risk. Nevertheless, Rasch analysis offers some excellent solutions to this issue (3), and further work is underway to provide clinicians with an efficient mechanism for maximizing the usability of UBS in everyday clinical settings.

Further limitations need to be considered carefully. Firstly, although UBS delivers linear estimates of balance that are sample-independent, all the traditional psychometric methods employed here are strictly sample-dependent (2). This may limit the generalizability of these results to other samples, thus prompting the need for replication in a much larger, multi-centre study involving different settings with different case mixes. Secondly, the estimation of fall risk was indirect and based on pooling the results from different studies. In doing this, we did not weight the results from different analyses and/or different tests, nor did we perform an extensive literature search, and thus we cannot exclude a publication bias. Therefore, these results will need confirmation by other prospective studies in which optimal cut-offs for fall risk will be estimated directly in other samples. Finally, the predictive validity analysis was not performed using logistic regression in view of the severe inequality of the outcome groups (few cases admitted to nursing home and with LOS > 6 weeks).

In conclusion, although further studies are warranted to generalize these results to different samples and to improve its usability, the UBS appears to be a very promising outcome measurement tool to evaluate the effectiveness both of fall reduction plans and of rehabilitation interventions aimed at improving balance.

ACKNOWLEDGEMENTS

The authors declare that they have no competing interests.

The authors are grateful to Stefano Gualdi, PT, Chiara Bosi, PT, and Matteo Maria Mariani, PT, for data collection.

REFERENCES

1. Jette AM, Haley SM. Contemporary measurement techniques for rehabilitation outcomes assessment. J Rehabil Med 2005; 37: 339–345.
2. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technol Assess 2009; 13: 1–177.
3. Andresen EM. Criteria for assessing the tools of disability outcomes research. Arch Phys Med Rehabil 2000; 81 Suppl 2: S15–S20.
4. Scientific Advisory Committee of the Medical Outcome Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res. 2002; 11: 193–205.
5. La Porta F, Franceschini M, Caselli S, Cavallini P, Susassi S, Tennant A. Unified Balance Scale: an activity-based, bed to community, and aetiology independent measure of balance calibrated with Rasch analysis. J Rehabil Med 2011; 43: 435–444.
6. Declaration of Helsinki: ethical principles for medical research involving human subjects [program]. Ferney-Voltaire: WMA; 2002.
7. Hobart JC, Lamping DL, Thompson AJ. Evaluating neurological outcome measures: the bare essentials. J Neurol Neurosurg Psychiatry 1996; 60: 127–130.
8. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989; 27 Suppl 3: S178–S189.
9. Hobart JC, Cano SJ, Thompson AJ. Effect sizes can be misleading: is it time to change the way we measure change? J Neurol Neurosurg Psychiatry 2010; 81: 1044–1048.
10. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. Biom J 2005; 47: 458–472.
11. Wright BD, Masters GN. Rating Scale Analysis. Chicago: MESA Press; 1982.
12. Weerdesteyn V, de Niet M, van Duijnhoven HJ, Geurts AC. Falls in individuals with stroke. J Rehabil Res Dev 2008; 45: 1195–1213.
13. Perell KL, Nelson A, Goldman RL, Luther SL, Prieto-Lewis N, Rubenstein LZ. Fall risk assessment measures: an analytic review. J Gerontol A Biol Sci Med Sci 2001; 56: M761–M766.
14. Hinderer SR, Hinderer KA. Principles and applications of measurement methods. In: DeLisa JA, Gans BM, Walsh NE, Bockenek WL, Frontera WR, Geiringer SR, et al., editors. Physical medicine and rehabilitation: principles and practice. 4th edn. Philadelphia: Lippincott, Williams & Wilkins; 2005, p. 1139–1162.
15. Pallant J. SPSS Survival manual – 2nd edition. Maidenhead, UK: Open University Press; 2005.
16. Cohen J. Statistical power analysis for the behavioral sciences. New York: Academic Press; 1969.
17. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assess 1994; 6: 284–290.
18. Shumway-Cook A, Baldwin M, Polissar NL, Gruber W. Predicting the probability for falls in community-dwelling older adults. Phys Ther 1997; 77: 812–819.
19. Bogle Thorbahn LD, Newton RA. Use of the Berg Balance Test to predict falls in elderly persons. Phys Ther 1996; 76: 576–583; discussion 584–585.
20. Hernandez D, Rose DJ. Predicting which older adults will or will not fall using the Fullerton Advanced Balance scale. Arch Phys Med Rehabil 2008; 89: 2309–2315.
21. Faber MJ, Bosscher RJ, van Wieringen PC. Clinimetric properties of the performance-oriented mobility assessment. Phys Ther 2006; 86: 944–954.
22. Murphy M, Olson S, Protas E, Overby A. Screening for falls in community-dwelling elderly. J Aging Phys Act 2003; 11: 66–80.
23. University of Missoury. Geriatric assessment toolkit. Avalible from: http://web.missouri.edu/~proste/tool/.
24. Rose DJ, Lucchese N, Wiersma LD. Development of a multidimensional balance scale for use with functionally independent older adults. Arch Phys Med Rehabil 2006; 87: 1478–1485.
25. Era P, Avlund K, Jokela J, Gause-Nilsson I, Heikkinen E, Steen B, et al. Postural balance and self-reported functional ability in 75-year-old men and women: a cross-national comparative study. J Am Geriatr Soc 1997; 45: 21–29.
26. Horak FB, Henry SM, Shumway-Cook A. Postural perturbations: new insights for treatment of balance disorders. Phys Ther 1997; 77: 517–533.
27. Berg KO, Wood-Dauphinee SL, Williams JI, Maki B. Measuring balance in the elderly: validation of an instrument. Can J Public Health 1992; 83 Suppl 2: S7–S11.
28. Berg K, Wood-Dauphinée S, Williams J, Gayton D. Measuring balance in the elderly: preliminary development of an instrument. Physiother Canada 1989; 41: 304–311.
29. Svensson E. Guidelines to statistical evaluation of data from rating scales and questionnaire. J Rehabil Med 2001; 33: 47–48.
30. Kersten P, White PJ, Tennant A. The visual analogue WOMAC 3.0 scale – internal validity and responsiveness of the VAS version. BMC Musculoskelet Disord 2010; 11: 80.
31. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 1998; 2: 1–74.
32. Riddle DL, Stratford PW. Interpreting validity indexes for diagnostic tests: an illustration using the Berg balance test. Phys Ther 1999; 79: 939–948.
33. Bewick V, Cheek L, Ball J. Statistics review 13: receiver operating characteristic curves. Crit Care 2004; 8: 508–512.
34. Bohning D, Bohning W, Holling H. Revisiting Youden's index as a useful measure of the misclassification error in meta-analysis of diagnostic studies. Stat Methods Med Res 2008; 17: 543–554.
35. Katz RT, Campagnolo DI, Goldberg G, Parker JC, Pine ZM, Whyte J. Crititical evaluation of Clinical Reserarch. Arch Phys Med Rehabil 1995; 76: 82–93.
36. Wee JY, Wong H, Palepu A. Validation of the Berg Balance Scale as a predictor of length of stay and discharge destination in stroke. Arch Phys Med Rehabil 2003; 84: 731–735.