

Developing a Short Form of the Berg Balance Scale for People With Stroke

Background and Purpose. To improve the utility of the Berg Balance Scale (BBS), the aim of this study was to develop a short form of the BBS (SFBBBS) that was psychometrically similar (including test reliability, validity, and responsiveness) to the original BBS for people with stroke. **Subjects and Methods.** A total of 226 subjects with stroke participated in this prospective study at 14 days after their stroke; 167 of these subjects also were examined at 90 days after their stroke. The BBS, Barthel Index, and Fugl-Meyer Motor Test were administered at these 2 time points. By reducing the number of tested items by more than half the number of items in the original BBS (ie, making 4-, 5-, 6-, and 7-item tests) and simplifying the scoring system of the original BBS (ie, collapsing the 5-level scale into a 3-level scale [BBS-3P]), we generated a total of 8 SFBBBSs. **Results.** The distributions of scores for all 8 SFBBBSs were acceptable but featured notable floor effects. The 4-item BBS, 5-item BBS, 5-item BBS-3P, and 7-item BBS-3P demonstrated good reliability. The subjects' scores on the 6-item BBS, 6-item BBS-3P, 7-item BBS, and 7-item BBS-3P showed excellent agreement with those on the original BBS. The 6-item BBS-3P and 7-item BBS-3P exhibited great responsiveness. Only the 7-item BBS-3P demonstrated both satisfactory and psychometric properties similar to those of the original BBS. **Discussion and Conclusion.** The 7-item BBS-3P was found to be psychometrically similar to the original BBS. The 7-item BBS-3P, compared with the original BBS, is simpler and faster to complete in either a clinical or a research setting and is recommended. [Chou CY, Chien CW, Hsueh IP, et al. Developing a short form of the Berg Balance Scale for people with stroke. *Phys Ther.* 2006;86:195–204.]

Key Words: *Balance, Cerebrovascular disorders, Item reduction.*

Chia-Yeh Chou, Chi-Wen Chien, I-Ping Hsueh, Ching-Fan Sheu, Chun-Hou Wang, Ching-Lin Hsieh

A balance measure that is deemed useful in a clinical setting must be both psychometrically sound and not lengthy to administer.¹⁻³ The Berg Balance Scale (BBS)⁴ has been used widely in order to evaluate balance performance for people with stroke.⁵⁻⁷ The BBS was previously shown to be psychometrically sound (including having high interrater reliability, high concurrent validity, and satisfactory responsiveness).^{5,6} However, 3 issues have been hampering the widespread utility of the BBS. First, the BBS may take about 20 minutes to complete⁸; such a procedure is quite time-consuming for daily clinical use and may place unreasonable demands upon respondents, especially in instances in which they may be seriously unwell, as in the case of people with stroke. Second, the BBS consists of 14 five-level items with scoring criteria varying from item to item. Such an inconsistency in scoring criteria could lead to difficulties for raters when making judgments about their patients' conditions, especially for raters with less training. Third, the extremely high internal consistency of the BBS (the Cronbach α coefficient has been found to be as high as .98)⁶ indicates, to some extent, item redundancy. These observations suggest that the BBS needs to be simplified in order to improve its utility.

The simplification of a measure may include reducing the number of items or shortening the levels of scaling, or both.^{2,9-12} It has been revealed that certain measures simplified by one or both of these methods are psychometrically similar to the original measures.^{2,5,9,11} There-

fore, the purpose of this study was to develop a short form of the BBS (SFBBs) that was psychometrically equivalent to the original BBS. We hypothesized that at least half of the items on the original BBS could be omitted and that the 5-level scaling could be reduced without sacrificing any psychometric properties. Thus, several SFBBs are proposed here, and the psychometric properties of the SFBBs were compared with those of the original BBS for a cohort of subjects who had had a stroke and who were evaluated from 14 days to 3 months after their stroke.

Method

Subjects

Data were retrieved from a prospective study (the Quality of Life After Stroke Study in Taiwan) initiated on December 1, 1999.⁶ For that study, subjects were recruited if they met the following criteria: diagnosis (clinical modification codes from the International Classification of Disease, 9th revision¹³) of cerebral hemorrhage (431), cerebral infarction (434), or other categories (430, 432, 433, 436, or 437); first onset of cerebrovascular accident without other major diseases; stroke onset within 14 days prior to hospital admission; ability to follow commands; and ability to provide informed consent personally or by proxy. Subjects were excluded if they had another stroke or other major disease(s) during the follow-up period or lived more than 64 km (40 miles) from the participating hospital.⁶

CY Chou, OT, MS, is Lecturer, Department of Rehabilitation, Jen-Teh Junior College of Medicine, Nursing and Management, Miaoli, Taiwan, and a doctoral student, Institute of Allied Health Sciences, College of Medicine, National Cheng Kung University, Tainan, Taiwan.

CW Chien, OT, BS, is Research Assistant, School of Occupational Therapy, College of Medicine, National Taiwan University, Taipei, Taiwan.

IP Hsueh, OT, MA, is Assistant Professor, School of Occupational Therapy, College of Medicine, National Taiwan University.

CF Sheu, PhD, is Professor, Department of Psychology, National Chung Cheng University, Chiayi, Taiwan.

CH Wang, PT, BS, is Associate Professor, School of Physical Therapy, College of Medical Technology, Chung Shan Medical University, and Department of Physical Therapy, Chung Shan Medical University Rehabilitation Hospital, Taichung, Taiwan. Address all correspondence to Mr Wang at School of Physical Therapy, College of Medical Technology, Chung Shan Medical University, 110, 1 Sec, Chien-Kuo N Rd, Taichung 402, Taiwan (chwang@csmu.edu.tw).

CL Hsieh, OT, PhD, is Professor and Chair, School of Occupational Therapy, College of Medicine, National Taiwan University.

Ms Chou, Mr Wang, and Dr Hsieh provided concept/idea/research design. Ms Chou, Mr Chien, and Ms Hsueh provided writing and data collection. Ms Hsueh and Dr Hsieh provided institutional liaisons, subjects, and project management. Mr Wang and Dr Hsieh provided fund procurement and clerical support. Dr Sheu and Dr Hsieh provided data analysis and consultation (including review of manuscript before submission).

This study was approved by an institutional review board of National Taiwan University Hospital.

This study was supported by a research grant from the National Science Council (NSC-90-2815-C-002-022-B), and National Health Research Institutes (NHRI-EX94-9204PP).

This article was received December 14, 2004, and was accepted July 18, 2005.

Measures

The BBS has 14 items, including 1 sitting item and 13 standing items.^{4,6} These items are based on a 5-level scale (0–4). Its total score ranges from 0 to 56. The BBS was originally developed to screen elderly people who are at risk for falling. The psychometric properties of the scale have been found to be satisfactory for people with stroke.^{5–7}

A simplified BBS with a 3-level scale (BBS-3P)⁵ was developed by collapsing the second, third, and fourth levels of the original scale into a single level. This collapsed level was scored when subjects met the criteria for the original second or higher level of the scale but not when subjects met the criteria for the highest level of the scale. The BBS-3P was found to feature psychometric properties similar to those of the original BBS. Thus, in the present study, both the BBS and the BBS-3P were used in the development of short forms with shortened scaling. For use of the BBS-3P in this study, the data retrieved for this study were recoded as 0-2-4 by collapsing the 3 middle levels of the original 5-level scale.

The Barthel Index (BI) was developed to measure the severity of disability.¹⁴ The BI evaluates 10 basic activities of daily living items: feeding, transferring, grooming, toileting, bathing, ambulation, stair climbing, dressing, bowel control, and bladder control.¹³ The total possible score of the BI ranges from 0 to 100. The BI was previously shown to yield scores with good interrater reliability (intraclass correlation coefficient [ICC]=.94) and high convergent validity (Spearman $\rho \geq .92$) for people with stroke.^{5,6,15,16} The BI was used to examine the convergent validity and predictive validity of data for the SFBBSSs proposed in this study.

The Fugl-Meyer Motor Test (FM)¹⁷ has been used to measure motor impairment following stroke. The FM consists of 50 items of upper- and lower-extremity motor function. Each item is graded on a 3-level scale. Its total possible score ranges from 0 to 100 points, and it has been shown to yield data with good interrater reliability (ICC $\geq .92$) and high concurrent validity ($r \geq .99$) for people with stroke.^{5,18,19} The FM was used to test the convergent validity of data for the SFBBSSs proposed in this study.

Procedure

Subjects consecutively enrolled in the Quality of Life After Stroke Study were examined at 14 days after the onset of stroke and reassessed at other specific time points (eg, 90 days) after stroke onset for up to 3 years after the stroke to characterize their recovery of neurologic function (eg, as measured by the FM), balance ability (eg, as measured by the BBS), functional abilities

Table 1.

Characteristics of Subjects With Stroke in Random Samples

Variable ^a	Calibration ^b	Validation ^c
Age, y, mean (SD)	68.2 (10.1)	68.1 (11.3)
Sex (no. of men/women)	73/40	64/49
BBS score, mean (SD)	23.4 (23)	23.3 (21.4)
BI score, mean (SD)	43.5 (31.5)	42.5 (30)
FM score, mean (SD)	53.2 (36)	54.4 (35.2)

^a BBS=Berg Balance Scale, BI=Barthel Index, FM=Fugl-Meyer Motor Test.

^b Group with which short forms of the Berg Balance Scale were developed (n=113).

^c Group with which short forms of the Berg Balance Scale were tested (n=113).

(eg, as measured by the BI), and health-related quality of life. The measures used in this study (ie, the BBS, the FM, and the BI) were administered by an occupational therapist who was not informed of the purpose of this study. The interrater reliabilities for the raters administering the BBS and the BI were satisfactory, with ICCs of .95 and .94, respectively.^{6,15,16}

Development of SFBBSSs

In this study, the method used to develop and validate the SFBBSSs mainly followed that proposed by Hobart and Thompson.² These authors selected items featuring the highest internal consistency (ie, minimizing measurement error) and the greatest responsiveness (ie, maximizing the ability to detect change). Thus, this method would appear to be especially useful for developing a measure for monitoring recovery after stroke and measuring outcome after treatment and was adopted in this study. The data retrieved for this study were randomly divided into 2 groups: a calibration group for developing the SFBBSSs and a validation group for comparing the psychometric properties of the various SFBBSSs with those of the original BBS.

To develop the SFBBSSs, the best items were determined by selecting the items with the lowest values from an overall item index of each item.² The overall item index of each item is the product of the 2 rank orders (ie, the rank order of the corrected item total correlation for an item and the rank order of the effect size for an item). The corrected item total correlation for an item is the correlation between the scores of an individual item and the sum of the scores of all of the items on the scale minus that item. The rank of the corrected item total correlation is useful in removing test items that have a lower correlation with the overall construct measured in the BBS. Furthermore, the effect size for an item is the mean change score (14–90 days after stroke) divided by the standard deviation of the scores at 14 days after stroke. The rank of the effect size is useful in removing test items that show little sensitivity to change. Finally, the corrected item total correlation for each item and

Table 2.
Item Analysis of Berg Balance Scale Scores at 14 Days After Stroke Onset

Item ^a	Item Total Correlation ^b (Rank Order) (n=113)	Effect Size ^c (Rank Order) (n=86)	Overall Item Index ^d (Rank Order)
Reaching forward with outstretched arm	.96 (1)	.72 (4)	4 (1)
Standing with eyes closed	.96 (1)	.68 (5)	5 (2)
Standing with one foot in front	.96 (1)	.66 (7)	7 (3)
Turning to look behind	.96 (1)	.65 (9)	9 (4)
Retrieving object from floor	.96 (1)	.61 (10)	10 (5)
Standing on one foot	.73 (12)	.86 (1)	12 (6)
Sitting to standing	.96 (1)	.46 (14)	14 (7)
Turning 360°	.83 (11)	.76 (2)	22 (8)
Standing unsupported	.95 (7)	.68 (5)	35 (9)
Placing alternate foot on stool	.72 (13)	.74 (3)	39 (10)
Transferring	.89 (10)	.66 (7)	70 (11)
Standing with feet together	.95 (7)	.61 (10)	70 (12)
Standing to sitting	.95 (7)	.53 (13)	91 (13)
Sitting unsupported	.72 (13)	.55 (12)	156 (14)

^aThe first 7 items were selected for developing the short forms of the Berg Balance Scale.

^bCalculated as the correlation between the score of each item and the total score of the remaining 13 items.

^cCalculated as the mean change score (the score at 14 days after onset minus the score at 90 days after onset) divided by the SD of the score at 14 days after onset. Rank order: 1=highest value, 14=lowest value.

^dProduct of rank order for item total correlation and rank order for effect size; for example, reaching forward with outstretched arm=1×4=4.

the effect size for each item were respectively ranked, and then the product of these rank orders was computed, that is, the overall item index of each item. For example, if the item total correlation rank of a given item is 1 and its effect size rank is 4, then its overall item index is 1×4=4. Lower values for the overall item index indicated better items.

We hypothesized that the use of 4 to 7 best items would be adequate for the SFBBSs. Four sets of SFBBSs were generated (ie, 4-item BBS, 5-item BBS, 6-item BBS, and 7-item BBS). We also used a technique to collapse the 3 levels in the middle of the BBS into a single level. Thus, we developed an additional 4 sets of SFBBSs (ie, 4-item BBS-3P, 5-item BBS-3P, 6-item BBS-3P, and 7-item BBS-3P). Therefore, a total of 8 SFBBSs were generated.

Data Analysis

To compare the psychometric properties of the 8 SFBBSs and the original BBS, we linearly transformed the scores of the SFBBSs into the same score range as that for the original BBS (0–56). The psychometric properties tested in this study included acceptability, reliability, validity, and responsiveness.

Acceptability is a determination of whether the score distributions of a measure can match the distribution corresponding to the subjects intended to be measured.²

A measure exhibiting good acceptability should reveal observable scores spanning the entire range of the scale, with a mean score near the scale midpoint, and featuring small floor and ceiling effects, that is, less than 15% of the subjects achieving the lowest or the highest scores.^{2,20}

Test reliability reflects the degree of precision of a measure; that is, high reliability requires a low rate of errors to be generated.^{21,22} To estimate test reliability, Hobart and Thompson² recommended examination of the internal consistency of a specific test by use of Cronbach α coefficients to determine the intercorrelations among the items.² It has been suggested that reliability estimations exceed .80 for group comparison studies and .95 for individual patient clinical decision making.^{2,21} Confidence intervals for the α coefficients were computed.^{2,23} Confidence intervals for individual scores for subjects with stroke were computed by calculating the standard error of measurement (SEM).²¹ The SEM indicates

the spread of scores.²⁴ The following 2 formulas were used: $SEM = (\text{standard deviation of sample scores}) \times \sqrt{1 - \text{reliability}}$ and 95% confidence intervals for individual scores = $\pm 1.96 \times SEM$.

Test validity indicates whether a measure actually determines what it has been constructed to determine.^{2,25} We examined the agreement between the results of the SFBBSs and the results of the original BBS at 14 days after stroke by using a random-effects model ICC and the method proposed by Bland and Altman,²⁶ which involves plotting the scores of the difference between the original BBS and the SFBBSs against those of the average between the original BBS and the SFBBSs.²⁶ Ideally, there should be no trend showing systematic bias in a Bland-Altman plot.²⁶ These results are useful for determining whether the SFBBSs and the original BBS can be used interchangeably.

In addition, 3 validity indicators were examined for the comparisons of the 8 SFBBSs and the original BBS. First, the concurrent validity at 14 days after stroke was examined by computing the intercorrelations between the scores of the SFBBSs and those of the original BBS. Second, the convergent validity for the scores of the SFBBSs, the FM, and the BI at 14 days after stroke also was examined. Third, the predictive validity of scores for the SFBBSs was determined by examining the relation-

Table 3. Comparison of Acceptability, Reliability, and Agreement of 8 Short Forms of the Berg Balance Scale at 14 Days After Stroke Onset (n = 113)

Psychometric Property	14-Item BBS ^a	7-Item ^b BBS [BBS-3P] ^c	6-Item ^d BBS [BBS-3P]	5-Item ^e BBS [BBS-3P]	4-Item ^f BBS [BBS-3P]
Acceptability					
Mean score (SD)	23.3 (21.4)	23.2 (23)	22.7 (22.9)	25.4 (25.2)	25.2 (24.9)
% floor/ceiling effect ^g	23.9/2.7	41.6/2.7 [41.6/2.7]	46.9/2.7 [46.9/2.7]	46.9/3.5 [46.9/3.5]	46.9/3.5 [46.9/3.5]
Reliability					
α (LL 95% CI) ^h	.98 (.89)	.97 (.79) [.96 (.8)]	.96 (.74) [.95 (.75)]	.98 (.86) [.97 (.84)]	.98 (.81) [.96 (.78)]
SEM ⁱ	3.3	4.1 [4.2]	4.6 [4.7]	3.6 [4.2]	3.7 [4.6]
95% CI ^j	±6.4	±8.1 [±8.3]	±8.9 [±9.2]	±7 [±8.3]	±7.3 [±9.1]
Agreement with 14-item BBS-5P					
ICC ^k		.99 [.99]	.98 [.98]	.96 [.97]	.96 [.97]
Mean difference (SD) ^l		0.2 (3.4) [1.2 (3)]	0.6 (3.9) [1.9 (3.6)]	-2.1 (6.2) [-0.5 (5.3)]	-1.9 (6.5) [-0.1 (5.5)]
Limits of agreement ^m		-6.5 to 6.8 [-4.8 to 7.1]	-7.1 to 8.2 [-5.2 to 9]	-14.3 to 10.1 [-10.9 to 9.9]	-14.7 to 10.8 [-10.8 to 10.7]

^a Original 5-level Berg Balance Scale (BBS).

^b The 7 items are reaching forward with outstretched arm, standing with eyes closed, standing with one foot in front, turning to look behind, retrieving object from floor, standing on one foot, and sitting to standing.

^c The data in brackets are the results of the BBS simplified in scale from 0-1-2-3-4 (BBS-5P) to 0-2-4 (BBS-3P).

^d The 6 items are reaching forward with outstretched arm, standing with eyes closed, standing with one foot in front, turning to look behind, retrieving object from floor, and standing on one foot.

^e The 5 items are reaching forward with outstretched arm, standing with eyes closed, standing with one foot in front, turning to look behind, and retrieving object from floor.

^f The 4 items are reaching forward with outstretched arm, standing with eyes closed, standing with one foot in front, and turning to look behind.

^g Percentage of sample scoring 0 (floor effect)/percentage of sample scoring 56 (ceiling effect).

^h Lower limit (LL) of the 95% confidence interval (95% CI) was calculated as $(\alpha - 1.96SE)$, where $SE = \sqrt{(SD \cdot ni) / [(k/2)(k-1)] - 1}$, SD = SD of item intercorrelations, and k = number of items in scale.

ⁱ Standard error of measurement (SEM) was calculated as $SD / (1 - \alpha)$.

^j Calculated as $1.96 \times SEM$.

^k ICC = intraclass correlation coefficient (random-effects model).

^l Calculated as the score on the 14-item BBS at 14 days after stroke onset minus the transformed scores on the short forms of the BBS at 14 days after stroke onset.

^m Mean difference \pm 1.96 SDs.

ships between the scores of the SFBBSSs at 14 days after stroke and those of the BI at 90 days after stroke.

Responsiveness reflects the effectiveness of a measure in detecting changes in the longitudinal follow-up of the participants.^{27,28} The extent of the responsiveness of the SFBBSSs was investigated by calculating effect sizes.^{22,25,29} Effect sizes were determined by computing the mean of the total score difference between 14 days and 90 days after stroke for each subject, divided by the standard deviation of the total score at 14 days after stroke.¹⁶ Larger values suggest greater responsiveness. Finally, we cross-validated the main psychometric properties of the best SFBBSS found by using 20 samples that were randomly and repeatedly drawn from the full sample.

Results

We examined 226 subjects at 14 days after stroke; 167 of these subjects were successfully examined at 90 days after stroke. The 226 subjects examined at day 14 were randomly divided into either a calibration group or a validation group, with each group consisting of 113 subjects. There was no significant difference between the ratios of male and female subjects for the calibration and validation groups, and the various scores for the BBS, BI, and FM proved to be very close for the 2 groups (Tab. 1).

Development of SFBBSSs

Table 2 shows that the corrected item total correlations ranged from .72 to .96 and that the effect sizes ranged from .46 to .86 for individual items. According to the overall item index listed in Table 2, the 7-item BBS and 7-item BBS-3P were developed by including the 7 best items (in a hierarchical order): reaching forward with outstretched arm, standing with eyes closed, standing with one foot in front, turning to look behind, retrieving object from floor, standing on one foot, and sitting to standing. The 6-item, 5-item, and 4-item BBS and the BBS-3P were developed by sequentially removing the worst items from those 7 best items. Thus, a total of 8 SFBBSSs were developed. Comprehensive evaluation of the psycho-

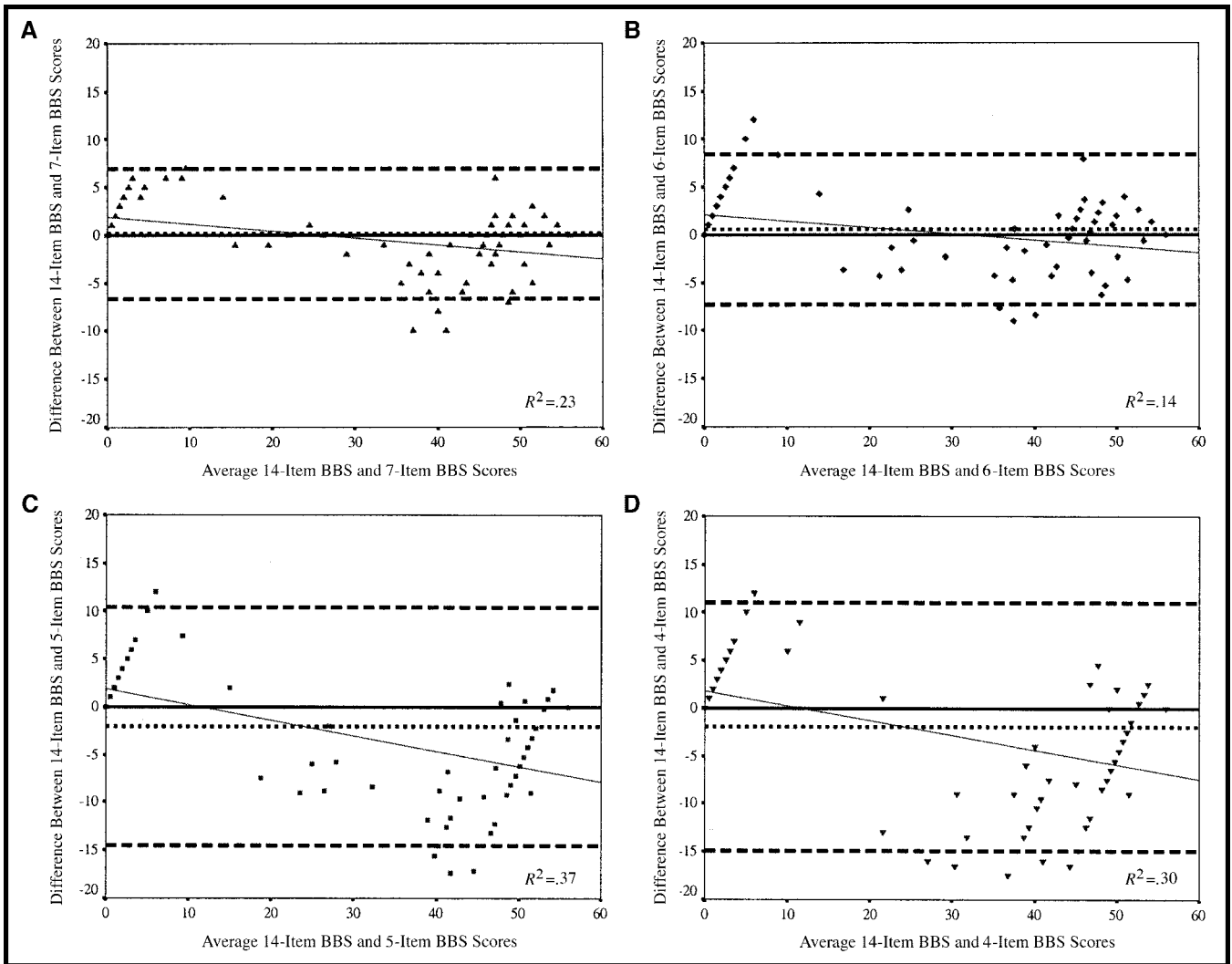


Figure 1. Bland-Altman method for plotting the difference of scores against the mean scores of the original Berg Balance Scale (BBS) and 4 short forms of the BBS, the 7-, 6-, 5-, and 4-item forms of the BBS (A, B, C, and D, respectively). The 2 bold dashed lines define the limits of agreement (mean of difference \pm 2 SDs).

metric properties of the 8 SFBBSs and the BBS revealed the following results.

Acceptability

All 8 SFBBSs investigated exhibited good variability, as the test scores spanned the full possible ranges of the scales. Mean scores (22.1–25.4) were slightly off the midpoint (28), and floor effects were notable ($\geq 41.6\%$ of the subjects) for the 8 SFBBSs (Tab. 3).

Reliability

All 8 SFBBSs had very high α coefficients ($\geq .95$), but only the 4-item BBS, 5-item BBS, 5-item BBS-3P, and 7-item BBS-3P had lower-limit confidence intervals that met the criterion of .80 (Tab. 3). The SEM of the 8 SFBBSs ranged from 3.6 to 4.7, values that were lower than 5.6 (ie, 10% the highest possible score of 56, such a score indicating clinical importance).³⁰

Validity

The ICCs for the original BBS and SFBBSs were high ($\geq .96$) (Tab. 3), indicating excellent agreement between the SFBBSs and the original BBS. The limits of agreement of the 6-item BBS, 6-item BBS-3P, 7-item BBS, and 7-item BBS-3P were about half those of the other SFBBSs, indicating that their scores for individual subjects were closer to the scores of the original BBS than to those of the other SFBBSs. Figures 1 and 2 show that only the 6-item BBS-3P and 7-item BBS-3P demonstrated no obvious systematic bias toward the BBS in the Bland-Altman plots ($r^2 \leq .04$).

Table 4 shows that scores for all 8 SFBBSs demonstrated very high concurrent validity with scores for the original BBS ($r \geq .97$). Moreover, scores for all of the SFBBSs exhibited equivalent and high convergent validity with scores for the BI ($r = .84$ – $.86$) and with scores for the FM

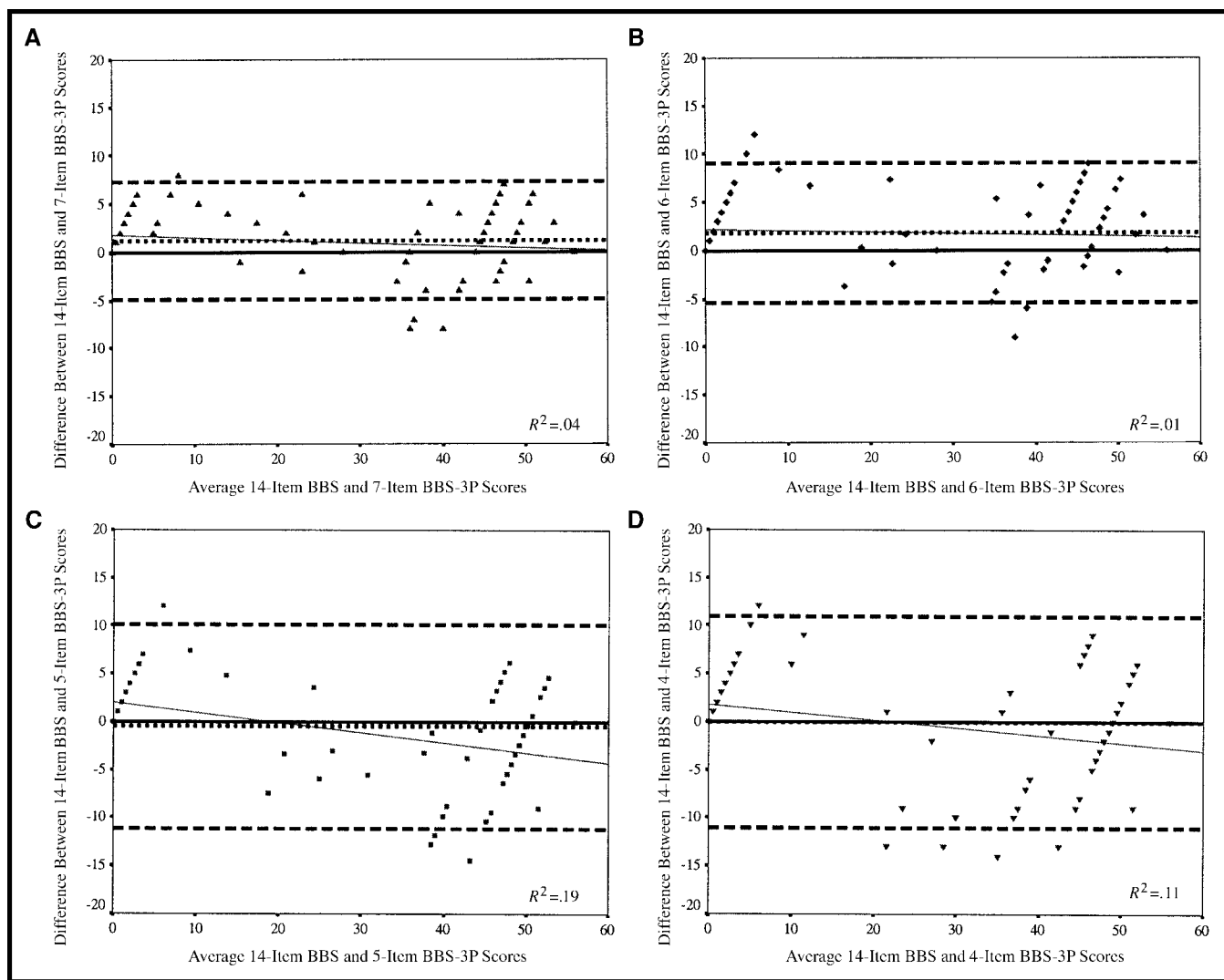


Figure 2.

Bland-Altman method for plotting the difference of scores against the mean scores of the original Berg Balance Scale (BBS) and 4 short forms of the BBS with 3 scales (BBS-3P), the 7-, 6-, 5-, and 4-item forms of the BBS-3P (A, B, C, and D, respectively). The 2 bold dashed lines define the limits of agreement (mean of difference \pm 2 SDs).

($r=.66-.68$). The extent to which each of the 8 SFBBSS was able to predict the score of the BI at 90 days after stroke also was similar to that of the original BBS and satisfactory ($r=.58-.60$).

Responsiveness

Table 4 shows that the 8 SFBBSS and the original BBS had similar and satisfactory effect sizes (.69-.85), especially the 6-item BBS-3P and 7-item BBS-3P, both of which had large effect sizes ($\geq .8$). We found that the 7-item BBS-3P was slightly superior to the 6-item BBS-3P in acceptability, reliability, and validity (Tabs. 3 and 4). Only the 7-item BBS-3P met all of the predefined psychometric criteria, with the exception of the floor effects. Furthermore, the findings of this study also supported the requirement that the 7-item BBS-3P demonstrate satisfactory internal consistency, concurrent

validity, and responsiveness relative to the original BBS for the 20 randomly reselected samples (Tab. 5).

Discussion

To the best of our knowledge, this study is the first to develop an SFBBSS with psychometric properties that are very similar to those of the original BBS. Simplifying the original BBS was achieved by comparing the psychometric properties of the original BBS with those of the 8 SFBBSS that were developed in this study. As a result, the 7-item BBS-3P did not appear to lose any psychometric properties compared with the original BBS; as a consequence, it is recommended for monitoring the recovery and measuring the outcome of patients with stroke.

Compared with the original BBS, the 7-item BBS-3P is improved in 3 significant aspects. First, the number of

Table 4. Comparison of Validity and Responsiveness of 8 Short Forms of the Berg Balance Scale

Psychometric Property	14-Item BBS ^a	7-Item ^b BBS [BBS-3P] ^c	6-Item ^d BBS [BBS-3P]	5-Item ^e BBS [BBS-3P]	4-Item ^f BBS [BBS-3P]
Validity ^g					
Concurrent (n=113)		.99 (98) [.99 (98)]	.99 (97) [.99 (97)]	.98 (95) [.98 (96)]	.97 (94) [.97 (95)]
14-item BBS (%V) ^h		.86 [.86]	.85 [.85]	.84 [.85]	.84 [.84]
Convergent (n=113)		.68 [.68]	.67 [.67]	.67 [.67]	.66 [.67]
BI ⁱ	.88				
FM ^j	.71				
Predictive (n=81)	.62	.60 [.60]	.59 [.58]	.59 [.59]	.59 [.59]
BI at 90 days after onset					
Responsiveness (n=81)		17.5 (20) [17.2 (18.9)]	17.5 (20.4) [17.3 (19.2)]	17.5 (22.1) [17.1 (20.6)]	17.1 (21.6) [16.6 (19.9)]
Change score, ^k mean (SD)		.78 [.80]	.78 [.81]	.70 [.73]	.69 [.72]
Effect size ^l	.85				

^a Original 5-level Berg Balance Scale (BBS).

^b The 7 items are reaching forward with outstretched arm, standing with eyes closed, standing with one foot in front, turning to look behind, retrieving object from floor, standing on one foot, and sitting to standing.

^c The 6 items are reaching forward with outstretched arm, standing with eyes closed, standing with one foot in front, turning to look behind, retrieving object from floor, and standing on one foot.

^d The data in brackets are the results of the BBS simplified in scale from 0-1,2,3-4 to 0-2-4 (BBS-3P).

^e The 5 items are reaching forward with outstretched arm, standing with eyes closed, standing with one foot in front, turning to look behind, and retrieving object from floor.

^f The 4 items are reaching forward with outstretched arm, standing with eyes closed, standing with one foot in front, and turning to look behind.

^g Production-moment correlations for the scores at 14 days after stroke onset.

^h Percent variance (%V) of 14-item BBS score explained.

ⁱ BI = Barthel Index.

^j FM = Fugl-Meyer Motor Test.

^k Calculated as the score at 14 days minus the score at 90 days. All change scores were statistically significant ($P < .001$).

^l Calculated as the mean change score divided by the standard deviation of the score at 14 days after stroke.

items is reduced by half. Second, the scoring levels are reduced from 5 to 3, thereby reducing the possibility of scoring inconsistency. Third, administration of the 7-item BBS-3P requires fewer assessment tools. For example, a stool was not necessary for the 7-item BBS-3P because of the removal of the item “placing alternate foot on stool.” All of these improvements allowed the raters to complete the SFBBS within half the time required to complete the original BBS (less than 10 of the original 20 minutes). This advantage of the 7-item BBS-3P decreases the possibility of incomplete data collection and contributes to efficiency in examination.

The use of the 7-item BBS-3P in clinical and research settings can be an improvement over the use of the original BBS given that the 7-item BBS-3P has excellent agreement with the original BBS. The Bland-Altman plot revealed that there was no notable trend between the difference and the average scores of the 7-item BBS-3P and the original BBS. Thus, the 7-item BBS-3P may be used interchangeably with the original BBS. The 7-item BBS-3P is especially useful when the time available for examination is short, such as at follow-up or when the clients are too weak to endure long examinations.

From the perspective of psychometric properties, up to 7 items (eg, standing unsupported and transferring) in the original BBS were found in our study to be redundant because their application did not provide any additional psychometric information. In earlier research, similar findings of item redundancy also were obtained for measures of some other domains, such as activities of daily living or quality of life.^{1,2,9,10,12,16} Therefore, it is worthwhile to explore in future studies whether there is any possibility of simplifying any of the other domains of conventional measures to decrease item redundancy in the measures and to promote the utility of clinical measures. However, from the clinical point of view, some important aspects of the balance performance of individual patients (eg, standing unsupported and transferring) are not recorded after the deletion of the items. Therefore, the 7-item BBS-3P may not be able to entirely replace the original BBS in the clinical setting, especially when the specific balance functions measured by the items deleted

Table 5.

Main Psychometric Properties of the Berg Balance Scale (BBS) and the 7-Item BBS With 3 Scales for 20 Randomly Selected Samples

Psychometric Property	14-Item BBS		7-Item BBS-3P ^a	
	\bar{X}	Range	\bar{X}	Range
Distribution, % floor effect ^b	25.0	19.5–31.4	42.7	37.2–48.2
Internal consistency, Cronbach α coefficient	.979	.977–.983	.968	.965–.974
Concurrent validity, ICC ^c			.988	.987–.991
Responsiveness, effect size ^d	.78	.62–.87	.75	.57–.85

^a Seven-item BBS simplified in scaling from 0-1-2-3-4 to 0-2-4 (BBS-3P). The 7 items are reaching forward with outstretched arm, standing with eyes closed, standing with one foot in front, turning to look behind, retrieving object from floor, standing on one foot, and sitting to standing.

^b Percentage of sample scoring 0 (floor effect).

^c ICC=intraclass correlation coefficient (random-effects model).

^d Calculated as the mean change score divided by the standard deviation of the score at 14 days after stroke.

from the original measure are deemed to be treatment goals.

In this study, we used the method described by Hobart and Thompson² to develop the 7-item BBS-3P. In that study, 6 of the first 7 items selected from the original BBS had been ranked as 1 (best) according to their corrected item total correlations, indicating that the corrected item total correlations were somewhat limited in discriminating the psychometric properties of the items of the BBS. Fortunately, this limitation did not interfere with the development of the 7-item BBS-3P. Future studies may add interrater reliability or test-retest reliability² as supplementary criteria when too many items are ranked the same in the results for the corrected item total correlations.

A rather notable floor effect that was revealed for the 7-item BBS-3P also was found for the original BBS to a lesser extent. This notable floor effect may have resulted from the removal of the easiest item (unsupported sitting) from the 14 items of the BBS. Removing this item from the original BBS could reduce the ability of the 7-item BBS-3P to detect changes in sitting balance. As a result, the floor effect could weaken the ability of the 7-item BBS-3P to differentiate small balance function differences between people with severe stroke. Moreover, the presence of just such a floor effect may potentially damage the relative responsiveness of such a measure. However, we found the responsiveness of the 7-item BBS-3P to be satisfactory and very similar to that of the original BBS. Thus, the floor effect of the BBS-3P may not necessarily restrict the use of the 7-item BBS-3P for detecting balance improvement. From another point of view, the 7-item BBS-3P would benefit people who are able to attain or maintain upright stance without support, because testing easy tasks (eg, unsupported sitting) appears to be irrelevant for these people.

The psychometric properties of the 7-item BBS-3P were internally validated by use of 20 randomly reselected samples. The results of such validation testing provided

strong evidence suggesting that the 7-item BBS-3P was psychometrically similar (including internal consistency, concurrent validity, and responsiveness) to the original BBS for people with stroke. Such results suggested that we did not “over fit” the results of the 7-item BBS-3P to this single data set and that the findings of this study were well supported.

Conclusion

The 7-item BBS-3P measure has sound psychometric properties and practical utility for use with people who have had a stroke. The 7-item BBS-3P, therefore, is suggested for use in people with stroke in both clinical and research settings.

References

- 1 Talley NJ, Verlinden M, Jones M. Quality of life in functional dyspepsia: responsiveness of the Nepean Dyspepsia Index and development of a new 10-item short form. *Aliment Pharmacol Ther*. 2001;15:207–216.
- 2 Hobart JC, Thompson AJ. The five item Barthel index. *J Neurol Neurosurg Psychiatry*. 2001;71:225–230.
- 3 Wang CH, Hsieh CL, Dai MH, et al. Inter-rater reliability and validity of the stroke rehabilitation assessment of movement (STREAM) instrument. *J Rehabil Med*. 2002;34:20–24.
- 4 Berg K, Wood-Dauphinee S, Williams JI, Gayton D. Measuring balance in the elderly: preliminary development of an instrument. *Physiother Can*. 1989;41:304–311.
- 5 Wang CH, Hsueh IP, Sheu CF, et al. Psychometric properties of 2 simplified 3-level balance scales used for patients with stroke. *Phys Ther*. 2004;84:430–438.
- 6 Mao HF, Hsueh IP, Tang PF, et al. Analysis and comparison of the psychometric properties of three balance measures for stroke patients. *Stroke*. 2002;33:1022–1027.
- 7 Berg K, Wood-Dauphinee S, Williams JI. The balance scale: reliability assessment with elderly residents and patients with an acute stroke. *Scand J Rehabil Med*. 1995;27:27–36.
- 8 Stevenson TJ. Detecting change in patients with stroke using the Berg Balance Scale. *Aust J Physiother*. 2001;47:29–38.
- 9 Jones G, Jenkinson C, Kennedy S. Development of the short form endometriosis health profile questionnaire: the EHP-5. *Qual Life Res*. 2004;13:695–704.

- 10** Wilde Larsson B, Larsson G. Development of a short form of the quality from the patient's perspective (QPP) questionnaire. *J Clin Nurs*. 2002;11:681–687.
- 11** Cash TF. The situational inventory of body-image dysphoria: psychometric evidence and development of a short form. *Int J Eating Disord*. 2002;32:362–366.
- 12** Chiou CF, Sherbourne CD, Ofman J, et al. Development and validation of Cedars-Sinai Health-Related Quality of Life in Rheumatoid Arthritis (CSHQ-RA) short form instrument. *Arthritis Rheum*. 2004;51:358–364.
- 13** Public Health Service and Health Care Financing Administration. *International Classification of Diseases, 9th Revision, Clinical Modification*. 4th ed. Washington, DC: Public Health Service; 1991.
- 14** Mahoney FI, Barthel DW. Functional evaluation: the Barthel index. *Md State Med J*. 1965;14:61–65.
- 15** Hsueh IP, Lee MM, Hsieh CL. Psychometric characteristics of the Barthel activities of daily living index in stroke patients. *J Formos Med Assoc*. 2001;100:526–532.
- 16** Hsueh IP, Lin JH, Jeng JS, Hsieh CL. Comparison of the psychometric characteristics of the Functional Independence Measure, 5 item Barthel Index, and 10 item Barthel Index in patients with stroke. *J Neurol Neurosurg Psychiatry*. 2002;73:188–190.
- 17** Fugl-Meyer AR, Jaasko L, Leyman I, et al. The post-stroke hemiplegic patient, 1: a method for evaluation of physical performance. *Scand J Rehabil Med*. 1975;7:13–31.
- 18** Duncan PW, Propst M, Nelson SG. Reliability of the Fugl-Meyer assessment of sensorimotor recovery following cerebrovascular accident. *Phys Ther*. 1983;63:1606–1610.
- 19** Sanford J, Moreland J, Swanson LR, et al. Reliability of the Fugl-Meyer assessment for testing motor performance in patients following stroke. *Phys Ther*. 1993;73:447–454.
- 20** McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res*. 1995;4:293–307.
- 21** Nunnally JC, Bernstein IH. *Psychometric Theory*. New York, NY: McGraw-Hill; 1994.
- 22** Lohr KN, Aaronson NK, Alonso J, et al. Evaluating quality-of-life and health status instruments: development of scientific review criteria. *Clin Ther*. 1996;18:979–992.
- 23** Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull*. 1955;52:281–302.
- 24** Guilford JP. *Psychometric Methods*. New York, NY: McGraw-Hill; 1954.
- 25** 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.
- 26** Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–310.
- 27** Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis*. 1987;40:171–178.
- 28** Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis*. 1985;38:27–36.
- 29** Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care*. 1989;27:178–189.
- 30** Hébert R, Spiegelhalter DJ, Brayne C. Setting the minimal metrically detectable change on disability rating scales. *Arch Phys Med Rehabil*. 1997;78:1305–1308.