Detecting change in patients with stroke using the Berg Balance Scale

Ted J Stevenson

St Boniface General Hospital, Winnipeg, Canada

The Berg Balance Scale (BBS) was designed to help determine change in functional standing balance over time. The purpose of this paper was to estimate the minimum detectable change score (MDC) using the standard error of measure (SEM), thereby providing a means to decide if genuine change had occurred. Calculation of the agreement regarding the presence of change as determined by the MDC and clinicians' perceptions was performed to give an indication of the validity of this criterion value. Forty-eight subjects who were receiving inpatient rehabilitation after stroke were assessed on consecutive days by two raters using the BBS. The MDC analysis suggests that a change of ±6 BBS points is necessary to be 90% confident of genuine change. Only 25/45 subjects showed agreement between the statistically derived presence of change and clinicians' perceptions of change. The lack of agreement may relate to the validity of the SEM/MDC methodology to determine the criterion BBS value, the heterogeneity of the subjects, or the use of clinician gestalt impressions of change. **[Stevenson TJ (2001): Detecting change in patients with stroke using the Berg Balance Scale. *Australian Journal of Physiotherapy* 47: 29-38]**

Key Words: Cerebrovascular Disorders; Equilibrium; Outcome Assessment (Health Care)

Introduction

Physiotherapists often prescribe interventions to address balance impairments that result from cerebrovascular accident (CVA; Carr et al 1994, Nilsson and Nordholm 1992). Many authors have proposed that physiotherapists adopt a problem solving approach to developing these treatment plans (Agency for Healthcare Research and Quality 1995, Carr and Shepherd 1982, Shumway-Cook et al 1996, Umphred 1985). A key feature of this approach is the need for careful assessment and reassessment in order to be confident of accurate and precise problem identification and correct recognition of a change in the individual's status. Standardised outcome measures can be used to help make these decisions.

Physiotherapists working in the Stroke Unit of St. Boniface General Hospital, Winnipeg, Canada (SBGH) routinely use the Berg Balance Scale (BBS; Berg et al 1989) to assess the standing balance of patients with CVA. Table 1 shows the BBS tasks. A subject's performance on each task is graded with a 5-point ordinal scale ranging from 0 to 4 with higher scores awarded on the basis of speed, stability or degree of assistance required for completion of the task.[a] The task scores are summed to give a total

BBS score out of a possible 56 points with higher scores representing better balance. The decision to adopt this tool was made following an examination of the process of development, determination of the feasibility of administration at SBGH and a review of the literature establishing reliability and validity.

Briefly, the BBS was developed through a process that used interviews with rehabilitation professionals and individuals with balance deficits to generate a pool of 38 balance items. Items were then excluded systematically on the basis of perceived usefulness and clarity of the items, consideration of internal consistency and examination of reliability until the final 14 tasks remained (Berg et al 1989). Since the BBS was designed explicitly for clinical use (Berg et al 1989), it is not surprising that it has been incorporated into clinical practice. Generally, it is applied within 20 minutes and requires a minimum of readily available equipment (step, stopwatch, ruler, chair).

The initial examination of reliability employed videotaped performances of the BBS graded by novice raters resulting in a calculated ICC = 0.98 for the total BBS score (Berg et al 1989). This finding of excellent reliability was later replicated by Berg et al

Table 1. Berg Balance Scale tasks.

Task

Sit to stand

Stand unsupported for 120 seconds

Sit unsupported for 120 seconds

Stand to sit

Transfers

Stand with eyes closed for 10 seconds

Stand with feet together for 60 seconds

Reach forward with an outstretched arm

Retrieve object from the ground

Turn to look behind

Turn 360 degrees

Place alternate foot on stool

Stand with one foot in front of the other for 30 seconds

Stand on one foot for 10 seconds

Table 2. The Functional Ambulation Classification (Holden et al 1984).

Category	Definition
0 Nonfunctional Ambulation	Patient cannot ambulate, ambulates in parallel bars only, or requires supervision or physical assistance from more than one person to ambulate safely outside of parallel bars.
1 Ambulator-Dependent for Physical Assistance - Level II	Patient requires manual contacts of no more than one person during ambulation on level surfaces to prevent falling. Manual contacts are continuous and necessary to support body weight as well as maintain balance and/or assist co-ordination.
2 Ambulator-Dependent for Physical Assistance - Level I	Patient requires manual contact of no more than one person during ambulation on level surfaces to prevent falling. Manual contact consists of continuous or intermittent light touch to assist balance or co-ordination.
3 Ambulator-Dependent for Supervision	Patient can physically ambulate on level surfaces without manual contact of another person but for safety requires standby guarding of no more than one person because of poor judgment, questionable cardiac status, or the need for verbal cueing to complete the task.
4 Ambulator-Independent Level Surfaces Only	Patient can ambulate independently on level surfaces but requires supervision or physical assistance to negotiate any of the following: stairs, inclines, or non-level surfaces.
5 Ambulator-Independent	Patient can ambulate independently on non-level and level surfaces, stairs, and inclines.

(1995) using a test-retest format to produce a within-rater ICC of 0.97 using 28 elderly subjects and a between-rater ICC of 0.98 using 35 subjects with acute CVA. Additionally, Liston and Brouwer (1996) have reported excellent test-retest reliability (ICC = 0.98) in 22 individuals with CVA, and Bogle Thorbahn and Newton (1996) inferred good between-rater reliability from a Spearman rho of 0.88 for 17 elderly individuals assessed by two raters.

A considerable amount of evidence suggests that the BBS is a valid measure of the standing balance of individuals with CVA. Liston and Brouwer (1996) found a Pearson correlation (r) of 0.81 between total BBS scores and gait speed, and Richards et al (1995) reported an r of 0.60 between BBS scores and gait speed expressed as a percentage of normal. Berg et al (1992b) reported correlations in excess of 0.80 between total BBS scores and total and sub-scale Barthel Index (BI) scores for individuals with CVA at various times post-CVA. Additionally, Berg et al (1992b) reported correlations in excess of 0.70 between total BBS scores and total Fugl-Meyer Scale (FMS) scores and in excess of 0.84 with scores on the Balance Subscale of the FMS. Finally, moderate correlations between total BBS scores and

performance on the Functional Independence Measure (FIM) were found by Juneja et al (1998; $r = 0.57$ at admission and 0.53 at discharge) and Wee et al (1999; $r = 0.76$). In the laboratory setting, total

BBS scores have shown statistically significant correlations on the order of 0.50 or greater with various force platform indicators of postural control in the static condition and during internally produced perturbations (Liston and Brouwer 1996, Niam et al 1999, Stevenson and Garland 1996). Evidence of the validity of the BBS as a measure of standing balance in elderly individuals with non-specific balance deficits has also been produced (Berg et al 1992a, Piotrowski and Cole 1994, Podsiadlo and Richardson 1991, Willems and Vandervoort 1994).

Since the introduction of this tool in the early 1990s, the BBS has been used widely for a variety of functions. Harada et al (1995a) investigated the usefulness of the BBS as a screening tool and found that BBS scores of less than 48/56 identified individuals who required further detailed physical therapy assessment with 91% sensitivity and 70% specificity. Bogle Thorbahn and Newton (1996) examined the usefulness of the BBS in a predictive function and reported that BBS scores of less than 45/56 predicted falls in elderly individuals with 53% sensitivity and 93 per cent specificity. Both Juneja et al (1998) and Wee et al (1999) found that BBS score on admission to acute inpatient rehabilitation was a significant predictor of length of stay. However, the strength of the relationship was insufficient for the regression equations to be clinically useful ($R^2 = 0.36$ and 0.28, respectively). Finally, the BBS has been used in an evaluative manner in projects examining the effectiveness of different interventions (Duncan et al 1998, Harada et al 1995b, Richards et al 1993, Shumway-Cook et al 1997).

At SBGH, physiotherapists use the BBS as a means through which change in patient status can be determined in two situations. The first situation occurs when the BBS is administered before and after some intervention is applied and the change score is considered when the clinician attempts to establish if the intervention is effective for that individual. The second situation occurs when the BBS is used to describe an individual's standing balance in transfer notes when the individual moves to the care of another physiotherapist who could then re-administer the BBS to determine if the individual's status has changed in the interim. In both cases, consideration of the change in BBS performance is required.

It seems reasonable to expect that the BBS would be responsive to change given the magnitude of the

ICCs, which indicate the variability in BBS scores due to measurement error is low, and, therefore, most change in BBS score could be attributable to genuine change. However, since the various models of the ICC use the ratio of within-subject and between-subject variances (Shrout and Fleiss 1979), the excellent ICCs reported may mask error that is clinically significant. Furthermore, the data from Berg et al (1995) showed greater agreement between raters at the extremes of the scale compared with that in the middle third of the scale. A chart review of patients in the SBGH Stroke Unit who had been assessed with the BBS showed that scores were generally greater than 25/56 while Berg et al (1995) included subjects with scores ranging from 0 to 56/56. Therefore, concern regarding the potential for clinically important error being concealed in the ICCs reported by Berg et al (1995) seems warranted.

Evidence that the BBS is sensitive to change is provided by Wood-Dauphinee et al (1997) who examined performance on the BBS, the FMS and the BI by individuals with CVA at two, six and 12 weeks post-onset of the neurological event. The BBS was found to have greater relative efficiency than the FMS (1.0 versus 0.65) but worse than the BI (1.24 versus 1.0) for the time period between two and twelve weeks post-CVA. The BBS had greater relative efficiency than both FMS and BI (1.0 versus 0.30 versus 0.68) for the six to 12 weeks post-CVA time period. From this, the investigators concluded that the BBS was responsive to change.

While the ICC and relative efficiencies may be of use for comparison of the relative strengths of different instruments, they cannot be used by clinicians to make decisions regarding individual patients such as those that face the SBGH physiotherapists described above. The standard error of measure (SEM) is a reliability index that may be of greater use to clinicians as it expresses the error inherent in total scores in the same units as the original measurement tool. Related to the SEM is the minimal detectable change (MDC), described by Stratford et al (1996b) as the criterion amount of change in a given measure that must occur for a clinician to conclude that genuine change has occurred. A clinician who is aware of the MDC can conclude reasonably that any change score that exceeds the MDC represents genuine change in performance on that tool rather than measurement error inherent in the score.

Therefore, the purpose of this project was to use SEM/MDC analysis to estimate the magnitude of change in BBS scores necessary for a clinician to conclude that genuine change had occurred. Comparison of the SEM/MDC findings across groups of individuals recovering from CVA-related movement disorders based on degree of independence in ambulation was performed to investigate if individuals with different balance abilities required different critical values. A preliminary evaluation of the validity of the statistically-derived critical change scores was performed through calculating the agreement between the clinicians' gestalt impression of change and the presence of change according to the MDC value.

Methods

Subjects All individuals admitted to the SBGH Stroke Unit for rehabilitation were potential subjects for this project. At the time of this project, admission to the Stroke Unit was limited to individuals aged 65 years or older and medically stable following an acute CVA. These individuals presented with residual deficits that prevented return to their previous living situation but were judged to have potential to recover sufficiently to permit community living. Individuals were approached to participate in this project if they were able to give consent and were to be assessed with the BBS by the responsible physiotherapist as a part of their treatment plan. The SBGH Stroke Unit does not have a defined policy regarding when the BBS should be administered, however, a chart review revealed that, in general, the BBS is not administered before the patient is walking with minimal to moderate assistance with or without an assistive device.

Raters All physiotherapists working in the Neuro-Geriatric Division of the Department of Rehabilitation Services, SBGH, were potential raters. These clinicians participate in an annual review of the application and scoring guidelines of the BBS as part of the physiotherapy department's quality assurance program. This review is conducted by senior physiotherapists and consists of a didactic review of the development of the tool and the interpretation of the results, viewing of a training videotape developed within the department for administering and scoring the BBS and a discussion session in which all BBS tasks and scoring are reviewed. All physiotherapists who contributed data to this project had completed their basic training a minimum of five years prior to the start of this project and were familiar with the administration and scoring of the BBS.

Protocol Following identification of potential subjects and receipt of their consent to participate in the project, the physiotherapist responsible for the patient's care administered the BBS as per established guidelines[a]. At SBGH, physiotherapists allow three attempts at each task with the individual's best performance used as the score for that task. At this time, the responsible physiotherapist also assigned a Functional Ambulation Classification (FAC; Holden et al 1984) score to the patient based on the overall assessment of the patient. Table 2 contains the descriptions for each of the FAC levels. Holden et al (1984) demonstrated that various temporal and spatial parameters of gait performance correlated with FAC score in individuals with neurological disorders suggesting it is a valid means of classifying ambulation status. Stevenson (1999) has shown that different raters can assign FAC scores in a consistent fashion with a Kappa of 0.85 describing the agreement for two raters. Demographic information and data pertaining to the stroke were gathered from the individual's medical record. A second rater, selected from the pool of available raters on the basis of convenience, re-applied the BBS within 24 hours of the initial assessment. Raters were formally blinded to each other's findings. The patient was then treated by the responsible physiotherapist for what was believed to be a reasonable time (usually one to two weeks). Since this project was not attempting to establish the effectiveness of any intervention, the type, intensity, or length of intervention was not standardised. Following this period of intervention, the responsible physiotherapist described the degree of change in the individual's balance since the initial BBS application on a 5-point scale with the following verbal anchors: much worse, slightly worse, no different, slightly better or much better. A third application of the BBS was then performed by the responsible physiotherapist after making this gestalt impression. The protocol was approved by the Committee on the Use of Human Subjects in Research, Faculty of Medicine, University of Manitoba and the SBGH Research Committee.

Analysis Subjects were classified into three groups based on FAC score: those with FAC scores of two or less were deemed to require physical assistance (ASSIST); those with an FAC scores of three were
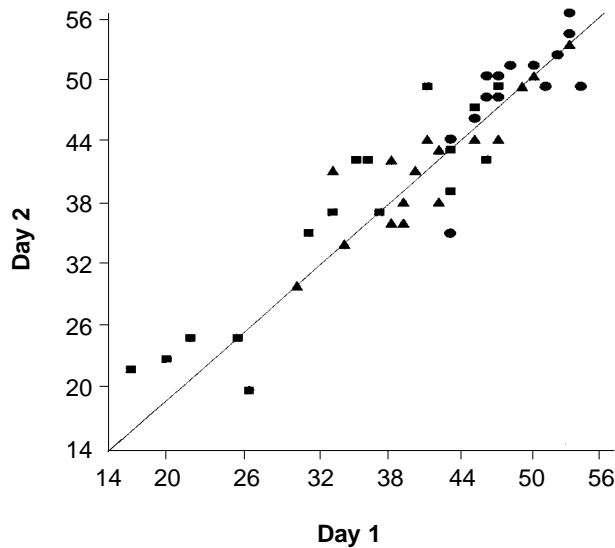
Figure 1. Scatterplot with line of equivalence of Berg Balance Scale scores (/56) on consecutive days for all subjects (■ denotes subjects who required assistance for ambulation; ▲ denotes subjects who required standby assistance for ambulation; ● denotes subjects who were independently ambulant).

deemed to require stand-by assistance (SBA); and, those with FAC scores of four or greater were deemed to be independently ambulant (INDEP). Comparison of BBS performance across functional sub-groups was done using Kruskal-Wallis ANOVA on ranks; Dunn's method was used for the post-hoc comparisons. Comparison of BBS performance within functional sub-groups from Time 1 to Time 3 was made using the Wilcoxon Signed Rank Test. The alpha level was set at 0.05. Non-parametric testing was selected for comparison between groups to acknowledge the ordinal nature of total BBS scores.

Although the use of a pool of second raters would normally imply the use of the (1,1) version of the ICC, the reliability of BBS scores between the raters for the total group was described with the (2,1) version of the ICC to permit direct comparison with the findings reported by Berg et al (1995). One way analysis of variance (ANOVA) tables for these calculations were generated using SigmaStat software; equations for generating the ICC (2,1) were taken from Shrout and Fleiss (1979). Using the method described by Stratford and Goldsmith (1997), the SEM and MDC values for the total group and sub-

groups were calculated. The MDC values at both 90 and 95% confidence levels (MDC_{90} and MDC_{95}, respectively) were calculated to provide clinicians with the option of being 90 or 95% confident that genuine change in BBS performance had occurred. The SEM is the square root of the error variance from the ANOVA table and the MDC is the product of the SEM, the tabled z-score for the desired confidence level and $\sqrt{2}$. While non-parametric methods were used to detect differences among groups, the SEM/MDC analysis requires the use of parametric methods (ANOVA). The distributions of the BBS data were tested for normality using the Kolmogorov-Smirnov (K-S) test with a critical value of 0.05 to minimise concern regarding the appropriateness of applying the ANOVA to the BBS data.

Because of the lack of variability in degree of improvement as perceived by the responsible physiotherapist at Time 3, the data were collapsed to give two groupings of subjects: those who improved and those who did not. Percentages of true positives and true negatives were calculated to describe the agreement between clinicians' perceptions of improvement and presence of improvement according to whether an individual's change in BBS performance from Time 1 to Time 3 exceeded the calculated MDCs.

Results

Forty-eight individuals consented to participate in this project, forming functional sub-groups of 15, 17 and 16 subjects for the INDEP, SBA, and ASSIST sub-groups. Three subjects (one from each sub-group) were not followed to the third application of the BBS because of unanticipated discharge from the hospital. Description of the total group and each sub-group is presented in Table 3. A total of nine physiotherapists made at least one BBS application with the majority of testing (78/96) done by the four physiotherapists who worked in the Stroke Unit over the time when data were collected.

Significant differences in BBS performance across functional sub-groups were found at Time 1 ($H_{(2)} = 20.7$, $p < 0.001$); pairwise comparisons using Dunn's method showed significant differences between INDEP vs ASSIST ($Q = 4.47$, $p < 0.05$) and INDEP vs SBA ($Q = 3.07$, $p < 0.05$) but not for SBA vs ASSIST ($Q = 1.49$, $p > 0.05$). There were no statistically significant differences for the total group

Table 3. Description of subjects as a total group and by functional grouping (mean values with standard deviations in parentheses except BBS Scores which are expressed as median values with interquartile ranges in parentheses). *: $n = 45$; †: $n = 14$; ‡: $n = 16$; #: $n = 15$; BBS = Berg Balance Scale.

	All Subjects ($n = 48$)	Independent ($n = 15$)	Standby Assist ($n = 17$)	Assist ($n = 16$)
Gender (M:F)	24:24	6:9	10:7	8:8
Age in years	73.5 (7.0)	71.0 (7.4)	75.2 (4.6)	74.1 (8.3)
CVA Location				
Right	19	8	8	3
Left	26	7	7	12
Brainstem	3	0	2	1
Time Post-CVA in days	30.3 (23.3)	18.3 (9.8)	27.5 (21.4)	44.6 (27.3)
BBS Scores (/56)				
Time 1	43.0 (36.5 - 47.0)	47.0 (45.3 - 51.8)	40.0 (38.0 - 45.3)	35.5 (25.5 - 43.0)
Time 2	42.5 (36.5 - 49.0)	49.0 (46.5 - 51.0)	41.0 (36.0 - 44.0)	38.0 (25.5 - 42.8)
Time 3	46.0 (42.0 - 51.3)*	52.5 (49.0 - 55.0)†	46.0 (43.0 - 48.0)‡	41.0 (29.0 - 47.5)#

or the functional sub-groups from Time 1 to Time 2 (ALL: $W = 187.0$, $p = 0.21$; INDEP: $W = 36.0$, $p = 0.27$; SBA: $W = 0.0$, $p = 1.0$; and ASSIST: $W = 49$, $p = 0.09$) although the difference in the ASSIST group approached significance at the 0.05 level.

All Time 1 and Time 2 data groups passed the K-S test for normality ($p > 0.05$) suggesting that the data were grossly normally distributed. The Time 1 versus Time 2 data for all subjects plotted in Figure 1 give an ICC (2,1) of 0.92. The SEM and MDC values for these data are reported in Table 4. The clinical interpretation of the MDC findings shows minor differences across the functional groups. At the 90% confidence level, the change scores necessary for the conclusion of genuine change are: ALL ± 6; INDEP ± 6; SBA ± 5; and ASSIST ± 7 BBS points.

All groups showed statistically significant increases in BBS performance from Time 1 to Time 3 (ALL: $W = 774.0$, $p < 0.001$; INDEP: $W = 78$, $p < 0.001$; SBA: $W = 92.0$, $p = 0.002$; and ASSIST: $W = 106$, $p = 0.001$). Clinicians perceived improvement in balance ability over this time frame in 34/45 (76%) subjects (INDEP = 10/14; SBA = 13/16; ASSIST =

11/15); no change perceived in 10/45 (22%) of subjects (INDEP = 4/14; SBA = 2/16; ASSIST = 4/15). A single SBA subject was perceived by the clinician to have deteriorated. Percentages of agreement between clinicians' perceptions and statistical determination of changes are presented in Table 4.

Discussion

Kirshner and Guyatt (1985) have described three purposes of health assessment instruments. First, the instrument can be used to discriminate among individuals to form homogenous groups. Second, instruments can be used to predict an individual's state at some time in the future. Finally, instruments may be used in a longitudinal manner to evaluate change in an individual's status. Kirshner and Guyatt (1985) rationalise further that different measurement issues will take on different degrees of importance depending on the proposed substantive application of the tool. For example, it is proposed that the ability to detect a difference when one is present (responsiveness) has greater importance for tools used for evaluative purposes than those used for discriminative or predictive purposes.

Table 4. Summary table of reliability indices and percent agreement between MDC-derived assignment of change and clinicians' perceptions of change (SEM = Standard Error of Measurement in BBS units; MDC = Minimal Detectable Change in BBS units for 90% and 95% confidence levels; *: $n = 45$; †: $n = 14$; ‡: $n = 16$; #: $n = 15$).

	All Subjects ($n = 48$)	Independent ($n = 15$)	Standby Assist ($n = 17$)	Assist ($n = 16$)
SEM	2.49	2.26	2.15	2.93
MDC_{90}	5.8	5.3	5.0	6.8
Agreement with Clinicians	25 (56%)*	7 (50%)†	8 (50%)‡	10 (67%)#
MDC_{95}	6.9	6.3	6.0	8.1
Agreement with Clinicians	23 (51%)*	6 (43%)†	7 (44%)‡	7 (47%)#

The BBS was designed with the explicit intention of providing a means to determine change in balance ability over time (Berg et al 1989). The reports of excellent between- and within-rater reliability (Berg et al 1995) and responsiveness to change (Wood-Dauphinee et al 1997) suggest that the BBS is suitable for this function. However, the ICCs, relative efficiencies, effect sizes and correlations between change scores do not provide an estimate of the change score necessary to conclude the presence of genuine change. Using the method described by Stratford and Goldsmith (1997), this project estimated the MDC to be 5.8 and 6.9 BBS points at the 90 and 95% confidence levels for subjects receiving rehabilitation for post-CVA deficits when assessed by two different raters within 24 hours. This is consistent with MDCs of 5.3 and 6.2 BBS points (90% and 95% confidence levels) estimated from the graphed between-rater data reported by Berg et al (1995). The clinical interpretation of these analyses holds that a minimum absolute change score of six BBS points is necessary to be 90% confident that an individual's BBS performance has changed when assessed by two different raters. This seems to be the first published attempt to establish or critically examine this criterion value for the BBS.

This project considered differences in MDCs across different functional groupings of the subjects, primarily because the between-rater data from Berg et al (1995) showed obvious differences in consistency of assessment with greater variability evident in the middle zones of the scale versus the extremes. One possible explanation for this comes from Stevenson and Garland (1996) who found that individuals with longstanding CVA who score approximately 25 to 45/56 on the BBS demonstrate greater inconsistencies in the postural strategies used during forward arm flexion compared with those individuals at the extremes of the scale. This could translate into increased variability in performance of balance tasks such as those included in the BBS. The strategy of using the best of three attempts as the subject's score for that BBS item attempts to minimise this possibility. Differences in MDC at different points within a scale is not without precedent. Stratford et al (1996a) have shown that the MDC of the Roland-Morris Questionnaire (RMQ) is conditional on the initial RMQ score.

The clinical relevance of this inconsistency in reliability across the BBS was demonstrated by the different MDCs calculated for subjects with different functional abilities. For example, the ASSIST group barely failed to demonstrate a statistically significant change in BBS scores from Time 1 to Time 2 ($p = 0.09$) suggesting that this group of patients was not as stable from Time 1 to Time 2 as the other functional sub-groups. The MDC analysis makes a more clinically useful translation of this by suggesting that individuals post-CVA who require assistance for ambulation must show a change of seven BBS points for the clinician to be 90% confident that genuine change has occurred. Conversely, the SBA group, who seemed to be more stable from Time 1 to Time 2, would require a change of only five BBS points for the clinician to be 90% confident of the presence of genuine change.

The lack of agreement between the clinicians'

perceptions of change and the statistical determination of change is noteworthy but not unexpected, given some nuances of the methodology. First, the MDC values were determined in a between-rater format while the clinicians' determinations of change was a within-rater situation. Assuming there would have been greater consistency in BBS scores had the MDC been calculated using a "within-rater" format, the present MDC probably overestimates what would be useful to clinicians using the BBS to evaluate change over time. The smaller "within-rater" MDC probably would result in increased agreement with clinicians' perceptions of genuine change. Second, the use of gestalt or global assessments of change has limitations (Norman et al 1997). Of particular interest to this project is the lack of evidence of the reliability and validity of the global rating scale and the difficulty in making an independent decision regarding the presence of change in a particular attribute. However, in the absence of an accepted methodology for determining the magnitude of change necessary for conclusion of genuine change, it seems this problem will require an accumulation of evidence before consensus on its solution is formed. Therefore, for this project, the gestalt rating provided one method for evaluating the legitimacy of the statistically derived conclusion.

The consensus of the physiotherapists at SBGH who make use of this tool is that a minimum of five BBS points is necessary before there is confidence in concluding there is a genuine change in patient status. Willems (1999) shares this opinion. Parenthetically, using the SEM found in this project, and the MDC equation listed in the methods section, a clinician can be 84% confident that a five point change in BBS performance represents genuine change. Using five BBS points as the criterion change value, there was agreement in 26/45 subjects (58%) with clinicians' perception of change. It is interesting to note that the SBGH physiotherapists, without knowledge of this project's findings, make allowances for scores at different sections of the scale with greater change necessary for lower BBS scores. For example, the SBGH clinicians were more likely to conclude genuine change had occurred when a patient's BBS score goes from 50 to 55/56 versus 20 to 25/56. This project confirms these clinicians' intuition through demonstration of different MDCs based on ambulatory ability with the largest MDC found for individuals requiring physical assistance for ambulation and, generally, lower BBS scores.

Future research As the clinical use of outcome measures becomes more routine, it seems logical that clinicians will demand that the utility of the information be clearly stated. This requires that both the designers of tools and the clinicians applying the tools move beyond describing a tool's measurement characteristics to demonstrating its usefulness in making clinical decisions. The statistical approach used by this project is limited to consideration of change scores in terms of variability of performance. It does not address the concept of clinically important difference defined by Sackett et al (1985) as the change score that is important to the patient and/or clinician. Further investigations regarding the functional correlates of both absolute and change scores for the BBS are indicated in order to comment on what constitutes a clinically significant difference.

Another aspect of the methodology that requires consideration relates to the implicit assumption that the calculated MDCs are symmetrical, ie the absolute change score necessary to conclude genuine change is the same for both improvement and deterioration in performance. This assumption should be questioned, given this project's bias towards improvement in performance as opposed to deterioration. Two aspects of the subject recruitment indicate this. First, all subjects were screened for the presence of positive indicators and absence of negative indicators for improvement before admission to the Stroke Unit. Therefore, individuals judged unlikely to benefit from stroke rehabilitation were less likely to be admitted to the unit. Second, the subjects were approached to participate in this project if the responsible physiotherapist had decided that a BBS would be helpful in developing a treatment plan. This indicates an expectation of improvement in the subject's balance. This bias towards improvement is demonstrated by the finding that only 11/45 subjects were perceived to have deteriorated or remained the same from Time 1 to Time 3. Therefore, caution should be exercised in applying these findings when deterioration of performance is under consideration. A future project including subjects expected to deteriorate as well as improve would increase confidence that the MDC values are symmetrical.

Further caution regarding the statistical approach to quantifying error employed in this project stems from the need for the analysed data to possess interval or ratio properties. This may be problematic for the BBS, since the total score represents the sum of the

ordinal scores of the 14 tasks. Merbitz et al (1989) argue that subjecting ordinal data to arithmetic operations is fundamentally illogical and lays the foundation for faulty decision making. Although total BBS scores have been largely treated as interval or ratio data in publications through the use of parametric statistics (ANOVA, *t*-test, *r*, ICC), there seems to be no evidence supporting this assumption. The finding that the present data were normally distributed does not justify the use of parametric methods but rather, gives an indication that their use was not excessively inappropriate. Further thought and investigation regarding the implications of misapplying the statistical methods is needed.

Finally, it is noteworthy that the American Physical Therapy Association Task Force on Standards for Measurement in Physical Therapy (1991) requires users of any assessment instrument to consider the limitations of the findings (see Standards U44.4 and U44.5). Therefore, regardless of the process through which the criterion value for concluding genuine change in BBS performance is generated, it is necessary that this value be examined for its sensitivity, specificity, positive predictive value and negative predictive value, to allow clinicians to use the results in an informed manner.

Conclusions

This project attempts to estimate the change score that is necessary for a clinician to conclude that genuine change has occurred in an individual patient. The results demonstrate that a difference of five to seven BBS points is necessary to conclude with 90% certainty that patients receiving rehabilitation following CVA have undergone a real change in BBS performance when assessed in a between-rater situation. However, the agreement between this benchmark and clinicians' perceptions was poor, possibly because of inconsistency in how change was determined (within- vs between-rater) and lack of standardisation of the manner in which clinicians attributed change. Further research to establish clinically useful criterion BBS scores is required.

Footnote (a) The Berg Balance Scale administration and scoring guidelines are available on the Internet at www.chcr.brown.edu/balance.htm.

Author Ted J Stevenson, Rehabilitation Services, St. Boniface General Hospital, 409 Taché, Winnipeg, Manitoba, Canada R2H 2A6. E-mail: tstevens@sbgh.mb.ca (for correspondence).

References

Agency for Healthcare Research and Quality (1995): Post-Stroke Rehabilitation. Clinical Practice Guidelines Number 16. Rockville: Agency for Health Care Policy and Research, Public Health Service, US Department of Health and Human Services.

Berg K, Wood-Dauphinee S and Williams JI (1995): The Balance Scale: reliability assessment with elderly residents and patients with an acute stroke. *Scandinavian Journal of Rehabilitation Medicine* 27: 27-36.

Berg K, Wood-Dauphinee S, Williams JI and Gayton D (1989): Measuring balance in the elderly: preliminary development of an instrument. *Physiotherapy Canada* 41: 304-311.

Berg KO, Maki BE, Williams JI, Holliday PJ and Wood-Dauphinee SL (1992a): Clinical and laboratory measures of postural balance in an elderly population. *Archives of Physical Medicine and Rehabilitation* 73: 1073-1080.

Berg KO, Wood-Dauphinee SL, Williams JI and Maki BE (1992b): Measuring balance in the elderly: validation of an instrument. *Canadian Journal of Public Health* 83(Suppl 2): S7-S11.

Bogle Thorbahn LD and Newton RA (1996): Use of the Berg balance test to predict falls in elderly persons. *Physical Therapy* 76: 576-585.

Carr JH, Mungovan SF, Shepherd RB, Dean CM, Nordholm D and Nordholm LA (1994): Physiotherapy in stroke rehabilitation: bases for Australian physiotherapists' choice of treatment. *Physiotherapy Theory and Practice* 10: 201-209.

Carr JH and Shepherd RB (1983): A Motor Relearning Programme for Stroke. Rockville: Aspen Publishers.

Duncan P, Richards L, Wallace D, Stoker-Yates J, Pohl P, Luchies C, Ogle A and Studenski S (1998): A randomized, controlled pilot study of a home-based exercise program for individuals with mild and moderate stroke. *Stroke* 29: 2055-2060.

Harada N, Chiu V, Damon-Rodriguez J, Fowler E, Siu A and Reuben DB (1995a): Screening for balance and mobility impairment in elderly individuals living in residential care facilities. *Physical Therapy* 75: 462-469.

Harada N, Chiu V, Fowler E, Lee M and Reuben DB (1995b): Physical therapy to improve functioning of older people in residential care facilities. *Physical Therapy* 75: 830-839.

Holden MK, Gill KM, Magliozzi MR, Nathan J and Piehl-Baker L (1984): Clinical gait assessment in the neurologically impaired. *Physical Therapy* 64: 35-40.

Juneja G, Czyrny JJ and Linn RT (1998): Admission balance and outcomes of patients admitted for acute inpatient rehabilitation. *American Journal of Physical Medicine and Rehabilitation* 77: 388-393.

Kirshner B and Guyatt G (1985): A methodological framework for assessing health indices. *Journal of Chronic Disease* 38: 27-36.

Liston RAL and Brouwer BJ (1996): Reliability and validity of measures obtained from stroke patients using the Balance Master. *Archives of Physical Medicine and Rehabilitation* 77: 425-430.

Merbitz C, Morris J and Grip JC (1989): Ordinal scales and foundations of misinference. *Archives of Physical Medicine and Rehabilitation* 70: 308-312.

Niam S, Cheung W, Sullivan PE, Kent S and Gu X (1999): Balance and physical impairment after stroke. *Archives of Physical Medicine and Rehabilitation* 80: 1227-1233.

Nilsson L and Nordholm L (1992): Physical therapy in stroke rehabilitation: bases for Swedish physiotherapists' choice of treatment. *Physiotherapy Theory and Practice* 8: 49-55.

Norman GR, Stratford PW and Regehr G (1997): Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *Journal of Clinical Epidemiology* 50: 869-879.

Piotrowski A and Cole J (1994): Clinical measures of balance and functional assessment in elderly persons. *Australian Journal of Physiotherapy* 40: 183-188.

Podsiadlo D and Richardson S (1991): The timed "up and go": a test of basic functional mobility for frail elderly persons. *Journal of the American Geriatric Society* 39: 142-148.

Richards CL, Malouin F, Wood-Dauphinee S, Williams JI, Bouchard J-P and Brunet D (1993): Task-specific physical therapy for optimization of gait recovery in acute stroke patients. *Archives of Physical Medicine and Rehabilitation* 74: 612-620.

Richards CL, Malouin F, Dumas F and Tardif D (1995): Gait velocity as an outcome measure of locomotor recovery after stroke. In Craik RL and Oatis CA (Eds) Gait Analysis: Theory and Applications. St Louis: Mosby, pp. 355-64.

Sackett DL, Haynes RB and Tugwell P (1985): Clinical Epidemiology. A Basic Science for Clinical Medicine. Boston: Little, Brown and Co.

Shrout PE and Fleiss J (1979): Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* 86: 420-428.

Shumway-Cook A, Gruber W, Baldwin M and Liao S (1997): The effect of multidimensional exercises on balance, mobility, and fall risk in community-dwelling older adults. *Physical Therapy* 77: 46-57.

Shumway-Cook A and Woollacott MH (1996): Motor Control: Theory and Practical Applications. Baltimore: Williams and Wilkins, pp. 99-116.

Stevenson TJ (1999): Using Impairment Inventory scores to determine ambulation status in individuals with stroke. *Physiotherapy Canada* 51: 168-174.

Stevenson TJ and Garland SJ (1996): Standing balance during internally produced perturbations in subjects with hemiplegia: validation of the balance scale. *Archives of Physical Medicine and Rehabilitation* 77: 656-672.

Stratford PW, Binkley J, Solomon P, Finch E, Gill C and Moreland J (1996a): Defining the minimum level of detectable change for the Roland-Morris Questionnaire. *Physical Therapy* 76: 359-365.

Stratford PW, Finch E, Solomon P, Binkley J, Gill C and Moreland J (1996b): Using the Roland-Morris Questionnaire to make decisions about individual patients. *Physiotherapy Canada* 48: 107-110.

Stratford PW and Goldsmith CH (1997): Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Physical Therapy* 77: 745-50.

Task Force on Standards for Measurement in Physical Therapy (1991): Standards for tests and measurements in physical therapy practice. *Physical Therapy* 71: 589-622.

Umphred DA (1985): Conceptual model: a framework for clinical problem solving. In Umphred DA (Ed.): Neurological Rehabilitation. St Louis: CV Mosby Company, pp. 3-25.

Wee JYM, Bagg SD and Palepu A (1999): The Berg Balance Scale as a predictor of length of stay and discharge destination in an acute stroke rehabilitation setting. *Archives of Physical Medicine and Rehabilitation* 80: 448-452.

Willems D and Vandervoort AA (1996): Balance as a contributing factor to gait speed in rehabilitation of the elderly. *Physiotherapy Canada* 48: 179-184.

Willems D (1999): Outcome measures chart. *Synapse: Newsletter of the Neurosciences Division. Canadian Physiotherapy Association* 19: 20-23.

Wood-Dauphinee SL, Berg K, Bravo G and Williams JI (1997): The Balance Scale: responsiveness to clinically meaningful changes. *Canadian Journal of Rehabilitation* 10: 35-50.