**Title page:**

**Responsiveness of the Berg Balance Scale in patients early after stroke**

Adam Saso PT, MSc1,2, Rolf Moe-Nilssen PT, PhD3, Mari Gunnes PT2, Torunn Askim PT, PhD2,4

1Clinical Services, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway.

2Department of Neuroscience, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway.

3Department of Global Public Health and Primary Care, Faculty of Medicine and Dentistry, University of Bergen, Bergen, Norway.

4Department of Physiotherapy, Faculty of health Science and Social Work, Sør-Trøndelag University Collage, Trondheim, Norway.

Address to corresponding author: Adam Saso, Dyre Halsesgate 10, 7042 Trondheim, NORWAY

ABSTRACT

The Berg Balance Scale (BBS) has previously shown good measurement properties. However, its ability to detect important change in patients early after stroke is still unknown. The purpose of the present study was to determine the minimal important change (MIC) and its relation to the minimal detectable change (MDC) for BBS in patients early after stroke. This prospective follow-up study, included patients within the first 2 weeks after onset of stroke. The BBS, Barthel Index and Scandinavian Stroke Scale were obtained at inclusion and one month later. At the follow-up assessment, the Patient Global Impression of Change was obtained. A Receiver Operating Characteristic (ROC) curve was used to calculate the cut-off value for the MIC. Fifty-two patients, (mean age of 78.7, SD 8.5 years), were included. All measures showed a significant improvement from baseline to follow-up. The ROC analysis identified a MIC of ≥ 6 BBS-points, while the MDC was 5.97 BBS-points at the 80 % confidence level. This study shows that a change of 6 BBS point or more can be considered an important change for patients in the sub-acute phase after stroke, which also represents a 80 % probability of exceeding the measurement error. 80 % of unchanged patients would display random fluctuations within the bounds of MDC80, while 20 % of unchanged patients would exceed MDC80.

INTRODUCTION

Despite good treatment in the acute phase, many patients who have suffered from a stroke have to live with their disabilities ([Collaboration, 2000](#_ENREF_12)). Disabilities after stroke varies, and could be motor impairments, loss of sensitivity, language problems, depression, reduced cognitive function and/or pain. The most frequent motor impairments are total paralysis or reduced power in the arm and/or leg in one side of the body, impaired motor control, reduced speed, disturbed automation, increased fatigue, uncoordinated movements, altered balance self-efficacy, reduced balance control and worry about falls (falls self-efficacy). Such poststroke mobility impairments are linked to decreased activity and participation ([Benaim et al., 1999](#_ENREF_4); [Lawrence et al., 2001](#_ENREF_33); [Lord et al., 2004](#_ENREF_37); [D'Alisa et al., 2005](#_ENREF_13); [Schmid and Rittman, 2009](#_ENREF_43)).

Balance control is a requirement that most motor tasks have in common and could be defined as the ability to maintain the position of the body within the limits of stability or base of support. However, the specific stability requirements vary according to the task and the environment ([Shumway-Cook and Woollacott, 2001](#_ENREF_44)). It is shown that stance stability, as well as the ability to compensate for external and internal perturbations and to control posture voluntarily, improves along with functional recovery during the first three months, while the recovery curve levels out in the later stages of stroke ([Jorgensen et al., 1995a](#_ENREF_27); [Geurts et al., 2005](#_ENREF_22)). It is also shown that the recovery curve varies between the different severity groups of stroke, with a faster and steeper curve for those with mild strokes compared to those who are severely affected ([Jorgensen et al., 1995c](#_ENREF_29)). This reality implies that the patients’ perception of an important change probably will vary between different severity groups and also change along the course of recovery.

The Berg Balance Scale (BBS) was developed to detect change in patients' balance performance over time ([Berg et al., 1989](#_ENREF_6)), and has been identified as the most commonly used assessment tool for evaluating balance in rehabilitation after stroke ([Blum and Korner-Bitensky, 2008](#_ENREF_9)). This extended utilization of the BBS in research and clinical practice demands a thorough exploration of psychometric properties, to be reported in a clinically relevant and applicable manner.

The BBS has demonstrated excellent inter-tester, intra-tester and test-retest reliability in patients with chronic stroke ([Liang, 2000](#_ENREF_35); [Mao et al., 2002](#_ENREF_39); [Revicki et al., 2008](#_ENREF_41)) and also high to excellent inter-tester and intra-tester reliability in patients with acute stroke ([Liang, 2000](#_ENREF_35); [Chou et al., 2006](#_ENREF_10); [Downs et al., 2013](#_ENREF_16)). Furthermore, the BBS has been found to be a valid instrument for evaluation of the effectiveness of interventions and for quantitative descriptions of function in clinical practice and research ([Berg et al., 1992](#_ENREF_7); [Stevenson, 2001](#_ENREF_46)). The internal consistency has shown to be excellent in several studies ([Liang, 2000](#_ENREF_35); [Chou et al., 2006](#_ENREF_10); [English et al., 2006](#_ENREF_17)). However, others have suggested directions for improving the internal validity by improving the rating scale structure, improving the standardization or excluding some items ([Kornetti et al., 2004](#_ENREF_31); [La Porta et al., 2012](#_ENREF_32); [Straube et al., 2013](#_ENREF_50)).

When interpreting clinical measures, it is important to consider that although small changes may be statistically significant, they may not be clinically relevant ([Hays and Woolley, 2000](#_ENREF_25); [Beaton et al., 2001](#_ENREF_3)). Small changes may also be due to measurement errors. Therefore smallest detectable change or minimal detectable change (MDC) is defined as the change beyond the measurement error at different confidence levels (95%, 90%, 80% and so on) ([de Vet et al., 2011](#_ENREF_14)). On the other hand, minimal important change (MIC) is defined by the international Delphy study to achieve consensus based standards for the selection of health measurement instruments (the COSMIN study) as the smallest change in score in the construct to be measured which patients (or clinicians) perceive as important ([de Vet et al., 2011](#_ENREF_14)).

Distribution-based methods, which are based on statistical characteristics of the sample, like the test-retest reliability, are most commonly used for calculation of MDC, while anchor-based methods, which uses an external criterion, are strongly recommended for calculation of MIC ([de Vet et al., 2011](#_ENREF_14)). The most commonly used external criteria are global rating scales assessing the amount of perceived change, i.e. asking the patient how much they have improved or deteriorated over a certain time period. In the next step a cut-off is set defining a certain amount of change as an important change ([de Vet et al., 2011](#_ENREF_14)). Responsiveness should be tested by relating the MDC to the MIC ([Terwee et al., 2007](#_ENREF_52); [Stratford and Riddle, 2012](#_ENREF_48)).

Responsiveness is a dynamic and context-specific concept and derivations of the MDC and MIC are usually estimated for a specific population at a particular stage of recovery ([Wells et al., 2001](#_ENREF_55)). An anchor-based method comparing a patient’s self-perceived improvement with a reference standard of change, has the potential of being understood more clearly because changes are related to a clinical observation ([Stratford et al., 1996](#_ENREF_47)).

The MDC of BBS has been determined in stroke patients undergoing inpatient rehabilitation showing that a change of 6 BBS points was needed to confirm that a change beyond the measurement error had taken place with 90% probability ([Stevenson, 2001](#_ENREF_46)). Giving it a 10 % chance of exceeding measurement error for MDC90.

Recently the anchor-based MIC has been estimated for BBS in patients with balance disorders following different neurological diseases, including patients with hemiparesis undergoing rehabilitation, giving a MIC of 7 BBS points for this compound group of patients ([Godi et al., 2013](#_ENREF_23)). However, the distribution based MDC and the anchor-based MIC has not been reported for the BBS in patients early after stroke, nor for stroke patients at different severity levels.

Hence, the primary aim of the present study was to determine the minimal important change (MIC) and its relation to the minimal detectable change (MDC) for BBS in patients early after stroke. In the present study, patients who rated themselves as ‘very much better’ or ‘much better’ were classified as responders experiencing an important change, while those who rated themselves as ‘slightly better’ or ‘unchanged’ were classified as non-responders.

The secondary aim was to determine the minimal important change for subgroups of patients with stroke.

METHODS

Study design

This study was a prospective follow-up study, with an initial assessment within the first two weeks after onset of stroke and a reassessment one month later. Measurements assessing balance, body function, activity level and global functions were applied at both occasions. For the patients' subjective experience of change in balance, an anchor instrument was used.

Subjects

All individuals admitted to the Stroke Unit at St. Olav`s Hospital, Trondheim University Hospital, who were available on two predefined study days every second week and diagnosed with stroke, were potential subjects for this project. Individuals were invited to participate in the project if the patient was admitted within 14 days after onset of symptoms, had reduced balance measured as BBS score < 56 and was able to give informed consent to participation. Patients were given oral and written information about the study. If needed the patient was given one day to consider participation in the study. Patients were excluded if they suffered from severe aphasia or received palliative care.

Measurements

At the level of body function, Scandinavian Stroke Scale (SSS) was used to examine the neurological outcome. It consists of nine items considering the patient's level of consciousness, orientation, motor impairments, language function and mobility. Total scores range from 0 to 58, where 58 represents normal function ([Christensen et al., 2005](#_ENREF_11)). Inter-tester reliability for the nine items vary with a weighted Kappa score (κw) from 0.688 to 0.912 ([Christensen et al., 2005](#_ENREF_11)). SSS predicts imposed destination and disability at discharge in all patients with stroke ([Jorgensen et al., 1995b](#_ENREF_28)). Patients with initial SSS scores ranging from 58 to 45 were classified with mild strokes, while patients with SSS scores ranging from 44 to 30 were classified with moderate strokes. Finally, patients with initial SSS scores ranging from 29 to 0 were classified with severe stroke.

At the level of activity, Berg Balance Scale (BBS), which was the measure of primary interest, was used to assess balance, while Barthel Index was used to evaluate activities of daily living. BBS consists of 14 different tasks assessing balance in sitting and standing position and in transfer ([Berg et al., 1989](#_ENREF_6)). Each motor task is rated on a five-point scale from 0 to 4. The total score ranges from 0-56, where 56 represents normal balance ([Berg et al., 1992](#_ENREF_7); [Berg et al., 1995](#_ENREF_5)). The psychometric properties of BBS have been thoroughly investigated, showing very good reliability ([Berg et al., 1995](#_ENREF_5); [Liston and Brouwer, 1996](#_ENREF_36); [Mao et al., 2002](#_ENREF_39); [Downs et al., 2013](#_ENREF_16)) and excellent internal consistency ([Berg et al., 1995](#_ENREF_5); [Mao et al., 2002](#_ENREF_39); [Chou et al., 2006](#_ENREF_10)). It is also shown that BBS has the ability to discriminate between sub groups of patients with stroke ([Berg et al., 1992](#_ENREF_7); [Stevenson, 2001](#_ENREF_46); [Au-Yeung et al., 2003](#_ENREF_2)). The Norwegian version of BBS applied in the present study has been translated according to international recommendations without any contextual changes of the 14 tasks. Furthermore, it has demonstrated high inter-tester reliability and internal consistency ([Halsaa et al., 2007](#_ENREF_24)).

Barthel Index (BI) consists of ten different items of activities of daily living. Total scores range from 0-100, where 100 represents normal scores on all test items ([Mahoney and Barthel, 1965](#_ENREF_38)). Wolfe et al. (1991) reported κw = 0.96 for inter-tester reliability and κw = 0.98 for test-retest reliability measured at two weeks intervals for patients with stroke of varying severity ([Wolfe et al., 1991](#_ENREF_56)).

Patient Global Impression of Change (PGIC) is a commonly used anchor instrument. It has shown to be valid for perceived change in different studies ([Jaeschke et al., 1989](#_ENREF_26); [Farrar et al., 2001](#_ENREF_19); [Farrar et al., 2010](#_ENREF_18)), also including gait and balance ([Smedal et al., 2006](#_ENREF_45)). PGIC is a seven-point scale in which the patient indicates how he/she experiences the current situation, compared to a given set point some time back. Examples of this could be a hospitalization, a previous measurement or the time before rehabilitation. In the present study patients were asked the following question; *“Compared to your stay at the stroke unit, how do you experience your balance today?”* The response categories were as follows; 1) Very much better; 2) Much better; 3) Slightly better; 4) Unchanged; 5) Slightly worse; 6) Much worse; 7) Very much worse. Furthermore, patients were dichotomized into two groups based on their ratings on PGIC. Patients who reported PGIC score of 1 'very much improved', 2 'much better', were categorized to have experienced an important change and are referred to as responders. Those with scores 3 'slightly better', 4 'unchanged' or 5 'slightly worse' were judged not to have experienced any significant change and were categorized as non-responders. Those with scores 6 'much worse' or 7 'very much worse' would be excluded from the study ([Terwee et al., 2010](#_ENREF_54)).

Procedure

Patients who met the inclusion criteria and gave informed consent were included and a baseline assessment was conducted. Demographic data and information on stroke was obtained from the patient and the patient's medical record. At baseline the SSS, BBS and BI were used. All tests were conducted according to the guidelines provided in the test manuals. At discharge from the stroke unit, the patient's future residency was recorded.

One-month follow-up assessment was conducted in hospital, at the rehab clinic or in the patients’ home, depending on patients’ discharge residency. SSS and BBS were assessed by clinical assessment, whereas BI was determined in consultation with the patient and caregivers. Finally, the PGIC was used to measure the patient's subjective experience of change in balance over the last month.

Four independent assessors, who were trained in performing all the measures applied in the present study, conducted the baseline and follow-up assessments. The follow-up assessors were blinded to the baseline assessment.

Ethical approval

Informed consent was obtained from all participants. The study was approved by the Regional Committee for Medical and Health Research Ethics in Norway (REK no 2011/2053).

Analysis

The data were analyzed in Statistical package for Social Sciences (IBM SPSS), version 19 and the free software programming language R. Clinical and demographic data were analyzed by descriptive statistics and all variables were tested for normality. Sensitivity analyses were performed to check for possible outliers. For variables that complied with the normal distribution a paired t-test was used while for variables that did not satisfy the requirement of normal distribution Wilcoxon signed rank test was used to analyze changes in test scores from baseline to one-month follow-up.

The Mann-Whitney U-test was used to analyze differences between responders and non-responders according to the dichotomized PGIC as described in the Measurements section. For variables satisfying the requirement of normal distribution Student's t-test for two independent samples was used, while Chi-square test was used to analyze differences between groups for categorical variables.

Receiver operating characteristic (ROC) curve was used to define the anchor-based MIC by identifying the point on the ROC curve nearest to the upper left hand corner, which is considered to be the best cutoff score for distinguishing responders from non-responders. The sensitivity of the test was defined as the proportion of positives that were correctly identified by the test, while the specificity was defined as the proportion of negatives that were correctly identified by the test ([Altman, 1990](#_ENREF_1)). In this study, this means the proportion of the patients with a defined BBS change score properly identified by the PGIC to be a responder. Furthermore, the positive predictive value was defined as the proportion of patients with positive test result who were correctly identified by the test, while the negative predictive value was defined as the proportion of patients with negative test result who were correctly identified by the test ([Altman, 1990](#_ENREF_1)). The predictive values standardized to a 50% pretest chance of responding were presented.

Area under the curve (AUC) is a measure of the average diagnostic accuracy found for all values of the test under investigation ([Fawcett, 2006](#_ENREF_20)). AUC gives an indication of the test's ability to discriminate and should be at least 0.70 ([Terwee et al., 2003](#_ENREF_53)). AUCs less than 0.70 has been defined as poor accuracy, 0.70 to 0.89 as adequate and AUC ≥ 0.90 as excellent accuracy ([StrokeEngine, 2011](#_ENREF_51)). Bootstrap confidence intervals were computed for the MIC values (Terwee et al., 2010), while the 95% confidence intervals (CI) have been computed for the sensitivity, specificity and predictive values, by use of CI calculator ([MedCalc, 2015](#_ENREF_40)).

The Standard Error of Measurement (SEM) is a measure of reliability expressed in the same units as the original measurement tool, giving an indication of absolute reliability ([Finch et al., 2002](#_ENREF_21)). This makes the SEM value more applicable for clinicians working with individual patients on a daily basis, because the measure is easy to interpret for the clinical decision-making. Based on the SEM value, the minimal detectable change (MDC) can be calculated.

The SEM value was calculated as the square-root of the mean within-subject variance from an ANOVA that included the subsample of non-responders in the present study. The factors were patients and occasions. Since the within-subject variances have a chi-square distribution (Armitage and Berry, 1994), a 95 % CI for the SEM was estimated following the procedures of Stratford and Goldsmith (1997). The critical values of the chi-square distribution at the 0.975 and 0.025 levels (df = 33) was used to estimate the 95% CI of the mean within-subject variance, from which the 95% CI of the SEM was obtained by taking the square root. Furthermore the Type 2.1 intraclass correlation coefficient (ICC) was calculated and 95 % CI was estimated based on the subsample of non-responders.

The measurement error can be calculated as ±1.96\*SEM, representing the 95% confidence interval for the SEM. The MDC can be calculated as repeatability, and is defined as or 2.77 SEM ([Bland and Altman, 1996](#_ENREF_8)). The difference between two measurements for the same subject is expected to be less than this MDC for 95 % of pairs of observations ([Finch et al., 2002](#_ENREF_21)). A similar approach was used to estimate MDCs for levels of 90%, 80%, 70% and 50%, where 2.77 in the formula for MDC was replaced with 2.32, 1.81, 1.47 and 0.95, respectively (calculated by replacing 1.96 with 1.64 for 90%, with 1.28 for 80%, 1.04 for 70% and with 0.67 for 50% as appropriate z-scores taken from a normal distribution table).

For an internal validating purpose, the correlation between the rating on PGIC and the change score on BBS was calculated by use of Spearman’s rank correlation*, rs*. The acceptability limits were defined as; poor ≤ 30; adequate 0.31-0.59 and excellent ≥ 0.60 ([StrokeEngine, 2011](#_ENREF_51)). Bootstrapped confidence interval with 5000 iterations was estimated for Spearman's correlation coefficient which has an awkward distribution that can deviate from a normal distribution also after Fisher's Z transformation. Prior to pooling, the Fisher Z transform is applied to the correlations. After pooling, the inverse Z transform is applied. The p-value for the hypothesis that *rs* = 0 will also be reported.

RESULTS

A total of 89 patients were screened for inclusion. Fourteen were excluded due to high BBS score (they had no deficits and scored 56/56), while eight were excluded due to severe aphasia. A total of 67 were enrolled and completed the baseline assessment. Thirteen patients did not complete the one-month follow-up assessment. The main reasons were; withdrew from the study (n=11), not available for testing due to a long stay abroad (n=1) and death (n=1). Furthermore, the sensitivity analyses revealed two extreme outliers. These two cases were removed from the dataset leaving a total of 52 participants, 23 women and 29 men.

*Insert table 1 about here*

Table 1 shows the baseline characteristics of the included patients. The mean age was 78 years. Eighty-nine percent suffered from a mild and moderate stroke and the median time from onset of stroke to inclusion was 6 days.

Statistically significant improvements in neurological outcome was found as measured by the SSS, in ADL function as measured by BI and in balance as measured by the BBS from baseline to one-month follow-up. BBS had an increase in the total score of 26.3% between the two test occasions (table 2). The mean BBS score at baseline was 24.9 (SD 20.0).

*Insert table 2 about here*

Table 3 shows that thirty-four patients had experienced some degree of subjective improvement of balance at the one-month follow-up compared to baseline, while sixteen patients experienced their balance as unchanged. Two patients experienced their balance as slightly worse. No patients found their balance as 'much worse' or 'very much worse'. Based on the pre-defined classification of PGIC 18 patients were categorized as responders, while 34 patients were categorized as non-responders (table 3).

The ICC(2,1) (95% CI) and SEM (95% CI) calculated from the non-responders were 0.98 (0.96 to 0.99) and √10.97059 = 3.3 (2.7 to 4.4) respectively, giving a MDC of 6.5 BBS points at the 95% confidence level and a MDC of 5.97 BBS points at the 80% confidence level. The MDC values for confidence levels from 95% to 50% are reported in table 4.

The correlation between the rating on PGIC and the BBS change score was adequate with *rs* = -0.428, 95% CI -0.152 to -0.637, p=0.002.

*Insert table 3 about here*

The ROC curve analysis specified an optimal cut off of 5.5 BBS change points, which gives an anchor-based MIC of ≥ 6 BBS points, with a sensitivity of 72.2% and specificity of 85.3%. The AUC (95% CI) was 0.777 (0.623 to 0.931), p=0.001 (figure 1 and table 4).

Thirteen out of the 18 responders revealed a BBS change score of 6 points or more. The positive and negative predictive values for a standardized prevalence of 50 % were 0.83 and 0.75 respectively.

*Insert table 4 and figure 1 about here*

Based on SSS measured at baseline, patients were stratified into three subgroups of severity of stroke (table 1). Patients classified with mild stroke showed corresponding results with the whole group, while the AUC (95% CI) for the moderately affected patients revealed no difference in the distribution of responders versus non-responders and a cut-off value was not determined. No patients classified with severe stroke were categorized as responders and the ROC curve analysis could not be performed. Results from the subgroup analysis are reported in table 4.

DISCUSSION

In the present study we have applied two different methods for estimating change threshold values, one anchor based method using ROC curve analysis to identify the patient’s perception of an important change and a distribution based method using the reliability methodology to identify measurable change beyond the measurement error. The main finding from the anchor-based method was that a 6 point change on BBS was rated as an important change by patients in the early phase after stroke. The distribution based method revealed a SEM of 3.3 BBS points, producing a MDC80 of 5.97 BBS points, which indicates that 20 % of truly unchanged patients will display random fluctuations on re-test greater than an apparent 6 point change is needed to be 80 % confident that the change is beyond the measurement error as well. The non-responder sample distribution used to estimate MDC is the same distribution on which specificity is estimated. Assuming this distribution is consistent with a normal distribution, a specificity of 85 % obtained by the anchor based method corresponds to a two-tailed 70 % confidence level or an MDC70 of 4.85. The value of the MDC indicates the percentage chance that unchanged patients would display random fluctuations within the bounds of the MDC value. For example, an MDC80 will give 80 % of unchanged patients would display random fluctuations within the bounds of an MDC80, while 20 % of unchanged patients would exceed MDC80.

The anchor-based MIC of 6 BBS points reported in the present study is one point less than the MIC reported by Godi et al (2013) in a group of elderly with balance deficits undergoing rehabilitation ([Godi et al., 2013](#_ENREF_23)). The median BBS score at baseline in Godi’s study was more than 15 points higher than the median BBS score at baseline reported in the present study. In general, patients with very poor balance in the early phase after stroke will be more likely to experience a significant improvement in their balance, as a result of both early rehabilitation and spontaneous recovery, compared to patients in the later phase when the recovery curve levels out.

The MDC of 7.65 BBS points revealed at the 90 % confidence level is somewhat higher than the MDC of 5.8 points at the 90% confidence level revealed by Stevenson (2001) ([Stevenson, 2001](#_ENREF_46)). These discrepancies are probably caused by poorer balance and greater variability, giving a greater SEM value in patients early after stroke compared to patients in the later phase (mean time from onset of stroke was 6 days in the present study compared to 30 days in the study by Stevenson (2001)). On the other hand these discrepancies might not represent a real difference but could be a result of sampling variation between the two studies.

On the contrary, previous subgroup analyses have shown that the MDC at the 95% confidence level varies between patients at different severity levels ([Downs et al., 2013](#_ENREF_16)), ranging from 8.1 points in stroke patients dependent on assistance in walking ([Stevenson, 2001](#_ENREF_46)) to 3.3 points in older people with a BBS score of 45 points or better ([Donoghue and Stokes, 2009](#_ENREF_15)). We were not able to confirm such differences in the present study. The subgroup analysis was performed by stratifying patients into three different severity groups according to SSS. The ROC curve estimated for those with mild and moderate stroke showed that the BBS had a better ability to detect changes in those with mild stroke compared to those with moderate stroke, who revealed no difference in the distribution for the AUC. Estimating a ROC curve was not feasible for those with severe stroke, because everyone in the group reported 'unchanged' on PGIC. If a cut-off value ≥ 6 BBS points is assumed, data in this group showed 100 % agreement, as all patients had a change in BBS < 6 points. Because of the small sample of only six patients with severe stroke, one should be careful in drawing conclusions based upon these findings. Furthermore, the subgroup analysis showed the same MIC for patients with mild stroke as for the whole group, except for a narrower confidence interval, which is in line with a previous study by Mao et al. (2002) showing moderate to high sensitivity to change for patients with mild and moderate stroke ([Mao et al., 2002](#_ENREF_39)). However, the MIC should be further assessed in subgroups of patients in future research.

There was a statistically significant improvement from the baseline assessment to one-month follow-up for all measuring instruments, including neurological impairments and activities of daily living and balance. This suggests that the improvement is not the result of chance, but the result of a real change. Most patients with stroke who receive treatment in a stroke unit have decreased function with an expected spontaneous recovery of neurological function and body activity during the first three months after stroke ([Jorgensen et al., 1995a](#_ENREF_27); [Jorgensen et al., 1995b](#_ENREF_28)). However, patients with severe stroke typically have a slower recovery curve ([Jorgensen et al., 1995c](#_ENREF_29)). This was also confirmed by the results from the present study showing that no patients with severe stroke experienced any important change in balance, while four out of 14 and 14 out of 32 patients with moderate and mild stroke respectively, experienced a change of their BBS scores. In the present study, a ROC curve was constructed to define a cut-off value of BBS that represents an important change. Out of several recommended methods to examine the MIC of a measuring device, ROC curve is recommended as the superior method ([Lehman and Velozo, 2010](#_ENREF_34); [Terwee et al., 2010](#_ENREF_54)). However, a possible limitation of this method is the risk of losing information, when transforming the PGIC ratings into dichotomous data in order to conduct the ROC analysis. A priori, important change was defined as ‘very much better’ or ‘much better’ on the patient's global impression of change in balance. One can also argue that 'slightly better' is an important change. However, to ensure that the change really represents an important change, those reporting 'slightly better' was defined as non-responders, which is in line with other studies ([Farrar et al., 2001](#_ENREF_19); [Salaffi et al., 2004](#_ENREF_42); [Terwee et al., 2010](#_ENREF_54)). On the other hand, defining the amount of change as an important change might represent a validity problem. Even though a large body of the literature have been defining amount of change as an important change, it has also been shown that these represent two different concepts ([Stratford and Riddle, 2013](#_ENREF_49)), indicating that further research is needed to more clearly distinct the amount of change from the importance of change.

Another limitation of the present study was the use of SEM based on the non-responders. A more sophisticated approach would have been to do a test-retest study within the first few days after inclusion. However, such a design is hard to perform in the acute hospital setting because patients are discharged upon a very short notice. A prerequisite for assessing reliability is that no change has occurred, and by excluding the two extreme outliers who rated themselves as ‘non-responders’ despite a change of 20 and 30 BBS points respectively, only five out of the remaining 34 non-responders obtained a BBS change score of 6 points or more. Even though the non-responders complied with the normality assumption the total sample did not comply so well, and the SEM and MDCs should therefore be interpreted with caution.

A final limitation is the collection of data from a single hospital, which reduces the external validity.

Three different meanings of health related measurement tools are described by Kirshner and Guyatt (1985). They believe that a measurement tool can be used to discriminate among individuals to produce homogeneous groups, or it can be used to predict an individual's condition ahead in time, or also the measuring instrument can be used to evaluate changes in an individual's condition which is tested ([Kirshner and Guyatt, 1985](#_ENREF_30)). A measuring instrument's properties, or lack of ability, will have varying degrees of importance depending on what it is used for. For measuring instruments used to evaluate effect of an intervention, the ability to detect a change when the change has actually taken place is an important trait ([Kirshner and Guyatt, 1985](#_ENREF_30)). BBS was designed precisely with this intention, namely to identify change in balance ([Berg et al., 1989](#_ENREF_6)). This study's purpose was to determine how much change is an important change to patients in the early phase after stroke. The anchor-based MIC of 6 BBS points revealed in the present study is consistent with results on MDC found in comparable populations ([Berg et al., 1995](#_ENREF_5); [Stevenson, 2001](#_ENREF_46)), indicating good consistency between measurable change and clinical meaningful change on BBS.

Based on these findings, the use of BBS in a clinical setting to assess whether a change in a patient's balance has occurred is strengthened. A 6 points change to determine whether an important change has taken place or not can be used to inform clinicians whether one should make a change in the intervention if desired effect has not occurred. Inevitably, clinical experiences and a thorough clinical assessment of the individual patient and the patient's perception of their change should be heavily emphasized and taken into account in any evidence based practice. In future research, findings from this study can be used to calculate the sample size in randomized controlled trials, including patients early after stroke, using the BBS as the primary outcome.

CONCLUSION

In the present study we have applied an anchor based method to identify the patient’s perception of an important change and a distribution based method to identify change beyond the measurement error at different confidence levels. The anchor based method showed that a change greater than or equal to 6 BBS points was considered to be an important change by patients early after stroke while the distribution based method showed that a change of 5.97 BBS points is needed to be 80 % confident that the change has exceeded the measurement error. These results support the validity of using the BBS in the clinic early after stroke and will be of great value to sample size calculations in future studies using BBS as the primary outcome.

Acknowledgements

The authors want to thank Associated Professor/Statistician Øyvind Salvesen, Department of Applied Clinical Science at NTNU for his statistical advices.

Declaration of interest

The authors report no declaration of interest. The authors alone are responsible for the content and writing of the article. MSc PT Adam Saso received support from Central Norway Health Authority RHF.

**REFERENCES**

Altman DG. *Practical statistics for medical research*. 1 ed: Chapman & Hall/CRC Texts in Statistical Science, 1990.

Armitage P, Berry G. Statistical methods in medical research. 3rd ed. Oxford: Blackwell Science Ltd; 1994.

Au-Yeung SS, Ng JT, Lo SK. Does balance or motor impairment of limbs discriminate the ambulatory status of stroke survivors? Am J Phys Med Rehabil 2003; 82 (4): 279-283.

Beaton DE, Bombardier C, Katz JN, Wright JG, Wells G, Boers M, Strand V, Shea B. Looking for important change/differences in studies of responsiveness. OMERACT MCID Working Group. Outcome Measures in Rheumatology. Minimal Clinically Important Difference. J Rheumatol 2001; 28 (2): 400-405.

Benaim C, Perennou DA, Villy J, Rousseaux M, Pelissier JY. Validation of a standardized assessment of postural control in stroke patients: the Postural Assessment Scale for Stroke Patients (PASS). Stroke 1999; 30 (9): 1862-1868.

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 (1): 27-36.

Berg K, Wood-Dauphinee S, Williams JI, Gayton D. Measuring balance in the elderly: preliminary development of an instrument. Physiotherapy Canada 1989; 41 (6): 304-311.

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: S7-11.

Bland JM, Altman DG. Measurement error. BMJ 1996; 313 (7059): 744.

Blum L, Korner-Bitensky N. Usefulness of the Berg Balance Scale in stroke rehabilitation: a systematic review. Phys Ther 2008; 88 (5): 559-566.

Chou CY, Chien CW, Hsueh IP, Sheu CF, Wang CH, Hsieh CL. Developing a short form of the Berg Balance Scale for people with stroke. Phys Ther 2006; 86 (2): 195-204.

Christensen H, Boysen G, Truelsen T. The Scandinavian stroke scale predicts outcome in patients with mild ischemic stroke. Cerebrovasc Dis 2005; 20 (1): 46-48.

Collaboration SUT. Organised inpatient (stroke unit) care for stroke. . Cochrane Database Syst Rev 2000; (2): CD000197.

D'Alisa S, Baudo S, Mauro A, Miscio G. How does stroke restrict participation in long-term post-stroke survivors? Acta Neurol Scand 2005; 112 (3): 157-162.

de Vet HC, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicine: A practical guide*: Cambridge University Press, 2011.

Donoghue D, Stokes EK. How much change is true change? The minimum detectable change of the Berg Balance Scale in elderly people. J Rehabil Med 2009; 41 (5): 343-346.

Downs S, Marquez J, Chiarelli P. The Berg Balance Scale has high intra- and inter-rater reliability but absolute reliability varies across the scale: a systematic review. J Physiother 2013; 59 (2): 93-99.

English CK, Hillier SL, Stiller K, Warden-Flood A. The sensitivity of three commonly used outcome measures to detect change amongst patients receiving inpatient rehabilitation following stroke. Clin Rehabil 2006; 20 (1): 52-55.

Farrar JT, Pritchett YL, Robinson M, Prakash A, Chappell A. The clinical importance of changes in the 0 to 10 numeric rating scale for worst, least, and average pain intensity: analyses of data from clinical trials of duloxetine in pain disorders. J Pain 2010; 11 (2): 109-118.

Farrar JT, Young JP, Jr., LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001; 94 (2): 149-158.

Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters 2006; 27 (8): 861-874.

Finch E, Brooks D, Stratford PW. *Physical rehabilitation outcome measures. A guide to enhanced clinical decision-making*. Hamilton, Ontario: Lippincott Williams & Wilkins, 2002.

Geurts AC, de Haart M, van Nes IJ, Duysens J. A review of standing balance recovery from stroke. Gait Posture 2005; 22 (3): 267-281.

Godi M, Franchignoni F, Caligari M, Giordano A, Turcato AM, Nardone A. Comparison of reliability, validity, and responsiveness of the mini-BESTest and Berg Balance Scale in patients with balance disorders. Phys Ther 2013; 93 (2): 158-167.

Halsaa KE, Brovold T, Graver V, Sandvik L, Bergland A. Assessments of interrater reliability and internal consistency of the Norwegian version of the Berg Balance Scale. Arch Phys Med Rehabil 2007; 88 (1): 94-98.

Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research. How meaningful is it? Pharmacoeconomics 2000; 18 (5): 419-423.

Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989; 10 (4): 407-415.

Jorgensen HS, Nakayama H, Raaschou HO, Olsen TS. Recovery of walking function in stroke patients: the Copenhagen Stroke Study. Arch Phys Med Rehabil 1995a; 76 (1): 27-32.

Jorgensen HS, Nakayama H, Raaschou HO, Vive-Larsen J, Stoier M, Olsen TS. Outcome and time course of recovery in stroke. Part I: Outcome. The Copenhagen Stroke Study. Arch Phys Med Rehabil 1995b; 76 (5): 399-405.

Jorgensen HS, Nakayama H, Raaschou HO, Vive-Larsen J, Stoier M, Olsen TS. Outcome and time course of recovery in stroke. Part II: Time course of recovery. The Copenhagen Stroke Study. Arch Phys Med Rehabil 1995c; 76 (5): 406-412.

Kirshner B, Guyatt G. A methodological framework for assessing health indices. J Chronic Dis 1985; 38 (1): 27-36.

Kornetti DL, Fritz SL, Chiu YP, Light KE, Velozo CA. Rating scale analysis of the Berg Balance Scale. Arch Phys Med Rehabil 2004; 85 (7): 1128-1135.

La Porta F, Caselli S, Susassi S, Cavallini P, Tennant A, Franceschini M. Is the Berg Balance Scale an internally valid and reliable measure of balance across different etiologies in neurorehabilitation? A revisited Rasch analysis study. Arch Phys Med Rehabil 2012; 93 (7): 1209-1216.

Lawrence ES, Coshall C, Dundas R, Stewart J, Rudd AG, Howard R, Wolfe CD. Estimates of the prevalence of acute stroke impairments and disability in a multiethnic population. Stroke 2001; 32 (6): 1279-1284.

Lehman LA, Velozo CA. Ability to detect change in patient function: responsiveness designs and methods of calculation. J Hand Ther 2010; 23 (4): 361-370; quiz 371.

Liang MH. Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments. Med Care 2000; 38 (9 Suppl): II84-90.

Liston RA, Brouwer BJ. Reliability and validity of measures obtained from stroke patients using the Balance Master. Arch Phys Med Rehabil 1996; 77 (5): 425-430.

Lord SE, McPherson K, McNaughton HK, Rochester L, Weatherall M. Community ambulation after stroke: how important and obtainable is it and what measures appear predictive? Arch Phys Med Rehabil 2004; 85 (2): 234-239.

Mahoney FI, Barthel DW. Functional Evaluation: The Barthel Index. Md State Med J 1965; 14: 61-65.

Mao HF, Hsueh IP, Tang PF, Sheu CF, Hsieh CL. Analysis and comparison of the psychometric properties of three balance measures for stroke patients. Stroke 2002; 33 (4): 1022-1027.

MedCalc. MedCalc easy-to-use statistical software. 2015; Available from: <http://www.medcalc.org/calc/diagnostic_test.php> [Accessed 03.03.15].

Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008; 61 (2): 102-109.

Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. Eur J Pain 2004; 8 (4): 283-291.

Schmid AA, Rittman M. Consequences of poststroke falls: activity limitation, increased dependence, and the development of fear of falling. Am J Occup Ther 2009; 63 (3): 310-316.

Shumway-Cook A, Woollacott M. *Motor Control: Theory and Practical Applications*. 2 ed: Lippincott Williams & Wilkins, 2001.

Smedal T, Lygren H, Myhr KM, Moe-Nilssen R, Gjelsvik B, Gjelsvik O, Strand LI. Balance and gait improved in patients with MS after physiotherapy based on the Bobath concept. Physiother Res Int 2006; 11 (2): 104-116.

Stevenson TJ. Detecting change in patients with stroke using the Berg Balance Scale. Aust J Physiother 2001; 47 (1): 29-38.

Stratford PW, Binkley JM, Riddle DL. Health status measures: strategies and analytic methods for assessing change scores. Phys Ther 1996; 76 (10): 1109-1123.

Stratford PW, Goldsmith CH. Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. Phys Ther 1997 Jul;77(7):745-50.

Stratford PW, Riddle DL. When minimal detectable change exceeds a diagnostic test-based threshold change value for an outcome measure: resolving the conflict. Phys Ther 2012; 92 (10): 1338-1347.

Stratford PW, Riddle DL. Assessing the amount of change in an outcome measure is not the same as assessing the importance of change. Physiother Can 2013; 65 (3): 244-247.

Straube D, Moore J, Leech K, Hornby TG. Item analysis of the berg balance scale in individuals with subacute and chronic stroke. Top Stroke Rehabil 2013; 20 (3): 241-249.

StrokeEngine. Statistical Evaluation Criteria for Outcome Measures. 2011; Available from: <http://strokengine.ca/assess/statistics-en.html> [Accessed 27.02.2015].

Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007; 60 (1): 34-42.

Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. Qual Life Res 2003; 12 (4): 349-362.

Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, Croft P, de Vet HC. Mind the MIC: large variation among populations and methods. J Clin Epidemiol 2010; 63 (5): 524-534.

Wells G, Beaton D, Shea B, Boers M, Simon L, Strand V, Brooks P, Tugwell P. Minimal clinically important differences: review of methods. J Rheumatol 2001; 28 (2): 406-412.

Wolfe CD, Taub NA, Woodrow EJ, Burney PG. Assessment of scales of disability and handicap for stroke patients. Stroke 1991; 22 (10): 1242-1244.

Table 1. Baseline characteristics

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ­­ | | |  |  | | Difference between groups (p-value) |
|  | | | All participants  n=52 | Responders  n=18 | Non-responders  n=34 |
| Age, years a | | |  |  |  |  |
|  | Mean (SD) | | 78.7 (8.6) | 74.9 (8.0) | 80.7(8.3) |  |
|  | Median | | 81.0 | 73.5 | 84.0 | 0.013 |
|  | 1st -3rd quartile | | 70.3-85.0 | 66.8-82.0 | 74.8-86.0 |  |
| Gender, n (%)  b | | | |  |  |  |
|  | Female | 23 (44.2) | | 6 (33.3) | 17 (50.0) | 0.379 |
|  | Male | 29 (55.8) | | 12 (66.7) | 17 (50.0) |  |
| First time stroke, n (%)  b | | | |  |  |  |
|  | Yes | 38 (73.1) | | 12 (66.7) | 26 (76.4) | 0.519 |
|  | No | 14 (26.9) | | 6 (33.3) | 8 (23.6) |  |
| *Time since stroke* c | | | |  |  |  |
|  | Mean | 6.4 (4.0) | | 6.4 (4.3) | 6.4 (3.8) |  |
|  | Median | 5.0 | | 6.0 | 5.0 | 0.913 |
|  | 1st -3rd quartile | 3.0-9.0 | | 3.0-9.5 | 4.0-9.5 |  |
| Affected side, n (%) b | | | |  |  |  |
|  | Right | 28 (53.8) | | 12 (66.7) | 16 (47.1) | 0.245 |
|  | Left | 24 (46.2) | | 6 (33.3) | 18 (52.9) |  |
| *Time in Stroke Unit* c | | | |  |  |  |
|  | Mean | 13.7 (8.7) | | 10.6 (8.8) | 15.3(8.3) |  |
|  | Median | 13.0 | | 8.5 | 14.0 | 0.020 |
|  | 1st -3rd quartile | 7.0-18.3 | | 4.75-13.5 | 9.0-21.0 |  |
| Discharge destination , n (%) b | | | | |  |  |
|  | Home | 24 (46.2) | | 12 (66.7) | 12 (35.3) | 0.043 |
|  | Rehab-unit | 14 (26.9) | | 5 (27.8) | 9 (26.5) | 1.000 |
|  | Nursing home | 14 (26.9) | | 1(5.5) | 13 (38.2) | 0.019 |
| *Scandinavian Stroke Scale* (SSS) c | | | | |  |  |
|  | Mean | 43.8 (11.6) | | 48.6 (5.6) | 41.2 (13.0) |  |
|  | Median | 48.5 | | 51.0 | 47.0 | 0.118 |
|  | 1st -3rd quartile | 36.3-52.0 | | 45.5-52.25 | 30.0-52.25 |  |
| Severity Groups , n (%) b | | | |  |  |  |
|  | SSS 45-58 Mild | 32 (61.6) | | 14 (77.8) | 18 (52.9) | 0.133 |
|  | SSS 30-44 Moderate | 14 (26.9) | | 4 (22.2) | 10 (29.4) | 0.746 |
|  | SSS 0-29 Severe | 6 (11.5) | | 0 (0.0) | 6 (17.7) | 0.081 |
| PGIC: Patient Global Impression of Change; SD: Standard deviation; aThe difference between groups was tested by Mann-Whitney U test b The difference between groups was tested by chi-square test | | | | | | |

Table 2. Functional outcomes at baseline and 1-month follow up

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Baseline | | | 1-month follow up | | | Change score | | | Difference  (p-value) a |
|  | | All participants (n=52) | Responders (n=18) | Non-responders (n=34) | All participants (n=52) | Responders (n=18) | Non-responders (n=34) | All participants (n=52) | Responders (n=18) | Non-responders (n=34) |  |
| Berg Balance Scale (BBS) | | | |  |  |  |  |  |  |  |  |
|  | Mean (SD) | 24.9(20.0) | 28.1(18.2) | 23.2(21.0) | 29.9(19.5) | 38.4(14.2) | 25.4(20.6) | 6.2(6.9) | 10.3(10.1) | 2.2(4.7) |  |
|  | Median  (1st - 3rd quartile) | 30.5  (0.0 - 56.0) | 31.5  (0.0-45.3) | 28.5  (0.0-43.0) | 36.0  (0.0 - 55.0) | 40.5  (32.5-51.3) | 32.0  (3.8-42.3) | 3.0  (-6.0 - 29.0) | 7.5  (3.5-20.5) | 3.0  (0.0-4.0) | 0.000 |
| Barthel Index | | |  |  |  |  |  |  |  |  |  |
|  | Mean (SD) | 59.0(29.3) | 69.4(24.1) | 53.5(30.6) | 73.0(28.2) | 85.8(15.1) | 66.2(31.2) | 13.9 (16.0) | 16.4(14.8) | 12.6(16.6) |  |
|  | Median  (1st - 3rd quartile) | 70.0  (35.0 - 80.0) | 72.5  (55.0-86.3) | 65.0  (25.0-80.0) | 82.5  (56.3 - 95.0) | 90.0  (70.0-100.0) | 75.0  (45.0-95.0) | 12.5  (0.0 - 20.0) | 15.0  (3.8-21.3) | 10.0  (0.0-20.0) | 0.000 |
| Scandinavian Stroke Scale (SSS) | | | |  |  |  |  |  |  |  |  |
|  | Mean (SD) | 43.8(11.6) | 48.6(5.6) | 41.2(13.0) | 48.0(11.7) | 53.3(4.4) | 45.2(13.3) | 4.3 (5.5) | 4.7(4.3) | 4.0(6.0) |  |
|  | Median  (1st - 3rd quartile) | 48.5  (36.3 - 52.0) | 51.0  (44.5-52.3) | 47.0  (30.0-52.3) | 53.00  (44.3 - 56.0) | 54.0  (49.0-57.0) | 49.0  (38.3-55.3) | 3.0  (1.0 - 8.0) | 3.5  (2.0-8.3) | 2.5  (0.0-6.5) | 0.000 |
| BBS by subgroups | | | |  |  |  |  |  |  |  |  |
| Mild stroke, SSS: 58 to 45 points | | | |  |  |  |  |  |  |  |  |
|  |  | (n=32) | (n=14) | (n=18) | (n=32) | (n=14) | (n=18) | (n=32) | (n=14) | (n=18) |  |
|  | Mean (SD) | 37(13.4) | 33.9(15.3) | 39.4(11.6) | 42.9(8.9) | 44.6(7.2) | 41.6(10.0) | 5.9 (9.0) | 10.6(10.7) | 2.2(5.2) |  |
|  | Median  (1st - 3rd quartile) | 40.0  (30.3 - 47.0) | 38.5  (17.5-46.3) | 42.0  (30.8-49.3) | 42.0  (36.0 - 51.8) | 42.0  (38.0-52.0) | 41.5  (34.0-50.5) | 3.5  (-0.8 - 10.3) | 7.5  (4.5-21.3) | 3.0  (-1.8-4.3) | 0.001 |
| Moderate stroke, SSS: 44 to 30 points | | | |  |  |  |  |  |  |  |  |
|  |  | (n=14) | (n=4) | (n=10) | (n=14) | (n=4) | (n=10) | (n=14) | (n=4) | (n=10) |  |
|  | Mean (SD) | 7.9 (13.8) | 7.8(12.4) | 8.0(14.9) | 12.7 (13.6) | 16.6(10.5) | 11.1(14.8) | 4.8 (5.7) | 9.0(9.0) | 3.1(3.0) |  |
|  | Median  (1st - 3rd quartile) | 0.5  (0.0 - 10.3) | 2.5  (0.0-20.8) | 0.5  (0.0-11.0) | 5.0  (4.0 - 23.5) | 17.5  (6.3-26.5) | 4.5  (3.8-14.8) | 4.0  (2.0 - 6.3) | 6.0  (2.5-18.5) | 4.0  (1.5-5.3) | 0.004 |
| Severe stroke, SSS: 30 to 0 points | | | |  |  |  |  |  |  |  |  |
|  |  | (n=6) | (n=0) | (n=6) | (n=6) | (n=0) | (n=6) | (n=6) | (n=0) | (n=6) |  |
|  | Mean (SD) | 0.0(0.0) | - | 0.0(0.0) | 1.0 (1.5) | - | 1.0(1.5) | 1.0 (1.5) | - | 1.0(1.5) |  |
|  | Median  (1st - 3rd quartile) | 0.0  (0.0 - 0.0) | - | 0.0  (0.0 - 0.0) | 0.0  (0.0 - 3.0) | - | 0.0  (0.0 - 3.0) | 0.0  (0.0 - 3.0) | - | 0.0  (0.0 - 3.0) | 0.157 |

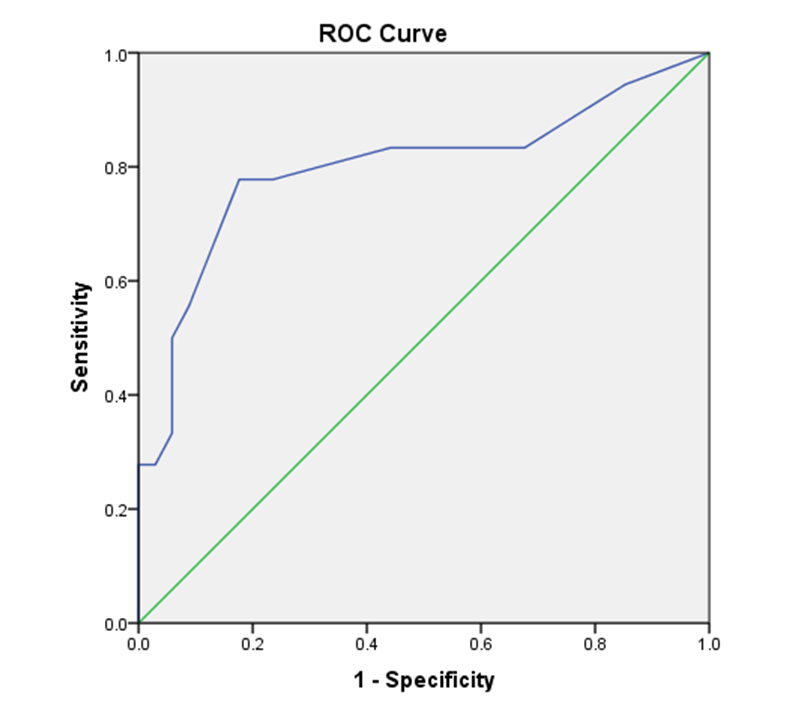
SD: Standard deviation;  aDifferences between baseline and 1-month follow up test for all participants by use of Wilcoxon Signed Rank test

Table 3. Distribution of Patient Global Impression of Change (PGIC) at 1-month follow up and median (1st to 3rd quartile) Berg Balance Scale (BBS) change score from baseline to 1-month follow up

|  |  |  |  |
| --- | --- | --- | --- |
| **PGIC**  **categories** | **Dichotomized**  **categories** | **n (%)** | **BBS**  **change score**  **Median**  **(1st to 3rd quartile)** |
| 1 Very much better | Responders | 1 (1.9) | 6.0 (6.0 to 6.0) |
| 2 Much better | 17 (32.7) | 8.0 (3.0 to 21.0) |
| 3 Slightly better | Non-responders | 16 (30.8) | 3.0 (-0.75 to 5.0) |
| 4 Unchanged | 16 (30.8) | * 1. (0.0 to 3.0) |
| 5 Slightly worse | 2 (3.8) | 4.0 (4.0 to 4.0) |
| 6 Much worse | Worse | 0 (0) | NA |
| 7 Very much worse | 0 (0) | NA |
| **Total** |  | **52 (100)** |  |

**Table 4.** Responsiveness indexes for Berg Balance Scale

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | | | **BBS** |
| **Minimal detectable change (MDC) at different confidence levels (n=34)** | | | | |
| MDC95 | | | | 9.15 |
| MDC90 | | | | 7.65 |
| MDC80 | | | | 5.97 |
| MDC70 | | | | 4.85 |
| MDC50 | | | | 3.92 |
| **Minimal important change (MIC)** | | | | |
| The ROC curve analyses | | |  | |
|  | All patients (n=52) | |  | |
|  |  | AUC (95% CI) | 0.78 (0.62 to 0.93) | |
|  |  | Sensitivity (95% CI) | 0.72 (0.47 to 0.90) | |
|  |  | Specificity (95% CI) | 0.85 (0.69 to 0.95) | |
|  |  | Optimal cutoff (95% CI) | 5.5 (3.5 to 6.5) | |
|  | Mild stroke (n=32) | |  | |
|  |  | AUC (95% CI) | 0.77 (0.59 to 0.95) | |
|  |  | Sensitivity (95% CI) | 0.79 (0.49 to 0.95) | |
|  |  | Specificity (95% CI) | 0.83 (0.59 to 0.96) | |
|  |  | Optimal cutoff (95% CI) | 5.5 (4.5 to 6.5) | |
|  | Moderate stroke (n=14) | |  | |
|  |  | AUC (95% CI) | 0.670 (0.335 to 1.00) | |
|  | Severe stroke (n=6) | | -- a | |
| MDC95=minimum detectable change at 95% confidence interval, MDC90=minimum detectable change at 90% confidence interval, MDC80=minimum detectable change at 80% confidence interval, MDC50=minimum detectable change at 50% confidence interval, AUC=area under the curve, ROC curve= receiver operating characteristic curve. aThe ROC curve could not be drawn because no patients experienced change in balance. | | | | |

****

**Figure 1.**  Receiver operating characteristics (ROC) curve generated with change in Berg Balance Scale (BBS) score versus dichotomized Patient Global Impression of Change (PGIC) score.