The Journal of Physical Therapy Science

Original Article

The test-retest reliability and minimal detectable change of the short-form Barthel Index (5 items) and its associations with chronic stroke-specific impairments

CHANG-SIK PARK, PhD, OT¹⁾

¹⁾ Department of Physical Therapy, Howon University: 64 Howondae 3gil, Impimyeon, Gunsan-si 573-932, Republic of Korea

Abstract. [Purpose] To establish the test-retest reliabilities, minimal detectable change of the Short form Barthel Index and associations with stroke-specific impairments. [Subjects and Methods] The Short form-Barthel Index assessment was tested on 24 chronic stroke patients twice, 7 days apart. A relative reliability index (ICC2,), Weighted Kappa Coefficients was used to examine the level of agreement of test-retest reliability for SF-BI, Absolute reliability indices, including the standard error of measurement and the minimal detectable change. The validity was demonstrated by spearman correlation of SF BI-total score with Postural Assessment Scale for Storke, Fugl Meyer Assessment. [Results] There was excellent agreement between test-retest for individual items of BI and total score $ICC_{2,1}=0.91$ and it all showed acceptable SEM and MDC were 2.83 score, 7.84 score respectively. The item-to-total correlations were all significant, ranging from r=0.83-0.92. SF-BI showed good internal consistency. Individual items also possessed high internal consistency 0.82-0.86. The SF-BI and total score were demonstrated high concurrent validity with the PASS, FMA. [Conclusion] This study has demonstrated that the SF-BI is a useful instrument with high test-retest reliability, Absolute reliability indices, internal consistency and validity. Key word: Activities daily living, Short form Barthel Index, Stroke

(This article was submitted Jan. 31, 2018, and was accepted Mar. 15, 2018)

INTRODUCTION

The low performance of stroke patients in activities of daily living is a major factor in deteriorating the life satisfaction levels of the patients and their caregivers¹). The improvement of independent activities of daily living is one of important therapeutic goals in occupational therapy. Therefore, it is essential for clinicians and therapists to select appropriate assessment instruments to effectively evaluate patients' performance of activities of daily living²⁻⁴). The Barthel Index⁵ and the Functional Independence Measure (FIM) have been most frequently used to assess impairment levels in domestic and foreign clinics⁶. The FIM is known to be a more comprehensive and responsive assessment instrument for scoring impairments than the BI7). However, the psychological characteristics of the two instruments are similar, suggesting that the FIM is not superior to the BI or has any particular advantages⁸). In addition, as both the instruments are composed of comprehensive and quantitative assessment items, the FIM test requires about 30 to 45 minutes and the BI test takes 20 minutes if the assessment of the performance of subjects is carried out through observation⁸). In particular, as the time required for assessment increases, the psychological burden (the decrease in muscle endurance and the increase in muscle tone) of therapists and stroke patients with neurological disorders is increased, which can leads to a problematic reliability of evaluation results³). Therefore, clinicians and researchers should be aware that if assessment takes a long time in data collection process, selective biases (random

Corresponding author. Chang-Sik Park (E-mail: 1629917@hanmail.net)

©2018 The Society of Physical Therapy Science. Published by IPEC Inc.



cc () () This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial No Deriva-NC ND tives (by-nc-nd) License. (CC-BY-NC-ND 4.0: https://creativecommons.org/licenses/by-nc-nd/4.0/)

measurement errors) may occur and the obtained scores may differ from the actual measured values^{9, 10)}. For this reason, short-form assessment instruments have been developed¹¹) by reducing unnecessary items within a range that minimizes the loss of data information without affecting the psychometric properties of the original assessment instruments. In addition, the assessment of subjects in clinical practice should be easy to apply; required time should be little; and obtained data should be quantifiable and easy to interpret¹¹). Also, psychological characteristics such as reliability, validity, and response rate should be well verified³). Hobart et al.⁸) derived the short form BI (5 items: transfers, bathing, toilet use, stair climbing and ambulation) from the original 10-item BI and examined its psychological characteristics. As a result, it was found that the SF-BI has an item internal consistency of Cronbach's a coefficient=0.888) with 0.71 at admission and 0.73 at discharge, respectively; an inter-rater reliability of ICC=0.90; and a standardized response mean (SRM) of 0.71 at average, which are the same as those of the original BI¹. In particular, the SF-BI (1.2) was reported to have a considerably high response rate compared with the original BI (1.2) and the FIM motor subscales (1.3)¹⁾. It was also reported that the concurrent validity between the SF-BI and the original BI (r=0.96) and the convergent and discriminant validities between the SF-BI and the FIM total scores (r=0.87) are significantly associated⁸⁾. In addition, the SF-BI was found to have psychological characteristics similar to those of the original BI¹) with concurrent validities of r=0.74 and 0.94 at admission and discharge, respectively, which are at satisfying levels. Nanayakkara and Lekamwasam¹²) proved that the 5-item SF-BI is a very sensitive assessment instrument that can predict or determine the independency levels of the elderly's activities of daily living (the original BI total scores for the ten items: <80 points=self-reported dependency, >80 points=self-reported independency). Thirty-seven points obtained in the SF-BI test (the maximum score: 55 points) is the cutoff value to distinguish between dependency and independency (sensitivity level 95%, specificity level 82%). Recent studies have emphasized the need for an assessment method that can identify the effects of treatment interventions and predict functional recovery in terms of treatment management for patients^{11, 13)}. The standard error measurement (SEM) and minimal detectable change (MDC) are the absolute reliability indices by which the two can be standardized and quantified¹⁴). They are used to estimate the sizes of potential random measurement errors caused by chance variations in case of score changes at the repeated same tests conducted on individuals or determine whether the measured values remain systematically consistent (95% confidence level)^{4, 14}). Particularly, the MDC can be used as important index data for clinical decision making because it can serve as the threshold value for both therapists and clinical researchers to determine the sizes and prognoses of post-treatment effects based on the actual changes in the scores of patients¹⁴). However, the test-retest reliability and MDC of the SF-BI have not been investigated in overseas studies and the concurrent validity for the posture control and upper and lower extremity movement control functions reflecting strokespecific impairments has not been known yet, either. Thus, this study was aimed at investigating the test-retest reliability and MDC of the SF-BI and its associations with stroke-specific impairments.

SUBJECTS AND METHODS

The subjects of this study were selected among chronic stroke patients who had been diagnosed with hemiplegia due to stroke and agreed to participate in this study. The subjects were those who were receiving regular medical services at G Hospital and the study period lasted from August 2016 to December 2016. The study purpose and methods were explained to the subjects, who provided informed consent according to the principles of the Declaration of Helsinki before participating. The criteria for selection of the subjects were: chronic stroke patients who had the onset of stroke more than six months ago; had obtained 24 or higher points for cognitive functions in the Mental State Examination-Korean Version (MMSE-K); and are able to understand verbal instructions. Patients unable to participate in functional performance assessment required by this study due to orthopaedic diseases were excluded. The general medical characteristics of the subjects including their ages, duration of illness, diagnoses, paralyzed areas and the MMSE-K scores were collected through admission notes and one-to-one interviews.

The sample size of this study was calculated to be 24 using the G Power (version 3.1) and through a repeated measures (twice) ANOVA at a power of the test of 95% (a significance level of 0.05) and an effect size of F=0.4. In order to minimize the effects of potential recovery on the BI in the test-retest agreement rates, patients with chronic stroke were selected as subjects¹⁰). The data were collected from the finally selected 24 patients after excluding two who did not meet the criteria for subject selection, two who were emergently discharged in the final data collection process and two who showed unreliable assessment results that caused data errors. For the test-retest reliability (ICC_{2,1}) of the BI, the agreement rates were compared in a total of two assessments performed at a weekly interval by occupational therapists with more than 16 years of clinical experiences¹⁵). The construct validity of the SF-BI was estimated by the correlation coefficient among the original BI, the PASS total scores and BBS total scores. The assessment of the subjects' functional performance was conducted by two physical therapists with 15 years of clinical experiences. The entire assessment process included two to five minute breaks as needed in order to minimize the decrease in performance due to excessive tension caused by fatigue and associated reactions along with learning effects.

The Barthel Index which is a basic daily activity assessment tool for stroke patients⁵⁾ was used. The BI consists of 10 items: personal hygiene, bathing, feeding, toilet use, stair climbing, dressing, bowel control, bladder control, ambulation or wheelchair mobility and chair/bed transfers. Each item has a five-stage scoring system according to the degree of external help, with the maximum score of 100 points. The 5-item SF-BI derived from the 10-item original BI consists of transfers,

bathing, toilet use, stair climbing and ambulation, with the maximum score of 55 points⁸⁾. It was reported that SF-BI has an item internal consistency of Cronbach's α coefficient=0.88 and an inter-rater reliability of ICC=0.90⁸).

The Postural Assessment Scale for Stroke (PASS) was developed to evaluate stroke patients' performance of postural control by modifying and supplementing the FM-B¹⁶). The PASS consists of 12 items (0–3 points) among which five items are for assessing postural maintenance and seven items for assessing postural changes in relation with three basic postures (lying, sitting and standing), and its maximum score is 36 points. It was reported that the PASS for chronic stroke patients has an inter-rater agreement of weighted kappa coefficient= $0.88 (0.61-0.96)^{17}$.

The Fugl Meyer Motor Assessment for the Upper and Lower Extremities (FMA-U/E, L/E) has been used to assess impairments in upper and lower extremity motor functions including the movements, coordination and reflexes of agonistic and synergic muscles among stroke patients. The FMA is a 3-point scale (0–2 points) and consists of 33 items to assess upper extremity motor functions (0–66 points) and 17 items to assess lower extremity motor functions (0–34 points), with the maximum score of 100 points¹⁸. The test-retest reliabilities of the FMA-U/E, L/E in stroke patients were reported to be ICC=0.98 and 0.95, respectively, with that of the total score of ICC=0.98³).

A statistical analysis was performed using the SPSS 18.0 for Windows 7 in this study. For the general characteristics of the subjects, a frequency analysis and descriptive statistics were carried out. The test-retest agreement rates of the SF-BI total scores were calculated using the intra class coefficient (ICC_{2,1}) and those of the SF-BI individual items using the Spearman correlation coefficient between the total scores of the SF-BI and of the original BI, while estimating the item internal consistency using the Cronbach's α coefficient. Also, in this study, the absolute reliability indices were calculated. To quantify random measurement errors, the standard error measurement (SEM=the standard deviation of all test-retest scores × [$\sqrt{1}$ -ICC]) was used, while the minimal detectable change (MDC=1.96 × SEM × $\sqrt{2}$) was used for the calculation to obtain the reference value determining whether the actual changed scores (treatment effect sizes) of each patient are maintained consistent within the 95% confidence interval¹⁴). The SEM is <15% of all test-retest average scores and the less its value is, the more acceptable it becomes. MDC values are reliable when they are <20% of the highest measured value¹⁰). To estimate the concurrent validity of the SF-BI total scores, the PASS total scores were used, while the correlation between the SF-BI total scores was calculated using the Spearman correlation coefficient. All statistical significance levels were set at α =0.05.

RESULTS

The SF-BI individual items' weighted kappa coefficients were as follows: transfers=0.67, bathing=0.61, toilet use=0.61, stair climbing=0.75, and ambulation=0.62. The agreement degree was good and the observed agreement rates were within a range of 71% to 80%, which was satisfactory.

The SF-BI total scores were acceptable with ICC=0.91 (0.86-0.95), SEM=2.83 (<10% of the average score, 41.66) and MDC=7.84 (<20% of the maximum score, 54).

The SF-BI individual items were found to be highly positively relevant to the SF-BI total scores (r=0.83-0.92), and to be highly relevant to the original BI total scores (r=0.78-0.86). The correlation coefficient of the SF-BI total scores and the original BI total scores was analyzed to be very high (r=0.95). The Cronbach's α coefficient of the SF-BI total scores was 0.87; and remained 0.82 to 0.86, which is an acceptable level, even when the individual items were deleted .

It was found that the SF-BI individual items were significantly relevant to the PASS total scores (r=0.75-0.78) and the FMA total scores (r=0.72-0.77) respectively, and the SF-BI total scores were significantly correlated to the PASS total scores (r=0.81) and the FMA total scores (r=0.76), respectively.

DISCUSSION

This study was aimed at identifying the SF-BI's test-retest reliability, absolute reliability (the SEM and MDC) and the associations with stroke-specific impairments. When it comes to the agreement degrees of the SF-BI individual items in this study, stair climbing showed the most excellent degree with 0.75, followed by transfers (0.67), ambulation (0.62), bathing (0.61) and toilet use (0.61). The agreement rates of the SF-BI individual items were 71% to 80%, which was a satisfactory level, and the agreement rate of the SF-BI total scores was confirmed to be high (ICC=0.91). These results are similar to those of the previous study by Hsueh et al.¹), where the weighted kappa coefficients of the original BI individual items were 0.53 to 0.94 (Median=0.72), which showed satisfactory or excellent agreement rates, but the agreement rate of "bathing" was 0.53 which was satisfactory and similar to the result of this study. In the present study, "bathing" and "toilet use" showed relatively low agreement rates, and this is due to the difference in the evaluation methods: by the subjects' self reporting rather than therapists' direct observation of their performance. Therefore, the evaluation method needs to be standardized and modulated for the two items¹). The SF-BI total scores of this study showed an ICC of 0.91 (0.86–0.95), which was consistent with 0.90 in the study of Hobart et al⁸). However, the result was in contrast to that of the study of Hsueh et al.¹ that presented ICCs of 0.55 at admission and 0.74 at discharge. In the case of the SF-BI, the floor effect is reduced from 46.6% at admission to 13.6% at discharge. Therefore, the SF-BI has a limitation in assessing the ADL of acute stroke patients. Nevertheless, it has been reported that it is appropriate to evaluate the independency levels of stroke patients as it has a response rate of 1.2,

which is similar to that of the original BI and the FIM-motor subscales¹). The ICC indicates the consistency of test scores in repeated measurements, but should be investigated using absolute reliability indices since the discrepancies in tests and measurement errors are unknown¹⁴). The SEM of the SF-BI in this study was 2.83, which was <10% of the average score of 41.66 and consistent with the SEM of 2.2 of the SF-BI in the study of Hobart et al⁸). However, the MDC of the SF-BI has never been reported. According to previous studies, the SEM and MDC of the original BI were 1.45 and 4.02, respectively²), similar to the SEM and MDC of the SF-BI of this study. The MDC of the SF-BI of this study was at an acceptable level, 7.84 points which is <20% of the maximum achievable score of 54. The score of 7.84 points of the subjects were maintained at the 95% confidence interval, which means that it is not a measurement error due to random variation and that there will be functional changes (improvement of independency levels) in each individual in the future. A highly reliable test should have a high ICC value and a low MDC value¹⁵), which was proved in this study. Thus, it was also confirmed that the SF-BI is a reliable assessment instrument for detecting and observing functional changes in patients over time in clinical settings. In this study, it was found that the individual items and total scores of the SF-BI are highly positively correlated (r=0.83-0.92) and that the SF-BI individual items and the total scores of the original BI are also highly correlated (r=0.78-0.86). The results of the study by Hobart et al.⁸⁾ were as follows: transfers (r=0.83), bathing (r=0.57), toilet use (r=0.83), stair climbing (r=0.68) and ambulation (r=0.77). The present study showed differences in bathing (r=0.83) and stair climbing (r=0.86). This is because the subjects of this study were stroke patients while the study of Hobart et al.⁸⁾ was conducted on patients with various diseases such as multiple sclerosis (45.6%), stroke (14.2%), spinal cord injuries (16.5%) and others (23.7%). In general, the application of one assessment instrument to patients with various diseases is inadequate for the functional assessment for patients and there is limitation in the estimation of recovery or response rates¹⁹. In this study, the concurrent validity (r=0.95) between the SF-BI and original BI total scores and the Cronbach's α coefficient (0.87) of the SF-BI were similar to those reported in the study of Hobart et al.⁸⁾ (r=0.96, 0.88). However, the results were quite different from those of the study of Hsuch et al.¹), which reported the concurrent validities of r=0.74 at admission and r=0.94 at discharge, and the Cronbach's α coefficients of 0.71 at admission and 0.73 at discharge, respectively. The average illness duration of the subjects in this study was 17.42 months and the 24 chronic stroke patients who had showed no natural recovery were selected as the subjects, while the study of Hsueh et al.¹) was conducted on 118 patients with acute stroke among which were included patients with a median FIM-motor score of 28 points (the maximum score: 91 points) and thus with severe impairments. In this case, selective biases and systematic errors that can affect the validity and reliability among variables due to the floor effect may be caused. In this study, the item internal consistency of the SF-BI total scores was considerably reliable, and even when individual items were deleted, it remained within a range of 0.82 to 0.86, which is at a high level. This means that the characteristics of the ADL of subjects are well reflected in the individual SF-BI items in terms of assessment of their ADL levels, which suggests that the individual items in the SF-BI are closely related to each other. The BI is an instrument to assess comprehensive ADL levels and it is known that the BI total scores at 14, 30, 90, and 180 days after the onset of stroke are significantly correlated with the FMA total scores (r=0.78-0.81)¹. However, the SF-BI is composed of the two items of toilet use and bathing (40%) derived from the original BI's seven self-management items and the three items of transfers, stair climbing and ambulation (60%) derived from the original BI's mobility items, which means that it is focused on the assessment of the ability to move. The correlations of the SF-BI with the PASS and the FMA have not been known yet. In this study, the individual items and total scores of the SF-BI were highly correlated with the PASS total scores (r=0.75-0.81). In a study on the correlation between the PASS and the FIM, which is the most similar to the BI, the PASS was reported to be associated with the FIMmotor items (r=0.82), ambulation (r=0.73) and the FIM total scores (r=0.73)¹⁶⁾. The PASS includes the items to assess standing on paralyzed and non-paralyzed sides (weight load, movement and balance) and picking objects from the floor in a standing position. It consists of tasks that can directly affect the mobility of stroke patients. Therefore, the PASS is considered to be composed of essential items to assess postural control and reactions prerequisite and necessary for the activities of daily living, so the two variables can be seen to be significantly correlated. The FMA reflects the selective separate movements and coordination of upper and lower extremities, the combined flexion and extension of the two extremities needed to perform activities of daily living²⁰). The FMA is mainly focused on the performance of upper extremity functions; however, the FMA total scores are significantly relevant to the BI total scores (r=0.75), personal hygiene (r=0.89), transfers (r=0.76), feeding (r=0.72) and dressing (R=0.76)²⁰. Regarding these characteristics, the individual items and total scores of the SF-BI in this study were found to be significantly related to the FMA total scores (=0.73-0.77), which is similar to the results of previous studies. Therefore, the SF-BI can be said to be suitable for the selective evaluation of ADL as it has high reliability and validity. The SF-BI has its advantages: allowing to easily collect data in a laboratory study and requiring only the least amount of time in managing data and interpreting results⁸). However, the reduction of evaluation items can cause floor effect and give limits to discrimination in uniformly evaluating ADL for all subjects with stroke. The limitations of this study are as follows: first, the SF-BI, which is focused on the items for assessing lower extremity motor functions and self-management, is inadequate to the patients with stroke who have severe disorders because there is floor effect; second, relationship between the SF-BI and upper extremity motor functions was not analyzed; and third, the results of this study cannot be generalized because this study's sample size is small, the average age of subjects is 60, and the FMA-L/E is 23.71 (23 to 28: moderate disorder) out of the maximum score, 3418), which suggests that this study is composed of relatively active subjects. In that sense, further studies are needed to check if the SF-BI can adequately evaluate ADL according to the degree of impairments in subjects with stroke and the duration of illness and to identify problems and limitations in the Rasch analysis and application of evaluation. The SF-BI was confirmed to be a useful evaluation instrument in that its test-retest agreement rate, absolute reliability, item internal consistency, and validity were high. Therefore, the SF-BI can be easily used in clinical practice, and both clinicians and researchers can use the selective ADL functions of patients with stroke and utilize them as useful information.

Funding

This paper was supported by research funds provided from the Howon University.

Conflict of interest

None.

REFERENCES

- Hsueh IP, Lin JH, Jeng JS, et al.: Comparison of the psychometric characteristics of the functional independence measure, 5 item Barthel index, and 10 item Barthel index in patients with stroke. J Neurol Neurosurg Psychiatry, 2002, 73: 188–190. [Medline] [CrossRef]
- Hsieh YW, Wang CH, Wu SC, et al.: Establishing the minimal clinically important difference of the Barthel Index in stroke patients. Neurorehabil Neural Repair, 2007, 21: 233–238. [Medline] [CrossRef]
- Hsueh IP, Hsu MJ, Sheu CF, et al.: Psychometric comparisons of 2 versions of the Fugl-Meyer motor scale and 2 versions of the stroke rehabilitation assessment of movement. Neurorehabil Neural Repair, 2008, 22: 737–744. [Medline] [CrossRef]
- Liaw LJ, Hsieh CL, Lo SK, et al.: The relative and absolute reliability of two balance performance measures in chronic stroke patients. Disabil Rehabil, 2008, 30: 656–661. [Medline] [CrossRef]
- 5) Hashidate H, Shiomi T, Sasamoto N: Effects of 3-month combined functional training at an adult day-care facility on lower extremity muscle strength and gait performance in community-dwelling people with chronic hemiplegia. J Phys Ther Sci, 2011, 23: 607–611. [CrossRef]
- Haigh R, Tennant A, Biering-Sørensen F, et al.: The use of outcome measures in physical medicine and rehabilitation within Europe. J Rehabil Med, 2001, 33: 273–278. [Medline] [CrossRef]
- 7) Cohen ME, Marino RJ: The tools of disability outcomes research functional status measures. Arch Phys Med Rehabil, 2000, 81: S21–S29. [Medline] [Cross-Ref]
- Hobart JC, Lamping DL, Freeman JA, et al.: Evidence-based measurement: which disability scale for neurologic rehabilitation? Neurology, 2001, 57: 639–644. [Medline] [CrossRef]
- 9) Chien CW, Lin JH, Wang CH, et al.: Developing a short form of the postural assessment scale for people with stroke. Neurorehabil Neural Repair, 2007, 21: 81–90. [Medline] [CrossRef]
- Liaw LJ, Hsieh CL, Hsu MJ, et al.: Test-retest reproducibility of two short-form balance measures used in individuals with stroke. Int J Rehabil Res, 2012, 35: 256–262. [Medline]
- Chen KL, Chen CT, Chou YT, et al.: Is the long form of the Fugl-Meyer motor scale more responsive than the short form in patients with stroke? Arch Phys Med Rehabil, 2014, 95: 941–949. [Medline] [CrossRef]
- Nanayakkara C, Lekamwasam S: Validity of the 5-item Barthel Index in the assessment of physical dependence in the elderly. Asian J Gerontol Geriatr, 2015, 10: 79–82.
- Hobart JC, Cano SJ, Thompson AJ: Effect sizes can be misleading: is it time to change the way we measure change? J Neurol Neurosurg Psychiatry, 2010, 81: 1044–1048. [Medline] [CrossRef]
- 14) Flansbjer UB, Blom J, Brogårdh C: The reproducibility of Berg Balance Scale and the Single-leg Stance in chronic stroke and the relationship between the two tests. PM R, 2012, 4: 165–170. [Medline] [CrossRef]
- 15) Chen HM, Hsieh CL, Liaw LJ, et al.: The test-retest reliability of 2 mobility performance tests in patients with chronic stroke. Neurorehabil Neural Repair, 2007, 21: 347–352. [Medline] [CrossRef]
- 16) Mitsutake T, Chuda Y, Oka S, et al.: The control of postural stability during standing is decreased in stroke patients during active head rotation. J Phys Ther Sci, 2014, 26: 1799–1801. [Medline] [CrossRef]
- 17) Mao HF, Hsueh IP, Tang PF, et al.: Analysis and comparison of the psychometric properties of three balance measures for stroke patients. Stroke, 2002, 33: 1022–1027. [Medline] [CrossRef]
- Fugl-Meyer AR, Jääskö L, Leyman I, et al.: The post-stroke hemiplegic patient. 1. A method for evaluation of physical performance. Scand J Rehabil Med, 1975, 7: 13–31. [Medline]
- Fisher AG, Short-DeGraff M: Improving functional assessment in occupational therapy: recommendations and philosophy for change. Am J Occup Ther, 1993, 47: 199–201. [Medline] [CrossRef]
- 20) Fugl-Meyer AR: Post-stroke hemiplegia assessment of physical properties. Scand J Rehabil Med Suppl, 1980, 7: 85–93. [Medline]