

## **Critical Evaluation of the Berg Balance Scale for Patients with Parkinson's disease**

**Zhiwei Yang\*<sup>2</sup>, Hui Wang<sup>1</sup>, Sheng Wang<sup>2</sup>**

<sup>1</sup>University of Nottingham, UK

<sup>2</sup>The affiliated Suzhou Science & Technology Town Hospital of Nanjing Medical University, Suzhou, China.

**\*Corresponding author:** Zhiwei Yang, the affiliated Suzhou Science & Technology Town Hospital of Nanjing Medical University, Suzhou, China.

Received: November 20, 2020

Published: December 15, 2020

Parkinson's disease (PD) is associated with the degeneration of the central nervous system. PD can lead to motor dysfunction which mainly manifests as bradykinesia, abnormal gait and postural instability (McNeely et al. 2012) [1]. Several sources of evidence report that motor dysfunction can directly result in the impairment of balance ability (Kerr et al. 2010; Lhomme et al. 2012) [2,3]. Due to balance dysfunction, individuals with PD commonly experience falls, which can lead to serious injuries. Therefore, it is important for therapists to have a measurement tool which can evaluate the balance ability of individuals with PD.

The Berg Balance Scale (BBS) is a tool which can evaluate balance for individuals with various disabilities (Downs 2015) [4]. The BBS contains 14 balance tasks and each one is scored on a Likert scale from 0-4. The total score can range from 0 to 56. A higher score stands for better balance function. If the result of the BBS is below 40, it indicates that the balance function of the individual is poor and the individual may have a higher risk of falling (Berg et al. 1992) [5]. The 14 tasks of the BBS include sitting and standing, standing unsupported, sitting unsupported, standing and sitting, transferring, standing with eyes closed, standing with foot together, reaching forward with outstretched arm, picking up an item from the floor, turning to look behind, turning 360°, placing alternate foot on a step, standing with one foot in front, and standing on one foot (Badura and Pietka 2016) [6]. Previous studies of the BBS have identified that it has been widely applied to assess the impairment of balance function caused by PD (Franchignoni and Vellozo 2005) [7]. However, the psychometric properties (validity, reliability, responsiveness and feasibility) of the BBS in assessing balance function of an individual with PD remain unclear.

A valid reliable measurement not only plays an important role in assessing whether the research results are acceptable but also demonstrating the effectiveness of treatment in clinical practice (Hicks 2009) [8]. This assignment will discuss these psychometric properties of the BBS for evaluating balance function of individuals with PD, identify the need for Outcome Measures (OM) in practice and conclude as to the relevance of the BBS in clinical practice with PD.

The term "berg balance scale", "Parkinson's disease", "valid-

ity", "reliability", "responsiveness" and "feasibility" will be used in a computer-based search strategy from the following databases: Springer Link, Ovid and Google scholar. The assignment will start by defining the terms utilized to describe an OM and use the SURE checklist to critique studies.

### **Validity**

The term "validity" is defined by Hicks (2009) [8] as the degree to which a measurement tests what it attempts to test. Generally, there are four main types of validity which are face validity, construct validity, content validity and predictive validity. Face validity can be defined as the test seemingly measuring what it should measure. The function of construct validity is to reflect whether a measurement can test a theoretical concept or construct. Convergent validity is a branch of construct validity, which can be defined as whether two different tests can measure a same theoretical concept or construct. The definition of content validity can be described as the correspondence between test tasks on a scale and the symptoms of a scenario being assessed. Predictive validity refers to the degree to which an instrument can predict future performance (Hicks 2009) [8].

Berg et al. (1989) [9] test the content validity of BBS in his original study which recruited 38 elderly Canadian patients (28% of them with PD) and 32 health-related staff from different healthy setting in Canada. 38 items were initially generated from three dimensions of balance (static balance, self-dynamic balance and dynamic balance) through discussing the relevance of all tasks of the BBS. Inappropriate movements were deleted due to difficulty to complete, such as "Turning head while sitting on pillow with legs crossed". Duplicate, unimportant and irrelevant items were deleted. Because of these rigorous screening methods, the content validity of the BBS should be excellent. However, although the background of health-related staff is different and the patients are elderly people with different balance-related diseases, they all come from two cities in Canada. Cultural bias may lead to a low generalization of research results (Wan 2001) [10].

70 stroke patients from acute care hospitals were recruited and tested using the BBS, Barthel Index (BI) and balance subscale of Fugl-Meyer (FM-B) (these were viewed as the gold standard at this time) in order to measure their construct validity (Berg

et al. 1992). All the samples were evaluated by these three instruments at baseline, 4th, 6th and 12th week. The excellent correlation between the BBS and the BI ( $r=0.87-0.93$ ), and good correlation between the BBS and the FM-B ( $r=0.84-0.94$ ) could be observed in this study. The correlation between the BBS and BI has no significant difference from the correlation between the BBS and FM-B. Because the BI is for assessing the daily activity and the FM-B is for assessing the balance function, this may imply a positive correlation between balance function and daily activity function.

There are currently no literature reviews addressing the construct validity and content validity of the BBS. Although the original research on the BBS investigates its validity, very few subjects of the research have PD. Most of the subjects are stroke patients and elderly people. Nonetheless, data from several investigations establish that impaired function of balance is a common problem among stroke patients, PD patients and elderly people (Byl et al. 2015; Beghi et al. 2018) [11,12]. Hence, the outcomes of content and construct validity from the BBS on measuring stroke and elderly subjects is partly transferable to PD subjects.

The high convergent validity (using the Pearson correlation coefficient) between the BBS and Mini-BESTest has been reported by Godi et al. (2013) [13] at baseline ( $r=0.85$ ). After the therapy programme, moderate correlation ( $r=0.58$ ) between the BBS and Mini-BESTest could be found. Godi et al. (2013) [13] suggest that the change of correlation between the BBS and Mini-BESTest during baseline and after therapy may result from the generally higher scores in the BBS after treatment. This phenomenon is regarded as the ceiling effect of the BBS (Schlenstedt et al. 2015) [14]. However, only one-quarter of the total sample size is PD patients, so the results of this study do not transfer well to all PD patients. Nonetheless, King et al. (2012) [15] recruited a total of 97 PD patients and indicated the excellent correlation between the BBS and Mini-BESTest ( $r=0.79$ ). Also, Bergström et al. (2012) [16] recruited a total of 9 PD patients and indicated a high correlation between the BBS and Mini-BESTest ( $r=0.94$ ). The above studies demonstrate the high convergent validity between the BBS and Mini-BESTest on measuring PD population. Therefore, both would appear valid to measure validity in the PD population.

Park and Lee (2016) [17] attempt to assess the predictive validity of the BBS on managing patients with balance issues in their systemic review. This systematic review finds that the predictive validity of the BBS is moderate when the BBS is used to predict fall risk among populations with different neuromuscular diseases. However, the population of this systematic review was not exclusively PD patients, stroke patients were also included. Although nearly one-third of the sample ( $n=1690$ ) were PD patients ( $n=513$ ). The presence of patients with other diseases affects the ability to conclude on validity of the BBS for predicting balance function of PD patients. Nonetheless, the population of this study has common symptoms of impaired balance. According to the high predictive validity of BBS from the outcome of this systematic review, the BBS can effectively predict the balance function of neuromuscular patients with balanced functional impairment.

There is a lack of research testing the validity of the BBS in PD

patients, although the construct of the BBS is proven to measure balance and be predictive of falls. There appears sufficient evidence by Godi et al to support its use with PD patients.

### Reliability

Reliability refers to whether an instrument can be administered repeatedly and achieve the same result. (i.e. is it consistent). If the measurement recorded is similar when measured by different individuals (inter-rater) and the same person on different occasions (test-retest), this indicates strong reliability of the tool or OM. To determine whether the measure is reliable or not, recordings are tested using correlation statistics (e.g. intraclass correlation (ICC) for parametric data and Spearman's or Pearson's test for nonparametric data) (Hicks 2009) [9]. If the correlation coefficient is higher than .75, it demonstrates high reliability otherwise .50 to .75 or 0 to .50 demonstrates medium or low reliability respectively (Portney and Watkins 2013) [19]. The BBS produces ordinal data within each section, however, the total score becomes interval ratio and therefore when critiquing the evidence, it is important to consider the score being evaluated.

The inter-rater and test-retest reliability of the BBS was assessed by Schlenstedt et al. (2015) [14]. For inter-rater reliability, 15 PD patients and three assessors were recruited. These three assessors were trained and had experience implementing the BBS. Measurements were carried out by them who took the measures once. For test-retest reliability, 17 participants were measured. One assessor evaluated each patient by taking one measure and a repeat measure  $3\pm 1$  days later. The repeat was taken by the same assessor at the same time of day for each patient. This study mainly tested the reliability of the total score of 14 items of the BBS, the total score is interval level data and therefore statistical testing can be analysed by parametric tests. Therefore, this study uses correct statistical testing i.e. (ICC). The ICC for the inter-rater reliability and test-retest reliability of the BBS was very high as the ICC was  $r=0.98$  and  $r=0.95$ , respectively. The recruitment of three experienced and trained assessors for testing the inter-rater reliability of the BBS can guarantee the accuracy of study outcomes and minimize human errors (Vincent 2006) [19]. However, this may question whether an inexperienced tester would also have good reliability. Also, all assessors scored the BBS independently and their results are blinded to each other. This blinding can avoid the bias in the study design (Pannucci and Wilkins 2010) [20]. Therefore, the outcome of inter-rater reliability of the BBS in this study appears to be convincing.

However, the outcome of test-retest reliability of the BBS in the study of Schlenstedt et al. (2015) [14] appears questionable. Although the researchers considered the impact of evaluation time (the same time of day) and antiparkinsonian medicine dosage and ensured there was no difference in the dosage of antiparkinsonian medication between the first assessment and the second assessment, the interval between two assessments is very short ( $3\pm 1$  days). The impact of the memory of patient and the assessor over a short period of time may result in a false high test-retest reliability of the BBS (Schatz and Maerlender 2013) [21]. Since PD is a chronic progressive disease, the balance function of the patient may not change significantly within three weeks (Lhommée et al. 2012) [3]. It is recommended that future studies extend the interval between

assessments.

32 patients (only eight PD) were recruited by Godi et al. (2013) [13] to assess the inter-rater reliability and test-retest reliability of the total score of 14 items of the BBS. Its use of ICC was suited to this study, which is a strength. This study indicates the excellent inter-rater ( $r=0.97$ ) and test-retest reliability ( $r=0.92$ ) of the BBS. Compared to Schlenstedt et al. (2015) [14], whereby the procedures for assessing the reliability is almost the same. However, the inclusion of only eight PD patients may mean that the sample is not representative of the whole Parkinson population. Hence, the outcome of this study can only be accepted with caution.

Berg et al. (1989) [9] demonstrate excellent inter-rater reliability (total score:  $r=0.98$ ; each item:  $r=0.71-0.99$ ) and test-retest reliability (total score:  $r=0.99$ ; each item:  $r=0.71-0.99$ ) of the BBS in the original study. All samples came from a geriatric ward and their balance function was affected by a variety of diseases. Combined with the above two studies, the good inter-rater and test-retest reliability of the BBS in assessing patients with PD and other balance disorders can be demonstrated.

### Internal Consistency

Internal consistency (IC) is a kind of reliability in which all components of a measure that propose to measure the same things actually produce similar scores. Cronbach's alpha is usually used for calculating internal consistency. The range of Cronbach's alpha is between 0.00 to 1.00. The value of Cronbach's alpha is required to be high and greater than 0.70 is satisfactory in demonstrating good IC (Connelly 2011) [22]. Halsaa et al. (2007) [23] report the IC of the Norwegian version of the BBS ( $r=.87$ ) is higher than .70 which could show relatively strong IC. The original study of the BBS's IC is very high ( $r=.96$ ), with a value  $>.90$  which has been described as a requirement for inclusion clinically use (Berg et al. 1989). The high IC shows that all items of the BBS have an excellent function for evaluating balance.

### Responsiveness and Feasibility

Responsiveness of an OM reflects the ability of a measurement to demonstrate and examine change over time (Portney and Watkins 2013) [18]. 26 PD patients were recruited by Lim et al. (2005) [24] for examining the responsiveness and feasibility of the BBS. Because there is very little literature on the study of responsiveness of the BBS, this study determines to apply the Smallest Detectable Difference (SDD) which is a kind of statistical data calculated by standard error (Beckerman et al. 2001; Beghi, E. et al. 2018; Bergström, M. et al. 2012; Morris, M. E. et al. 1994) [25-28]. However, the SDD from this study does not show the extent to which the variable changes of BBS. Moreover, the results from the SDD were subjectively applied for determining the responsiveness of the BBS. Therefore, the responsiveness of the BBS remains unclear. This study also focuses on the feasibility of the BBS which is determined by the length of one assessment time in a different setting. Assessment time in the clinical setting for one patient is 20-25 minutes and 20-30 minutes in the home environment, indicating that the BBS assessment is feasible in clinical and home circumstances. In addition, the equipment for assessing the BBS is just a chair, a watch, a ruler and a step. The cost of administering the BBS is not expensive. Therefore, the BBS is

a feasible tool.

Taken together, importance of using valid reliable OM in physio is to prove the effectiveness of treatment. The BBS is indeed a reliable and valid instrument that can be used to assess PD patients with impaired balance. One assessment time of the BBS in the clinical and home environment is short and stable, and its cost is low. However, due to its lack of literature on investigating the responsiveness of the BBS, this remains unclear. Also, some issues from previous studies include complicated study populations (not all PD patients) and methodological flaws (short interval between assessments). Attempts to correct the above issues would help to establish a greater degree of certainty when assessing balance in PD patients using the BBS. Nonetheless, the BBS remains an appropriate assessment of balance for the clinical PD setting.

### References

1. McNeely ME, et al. Medication improves balance and complex gait performance in Parkinson disease. *Gait & Posture* 2012; 36(1): pp. 144-148. doi: <https://doi.org/10.1016/j.gaitpost.2012.02.009>
2. Kerr GK, et al. Predictors of future falls in Parkinson disease. *Neurology* 2010; 75(2): pp. 116-124.
3. Lhommée E, et al. Subthalamic stimulation in Parkinson's disease: restoring the balance of motivated behaviours. *Brain* 2012; 135(5), pp. 1463-1477.
4. Downs S. The Berg Balance Scale. *Journal of Physiotherapy* 2015; 61(1), p. 46. doi: <https://doi.org/10.1016/j.jphys.2014.10.002>
5. Berg KO, et al. 1992. Measuring balance in the elderly: validation of an instrument. *Canadian journal of public health= Revue canadienne de sante publique* 1992; 83, pp. S7-11.
6. Badura P, Pietka E. Automatic Berg Balance Scale assessment system based on accelerometric signals. *Biomedical Signal Processing and Control*, 2017; 24(Supplement C): pp. 114-119. doi: <https://doi.org/10.1016/j.bspc.2015.10.005>
7. Franchignoni F, Vellozo CA. Use of the Berg Balance Scale in Rehabilitation Evaluation of Patients With Parkinson's Disease. *Archives of Physical Medicine and Rehabilitation* 2005; 86(11), pp. 2225- 2226. doi: <https://doi.org/10.1016/j.apmr.2005.09.006>
8. Hicks C. Research methods for clinical therapists: applied project design and analysis. 5th ed. ed. Edinburgh: Edinburgh: Churchill Livingstone. 2009.
9. Berg K, et al. Measuring balance in the elderly: preliminary development of an instrument. *Physiotherapy Canada* 1989; 41(6), pp. 304-311.
10. Wan MW. Ethnic culture, distress and clinical measurement: A CORE outcome comparison between the British Chinese and white Europeans. *Journal of Mental Health* 2001; 10(3): pp. 301-315. doi: 10.1080/09638230124768
11. Byl N, et al. Clinical impact of gait training enhanced with visual kinematic biofeedback: Patients with Parkinson's disease and patients stable post stroke. *Neuropsychologia* 2015; 79, pp. 332-343. doi: <https://doi.org/10.1016/j.neuropsychologia.2015.04.020>
12. Beghi E, et al. Prediction of Falls in Subjects Suffering From Parkinson Disease, Multiple Sclerosis, and Stroke. *Archives of Physical Medicine and Rehabilitation* 2018; 99(4): pp. 641-651. doi: <https://doi.org/10.1016/j.apmr.2017.10.009>
13. Godi M, et al. Comparison of Reliability, Validity, and Responsiveness of the Mini- BESTest and Berg Balance Scale in Patients With Balance Disorders. *Physical Therapy* 2013; 93(2), pp. 158-167. doi: 10.2522/pti.20120171
14. Schlenstedt C, et al. Comparing the Fullerton Advanced Balance Scale With the Mini-BESTest and Berg Balance Scale to Assess Postural Control in Patients With Parkinson Disease. *Archives of Physical Medicine and Rehabilitation* 2015; 96(2), pp. 218-225. doi: <https://doi.org/10.1016/j.apmr.2014.09.002>
15. King LA, et al. Comparing the Mini-BESTest with the Berg Balance Scale to Evaluate Balance Disorders in Parkinson's Disease. *Parkinson's Disease* 2012; p. 375419. doi: 10.1155/2012/375419

16. Bergström M, et al. Translation and validation of the Swedish version of the mini-BESTest in subjects with Parkinson's disease or stroke: A pilot study. *Physiotherapy Theory and Practice* 2012; 28(7), pp. 509-514. doi: 10.3109/09593985.2011.653707
17. Park SH, Lee YS. The Diagnostic Accuracy of the Berg Balance Scale in Predicting Falls. *Western Journal of Nursing Research* 2016; 39(11): pp. 1502-1525. doi: 10.1177/0193945916670894
18. Portney LG, Watkins MP. *Foundations of Clinical Research: Pearson New International Edition: Applications to Practice*. Pearson Education M.U.A. 2013.
19. Vincent C. *Patient safety*. Churchill Livingstone Edinburgh. 2006.
20. Pannucci CJ, Wilkins EG. Identifying and avoiding bias in research. *Plastic and reconstructive surgery* 2020; 126(2), p. 619.
21. Schatz P, Maerlender A. A Two-Factor Theory for Concussion Assessment Using ImPACT: Memory and Speed. *Archives of Clinical Neuropsychology* 2013; 28(8): pp. 791-797. doi: 10.1093/arclin/act077
22. Connelly LM. Cronbach's alpha. *Medsurg nursing* 2011; 20(1): pp. 45-47.
23. Halsaa KE, et al. Assessments of Interrater Reliability and Internal Consistency of the Norwegian Version of the Berg Balance Scale. *Archives of Physical Medicine and Rehabilitation* 2007; 88(1): pp. 94-98. doi: <https://doi.org/10.1016/j.apmr.2006.10.016>
24. Lim LIIK, et al. Measuring gait and gait-related activities in Parkinson's patients own home environment: a reliability, responsiveness and feasibility study. *Parkinsonism & Related Disorders* 2005; 11(1): pp. 19-24. doi: <https://doi.org/10.1016/j.parkreldis.2004.06.003>
25. Beckerman H, et al. Smallest real difference, a link between reproducibility and responsiveness. *Quality of Life Research* 2001; 10(7): pp. 571-578.
26. Beghi E, et al. Prediction of falls in Subjects Suffering from Parkinson Disease, Multiple Sclerosis, and Stroke. *Archives of Physical Medicine and Rehabilitation* 2018; 99(4): pp. 641-651. doi: <https://doi.org/10.1016/j.apmr.2017.10.009>
27. Bergström M, et al. Translation and validation of the Swedish version of the mini-BESTest in subjects with Parkinson's disease or stroke: A pilot study. *Physiotherapy Theory and Practice* 2012; 28(7): pp. 509-514. doi: 10.3109/09593985.2011.653707
28. Morris ME, et al. Ability to modulate walking cadence remains intact in Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry* 1994; 57(12): pp. 1532-1534.