



Clinical assessment of balance using BBS and SARAbal in cerebellar ataxia: Synthesis of findings of a psychometric property analysis

Stanley John Winser^{1,*}, Catherine M Smith², Leigh A Hale², Leica S Claydon³ and Susan L Whitney^{4,5}

¹*Department of Rehabilitation Sciences, Hong Kong Polytechnic University, Hong Kong*

²*School of Physiotherapy, University of Otago, New Zealand*

³*Department of Allied and Public Health, Anglia Ruskin University, UK*

⁴*School of Health and Rehabilitation Sciences, Department of Physical Therapy University of Pittsburgh, USA*

⁵*Rehabilitation Research Chair at King Saud University, Saudi Arabia*

**stanley.j.winsler@polyu.edu.hk*

Received 24 August 2016; Accepted 19 July 2017; Published 4 April 2018

Background: In the previous psychometric analysis paper in our series for identifying the core set of balance measures for the assessment of balance, we recommended the Berg Balance Scale (BBS) and balance sub-components of the Scale for the assessment and rating of ataxia (SARAbal) as psychometrically sound measures of balance for people with cerebellar ataxia (CA) secondary to multiple sclerosis.

Objective: The present study further examined the suitability of BBS and SARAbal for the assessment of balance in CA with regard to psychometric property strength, appropriateness, interpretability, precision, acceptability and feasibility.

Methods: Criteria to fulfill each factor was defined according to the framework of Fitzpatrick *et al.* (1998). Based on the findings of our previous psychometric analysis, each criterion was further analyzed.

*Corresponding author.

Results: The psychometric analysis reported good reliability and validity estimates for the BBS and SARAbal recommending them as psychometrically sound measures; they fulfilled both criteria for appropriateness and interpretability, the measures showed evidence for precision and acceptability, and they were found to be feasible in terms of the time and cost involved for the balance assessment.

Conclusion: We have provided evidence for the use of the BBS and SARAbal for the assessment of balance among people with CA.

Keywords: Balance; cerebellar ataxia; multiple sclerosis; psychometric analysis.

Introduction

Poor balance and gait difficulties are hallmarks of health conditions that result in cerebellar ataxia (CA).¹ Assessment of balance and gait in CA is challenging as there are no standardized measures of balance available. Previously, a series of studies by our research group recommended a set of core measures. A systematic review² and a Delphi survey³ reported the Berg Balance scale (BBS), the Timed Up and Go (TUG) test, posture and gait sub-component of the International Co-operative Ataxia Rating Scale (PG-ICARS) and the gait, stance and sit sub-components of the Scale for the Assessment and Rating of Ataxia (SARAbal) as appropriate measures of balance in CA.^{2,3} Further, a psychometric property analysis was done to estimate constructs of reliability and validity of these four measures among people with CA secondary to multiple sclerosis in New Zealand and the United States of America. The study aimed at proposing the best outcome measures based on the findings of the psychometric analysis. The BBS and SARAbal were recommended as the optimal measures of balance in people with CA secondary to multiple sclerosis.⁴

Fitzpatrick *et al.* reported eight factors to be addressed while selecting an outcome measure for clinical trials.⁵ In the process of choosing a standardized set of measures for balance in people with CA, these eight factors were considered. The present study therefore aimed to examine the psychometric properties, appropriateness, interpretability, precision, acceptability and feasibility of the BBS and SARAbal for people with CA based on the findings of the psychometric property analysis done by our research team earlier.⁴

Methods

This paper examined eight factors in light with Fitzpatrick's framework of evaluating a suitable

outcome measure for clinical trials and clinical practice.⁵ The findings of the present study were based on the outcomes of a psychometric property analysis of four outcome measures of balance tested in people with CA secondary to multiple sclerosis.⁴ For the present study, we grouped reliability, validity and responsiveness as psychometric properties.⁴ The factors are analyzed and their definitions are listed in [Table 1](#).

Each factor was analyzed based on the set criteria outlined as follows.

Key findings of the psychometric analysis of the BBS and SARAbal were summarized to report the reliability and validity of these measures in people with CA. The detailed methodology and results of this psychometric property analyses are published elsewhere.^{4,6} The other reported factors including appropriateness, interpretability, precision, acceptability and feasibility we based on the experience gained during the data collection and interpretation of results of our previous psychometric analysis study. To summarize, 60 participants aged 18–65 years with CA secondary to multiple sclerosis were recruited. Data were collected at four outpatient units in New Zealand and the United States of America. All included participants underwent balance assessment using the BBS, TUG, SARAbal and PG-ICARS. The participants were assessed on a single occasion and during the assessment, a video recording was done. The video recording was later used to estimate the intra-rater and inter-rater reliabilities. The Barthel Index, the Expanded Disability Status Scale (EDSS), the full scales of the ICARS and the SARA were also assessed and disease duration was recorded. The EDSS was completed by a neurologist. To investigate the intra-rater and inter-rater reliabilities, a repeat assessment was performed by the same physiotherapist (intra-rater) or a second physiotherapist (inter-rater) from the video recording.

Table 1. Descriptors of the factors analyzed.

Factor	Descriptor
Psychometric properties	Common term that includes reliability, validity and responsiveness of the outcome measures
Appropriateness	Described as how suitable the contents of the instrument are for use in people with CA ⁵
Interpretability	Indicates how meaningful are the scores obtained from the outcome measures ⁵
Precision	Defined as the accuracy of the instrument in categorizing sub-groups and distribution of numerical value
Acceptability	Defined as the level to which the outcome measure is tolerable for its use in people with CA ⁵
Feasibility	Described as the ease of use of the outcome measure in terms of administering it, and the associated financial cost ⁵

In this study, appropriateness was analyzed based on two criteria: (i) to judge whether the contents of the outcome measure suit the target population and (ii) if the recommended set of outcome measures has a combination of a condition-specific tool and a generic tool for the assessment of balance. Interpretability was analyzed based on two criteria: (i) to determine how meaningful the obtained scores were, using the BBS and the SARAbal and (ii) to determine if the outcome measures have established normative data. Precision was analyzed based on two criteria: (i) to determine if the instrument is able to discriminate between two known sub-groups within the collected samples and (ii) the accuracy of distribution of numerical values assessed using Rasch analysis or estimation of unidimensionality of the testing items using factorial analysis. Acceptability was determined by estimating the response rate of the participants to the items of the outcome measures. In general, the lesser the missing items, the better the acceptability.⁵ Feasibility was assessed by observing the ease of use, cost involved for the assessment, time taken to complete and training required for the assessor to complete the balance assessment using the two outcome measures. The criteria were organized into a tabular column and the reviewer marked either “yes” if the criteria were met or “no” if the criteria were not met or “unclear” if the answer was ambiguous. Each of the criterion was independently reviewed by two authors (SW and CS) and discrepancies in findings

were discussed. A third reviewer (LC) was involved for unresolved discrepancies in the findings between the first two reviewers or if the reviewers marked “unclear” for the criteria.

Balance measures

The BBS is a performance-based measure of balance⁷ and has been reported to be the most commonly used balance tool by physiotherapists.⁸ The BBS is a five-point ordinal scale scored between 0 and 4 for each task and has 14 tasks in total. The highest total score a participant may obtain is 56. This measure is interpreted as better balance with higher scores. Normative scores for the BBS have been established among community dwelling older adults.⁹ This measure has good inter-rater (ICC = 0.96) and test retest (ICC = 0.94) reliabilities and low standard error of measurement (SEM).¹⁰ The BBS is found to have acceptable concurrent validity in assessing balance and poor in discriminating between fallers and non-fallers in people with multiple sclerosis.¹¹

The SARA is an ataxia severity rating measure.¹² It consists of eight items among which gait, sitting and the standing sub-components are related to balance. The full scale is scored out of 40. The three sub-components of balance are scored out of 18 (SARAbal). Scoring of the eight sub-components do not have equal weighting, with scores ranging between eight for the “gait” sub-component and four for the “heel-shin glide”. The higher the score obtained, the worse the condition. The SARA has high test re-test reliability (ICC = 0.90), inter-rater reliability (ICC = 0.97) and internal consistency ($\alpha = 0.93$).¹² Structural validity has been reported,¹³ satisfactory convergent validity when correlated with other ataxia rating scales¹² and adequate responsiveness has been demonstrated.¹⁴ The testing has been done and conducted with both genetic and acquired forms of cerebellar disorders.

Results

The review of criteria for each factor resulted in 100% agreement between the reviewers and therefore the third reviewer was not approached. The reliability and validity of the measures were found to be strong and the responsiveness was not estimated. A summary of the findings on the

psychometric properties of the BBS and the SARAbal are highlighted in Table 2. For appropriateness, the measures met both the criteria. With regards to interpretability, off the two required criteria, both the measures met the first criteria whereas the BBS met the second criteria and SARAbal did not. The first criteria for precision were met by both the measures however, the second criteria were not established as Rasch analysis and factor analysis were outside the scope of the psychometric analysis. Both the measures met the criteria for acceptability and in addition, they were found to be feasible.

Discussion

This study aimed at identifying the suitability of using the BBS and SARAbal for the clinical assessment of balance in people with CA. The framework of Fitzpatrick *et al.*⁵ was used to address eight independent factors for this recommendation. We have provided evidence for most of the factors and in addition, recommendations for future research for strengthening the present findings have been provided.

Psychometric properties of the measures of balance

The BBS and SARAbal reported good intra-rater, inter-rater reliabilities and internal consistency.⁴ The criterion validity was found to be good for both the measures ($\rho_S > 0.80$). The measures were correlated against disease duration, disease severity and functional independence to determine construct validity and correlation was moderate ($\rho_S > 0.55$). The measures were correlated against ataxia severity rating scales to estimate convergent validity which was found to be good. The study participants were sub-divided into assistive walking device users and non-users. The ability of the measures of balance to differentiate between users and non-users of assistive devices was studied to determine the discriminant validity. The balance scores showed a statistically significant difference between the scores of assistive device users and non-users showing evidence for discriminant validity. In summary, both the BBS and SARAbal have good reliability and acceptable validity for the assessment of balance among people with CA secondary to multiple sclerosis. The structural

validity and responsiveness of the measures of balance were not determined.

Appropriateness of the measures of balance

A straightforward method to determine if the contents of the outcome measure suit the target population is to obtain feedback from end users, the clinicians. The psychometric analysis involved testing four balance measures of which three were endorsed by experts through the Delphi survey done earlier by our research team.³ In the Delphi survey, neurologists and physiotherapists involved in research and clinical practice of CA were interviewed. They were asked to indicate the most appropriate measure of balance they might use to quantify balance deficits relating to CA. The internet-based survey went on for two rounds and the participants came to a consensus on the use of the BBS, TUG and SARAbal as the most appropriate choice of assessment tool. Two of the measures recommended as the core set were those endorsed by the clinical experts in the Delphi study providing evidence for appropriateness.³

Secondly, it is recommended that an appropriate set of patient outcome measures should have one condition-specific measure and a generic measure.⁵ A condition-specific measure identifies changes that are in close relation or “proximal” to the disease such as difficulty in performing tandem walking in CA and the generic measure identifies changes that are slightly less proximal or “distal” to the health condition,¹⁸ such as altered stepping secondary to coordination deficits in CA. Among the core set of measures, the SARAbal is condition-specific and the BBS is a generic measure of balance.²

Interpretability of the measures of balance

In order to identify a meaningful score, the most significant approach may be to relate the scores achieved to the minimal clinically important difference (MCID).⁵ The MCID is described as the smallest difference in the score following an intervention that the patient perceives as beneficial.¹⁹ Since the psychometric analysis did not involve a repeat assessment, where arguably a change in score could be expected, determining the MCID score was not possible. However, we established the

Table 2. Definition, accepted statistical analysis, interpretation and findings of the psychometric properties considered.

Psychometric property	Description	Statistical analysis	Interpretation	Results
<i>Reliability</i>				
Internal consistency	Defined as the degree of interrelatedness between the test items within each outcome measures considered. ¹⁵	Cronbach alpha	There are no universal guidelines for interpreting reliability, in general, higher the value towards 1, greater the reliability. We interpreted as follows: $\alpha > 0.80$: good, α between 0.5 and 0.79: moderate, $\alpha < 0.50$: poor	$\alpha = 0.94$ (BBS) $\alpha = 0.72$ (SARAbal)
Inter-rater reliability	Defined as the proportion of variation in the scores of the participant done by two different investigators. ¹⁶	Continuous scores: ICC Dichotomous/nominal/ordinal scores: kappa (κ) or weighted kappa	ICC > 0.80 : good, ICC between 0.5 and 0.79: moderate, $A < 0.50$: poor	ICC = 0.97 (BBS) ICC = 0.96 (SARAbal)
Intra-rater reliability	Defined as the proportion of variation in the scores of the participant done by the same investigator with an interval of 7–10 days. ¹⁵	Same as inter-rater reliability		ICC = 0.99 (BBS) ICC = 0.98 (SARAbal)
<i>Validity</i>				
Criterion validity	Defined as the degree to which the scores of the measure under investigation are an adequate reflection of a “gold standard”. ¹⁶	Spearman or Pearson correlation co-efficient. Since the outcome measures considered were ordinal, we used the Spearman correlation co-efficient (ρS). Since “gold standard” was not available for balance assessment, we correlated the measures of balance (BBS, TUG, PGICARS and SARAbal) against each other.	$\rho S > 0.80$: good, ρS between 0.5 and 0.79: moderate, $\rho S < 0.50$: poor	BBS versus TUG: -0.88 PGICARS: -0.80 SARAbal: -0.92 SARAbal versus BBS: -0.92 TUG: 0.72 PGICARS: 0.92
Hypothesis testing	Defined as the degree to which the scores of the measures under investigation are consistent with the hypotheses. ¹⁶ Convergent, divergent, external and construct validity are grouped under hypothesis testing.	Spearman correlation co-efficient (ρS).	Same as above	
Convergent validity	Indicates that two measures examining similar underlying phenomenon will provide similar results. For example, high correlation can be anticipated between the results of two outcome measures assessing balance.	Spearman correlation co-efficient (ρS). The measures of balance were correlated with two ataxia rating scales (ICARS and SARA)	Same as above	BBS versus ICARS: -0.76 SARA: -0.82 SARAbal versus ICARS: 0.79 SARA: 0.85

Table 2. (*Continued*)

Psychometric property	Description	Statistical analysis	Interpretation	Results
External validity	Defined as the degree to which the outcome measure under investigation correlates with other instruments or other constructs, for example ADL, disease severity or disease duration.	Spearman correlation co-efficient (ρS). The measures of balance were correlated with ADL status, disease duration and disease severity. ADL was assessed using BI and disease severity using the EDSS.	Same as above	BBS versus EDSS: -0.78 BI: 0.55 Disease duration: -0.61 SARAbal versus EDSS: 0.76 BI: -0.44 Disease duration: 0.58
Discriminant validity	Defined as the ability of the outcome measures to differentiate between two-known groups within the study population.	Group differences of scores between users and non-users of assistive walking devices were considered for establishing discriminant validity. We used Mann-Whitney U test.	Statistically significant difference ($p < 0.05$) in the scores between groups was considered evidence for discriminant validity.	Mean, SD and p value BBS ADU: 34.6 (11.8) ADNU: 52.19 (4.43) $p < 0.01$ SARAbal ADU: 7.0(2.8) ADNU: 1.71 (1.37)
Cut-off score, sensitivity and specificity for assistive walking device use	Sensitivity is an indication that the outcome measure is capable of identifying certain trait that is really present in the given population. Specificity is an indication that the outcome measure is capable of identifying the lack of certain trait that is really absent in the given population.	Receiver operating characteristics (ROC) curve was constructed to determine the cut-off score, sensitivity and specificity of the measures to predict the users of an assistive walking device. In addition, to determine and quantify which measure had a better predictive ability, the "Area Under the Curve" (AUC) was used.	The examiner makes a logical decision based on the needs for the cut-off score. In this case, the score needs to precisely identify an assistive device user more than identifying a non-user. Thereof, the sensitivity was kept high and constant at 90% and the corresponding cut-off score and the highest specificity at 90% of sensitivity were derived.	$p < 0.01$ BBS: cut-off < 44 out of 56 sensitivity 90% specificity 94% SARAbal: cut-off > 5 out of 18 sensitivity 90% specificity 100%
<i>Responsiveness</i>	Described as the ability of the outcome measure to detect changes over time. ¹⁶	Can be determined using different approaches. Some commonly adopted analysis include ROC (distribution-based approach) or relating the change of score to "Global Rating of Change" score (anchor-based approach). ¹⁷		Responsiveness was not estimated.

Notes: α — Cronbach's alpha, ICC — intra class correlation co-efficient, ρS — Spearman's Rho, BBS — Berg Balance Scale, SARAbal, gait, sit and stance sub-component of the SARA, PGICARS — Posture and gait sub-component of the ICARS, SARA — Scale for the Assessment and Rating of Ataxia, ICARS — International Co-operative Ataxia Rating Scale, ADI — activities of daily living, BI — Barthel Index, EDSS — Expanded Disability Status Scale, ADU — assistive device user, ADNU — assistive device non-user SD — standard deviation, p — level of significance.

minimal detectable change (MDC) for the BBS and the SARAbal.

The MDC is described as the smallest change that an outcome measure detects due to a notable change in the participants' performance. The established MDC is a reflection of the SEM for the measures of balance and could be considered as a "proxy" for the MDC. The term "proxy" in statistics refers to a value that is probably not in itself of any great interest, but from which a variable of interest can be obtained. The MDC was estimated using a data-driven method proposed by Wyrwich *et al.*²⁰ The Cronbach alpha of the measures of balance was used to estimate the SEM that reflected the MDC. Therefore, the derived MDC provides meaningful information on the expected change in score that may be perceived to be clinically meaningful for the patient following intervention. Future studies may use the obtained MDC as reference scores for reporting their results.

The second method of assessing interpretability is to compare the scores with normative data in a way that the difference in the score reflects the magnitude of difference between the tested sample and an age-matched healthy peer. The BBS has established normative data among community dwelling healthy older adults.⁹ Being condition-specific and relatively new, the SARA does not have established normative data. Future studies are recommended to establish the normative scores for the SARAbal among healthy older adults.

Acceptability of the measures of balance

The response rate to the outcome measures was high for the psychometric analysis and there were no missing items in our data providing evidence for acceptability.⁵ Acceptability can also be demonstrated by determining the floor and ceiling effect of the tool. These estimates report on the level of ease to complete the items i.e., were the contents of the tool too easy or too difficult or tolerable for the tested population? Determining the acceptability was outside the scope of the psychometric analysis; however, based on the findings of the psychometric analysis, the answer to this question may be partially resolved. Of the 60 participants, only one (2%) had difficulty in completing all four assessments due to fatigue providing some evidence for acceptable floor effect. The participants were able to complete all four tests. Eight (13%) participants obtained full score for BBS and

five (8%) for the SARAbal. However, we hesitate to comment on the question "were the contents too easy to complete?" Future studies are recommended to estimate the floor and ceiling effect for these measures of balance.

Precision of the measures of balance

The psychometric property analysis estimated discriminant validity by sub-dividing the participants into assistive device users and non-users. Mann-Whitney *U* test was used to determine the group differences between the two known groups (assistive device users and non-users). The findings of this analysis revealed a statistically significant ($p < 0.01$) difference between the two groups for both the measures of balance providing evidence for precision.⁴ Secondly, it is recommended that the precision could be derived by estimating the unidimensionality of the measures under consideration. However, unidimensionality estimation was outside the scope of the psychometric analysis. Therefore, we recommend future studies to conduct Rasch analysis or factorial analysis to provide evidence for unidimensionality for these measures in future.

Feasibility of the measures of balance

Based on the experience gained during data collection, the two measures of balance took 15–20 min to complete. They did not require the use of sophisticated equipment and are available at free of cost. In addition, formal training is not needed to perform these tests (the measures include instruction). However, the examiners who conducted these tests were qualified physiotherapists and therefore, the feasibility of administration is limited to qualified physiotherapists. With regard to patient safety, it is recommended that the assessment room is well-lighted, surface is non-slippery, and adequate rest breaks are given between the assessment sessions. There were no adverse events documented during data collection providing evidence for feasibility of the measures.

Generalizability of the findings

The findings of the present study are based on the outcomes of a psychometric analysis conducted earlier. As reported, the psychometric property

analysis recruited people with CA secondary to multiple sclerosis. The recruited sample was heterogeneous in terms of disease course of multiple sclerosis which enables the generalizability of findings to all types of multiple sclerosis. In addition, the sample was homogenous in terms of the type of lesion. The included participants with multiple sclerosis were restricted to primary cerebellar impairment. Therefore, these recommendations may be considered for people with other types of cerebellar ataxic lesions.

Conclusion

The findings of this study suggests that the BBS and SARAbal are psychometrically sound, appropriate, interpretable, precise, acceptable and feasible for the assessment of balance in people with CA and multiple sclerosis. Future studies are warranted to estimate the structural validity, responsiveness, MCID, plus floor and ceiling effect for these measures to strengthen the present findings.

Conflict of Interest

The authors declare that there is no conflict of interest relevant to the study.

Funding/Support

This research was funded by the University of Otago, New Zealand Doctoral Scholarship, School of Health and Rehabilitation Research, University of Pittsburgh, USA and Division of Health Science, University of Otago, New Zealand, travel grants, Maurice and Phyllis Paykel Trust (MPPT) fund Physiotherapy New Zealand, Otago Branch Educational Fund, Physiotherapy New Zealand's Neurology Special Interest Group Grant, School of Physiotherapy, Research fund, and was partially and financially supported by King Saud University, through Vice Deanship of Research Chairs, Rehabilitation Research Chair.

Author Contributions

Dr. Stanley John Winser contributed to structuring study design, write up for funding, subject recruitment, data collection, data review, data analysis, data interpretation, project management, writing

the manuscript and revising the manuscript. Dr. Catherine Smith contributed to write up for funding, data review, data interpretation, project management and writing the manuscript. Prof. Leigh A Hale contributed to write up for funding, data interpretation, project management and writing the manuscript. Dr. Leica S Claydon contributed to write up for funding, data review, data interpretation, project management and writing the manuscript. Prof. Susan L Whitney contributed to data interpretation, project management and writing the manuscript.

References

1. Palliyath S, Hallett M, Thomas SL, Lebedowska MK. Gait in patients with cerebellar ataxia. *Mov Disord* 1998;13(6):958–64.
2. Winser SJ, Smith CM, Hale LA, Claydon LS, Whitney SL, Mehta P. Systematic review of the psychometric properties of balance measures for cerebellar ataxia. *Clin Rehabil* 2015;29:69–79.
3. Winser SJ, Smith C, Hale LA, Claydon LS, Whitney SL. Balance outcome measures in cerebellar ataxia: A Delphi survey. *Disabil Rehabil* 2015;37:165–70.
4. Winser SJ, Smith C, Hale LH, et al., Psychometric properties of a core set of measures of balance for people with cerebellar ataxia secondary to multiple sclerosis. *Arch Phys Med Rehabil* 2016; in press.
5. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess* 1998;2:1–74.
6. Winser SJ, Smith C, Hale LH, et al., Balance assessment in multiple sclerosis and cerebellar ataxia: Rationale, protocol and demographic data. *Phys Med Rehabil Int* 2014;1(5):6.
7. Berg K, Wood-Dauphinee S, Williams JI. The balance scale: Reliability assessment with elderly residents and patients with an acute stroke. *Scand J Rehabil Med* 1995;27(1):27–36.
8. Korner-Bitensky N, Wood-Dauphinee S, Teasell R, et al., Best versus actual practices in stroke rehabilitation: Results of the Canadian national survey. *Stroke* 2006;37:631.
9. Steffen TM, Hacker TA, Mollinger L. Age-and gender-related test performance in community-dwelling elderly people: Six-minute walk test, Berg balance scale, timed up & go Test, and gait speeds. *Phys Ther* 2002;82(2):128–37.
10. Cattaneo D, Jonsdottir J, Repetti S. Reliability of four scales on balance disorders in persons with multiple sclerosis. *Disabil Rehabil* 2007;29:1920–5.
11. Cattaneo D, Regola A, Meotti M. Validity of six balance disorders scales in persons with multiple sclerosis. *Disabil Rehabil* 2006;28(12):789–95.

12. Schmitz-Hübsch T, du Montcel ST, Baliko L, et al., Scale for the assessment and rating of ataxia: Development of a new clinical scale. *Neurology* 2006;66(11):1717–20.
13. Bürk K, Mälzig U, Wolf S, et al., Comparison of three clinical rating scales in Friedreich ataxia (FRDA). *Mov Disord* 2009;24(12):1779–84.
14. Lee YC, Liao YC, Wang PS, Lee IH, Lin KP, Soong BW. Comparison of cerebellar ataxias: A three-year prospective longitudinal assessment. *Mov Disord* 2011;26(11):2081–7.
15. Portney LG, Watkins MP. Reliability of measurements. In: *Foundations of Clinical Research Application to Practice*. 3rd ed. New Jersey, USA: Julie Levin Alexander, 2009:77–96.
16. Mokkink LB, Terwee CB, Patrick DL, et al., The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63(7):737–45.
17. Portney LG, Watkins MP. Statistical measures of validity. In: *Foundations of Clinical Research Application to Practice*. 3rd ed. New Jersey, USA: Julie Levin Alexander, 2009:620–25.
18. Brenner MH, Curbow B, Legro MW. The proximal–distal continuum of multiple health outcome measures: The case of cataract surgery. *Med Care* 1995;AS236–44.
19. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10(4):407–15.
20. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol* 1999;52(9):861–73.