Medicine®

OPEN

Reliability of functional gait assessment in patients with Parkinson disease
Interrater and intrarater reliability and internal consistency

Yaqin Yang, MD, Yongjun Wang, MD*, Yanan Zhou, BD, Chen Chen, BD, Deli Xing, BD

Abstract

To determine the reliability of the functional gait assessment (FGA) as a measure of balance and gait in patients with Parkinson disease (PD).

This study involved121 inpatients with PD (mean age 61.9 years). The participants were scored using the FGA by 2 raters, and the testing procedure was videotaped. One of the raters re-assessed the same FGA test via review of the videotaped test 4 weeks later.

The interrater and intrarater reliability of the total FGA score was found to be excellent, with an intraclass correlation coefficient of 0.99. The reliability of single items of the FGA ranged from 0.49 to 0.99. The internal consistency of the FGA scores was 0.94.

The FGA has high inter-rater and intrarater reliability, and internal consistency for evaluating balance and gait disorders in patients with PD.

Abbreviations: BBS = Berg Balance Scale, BESTest = Balance Evaluation Systems Test, CI = confidence interval, DGI = Dynamic Gait Index, FGA = functional gait assessment, HY = modified Hoehn and Yahr scale, ICC = intraclass correlation coefficient, PD = Parkinson disease.

Keywords: assessment, balance, gait, Parkinson disease, reliability

1. Introduction

Parkinson disease (PD) is a common neurodegenerative disease that is typically manifested clinically by dyskinesia, including tremors, muscle rigidity, bradykinesia, and gait abnormalities. Postural instability and balance problems are troublesome symptoms for patients with PD, as they seriously affect quality of life and may lead to falls and a subsequent increase in morbidity and mortality.[1–4] Many clinical scales have been developed to objectively evaluate balance and gait disorders. The functional gait assessment (FGA) proposed by Wrisley et al[5] in 2004 is one such scale used to measure disturbances in balance and gait.

The FGA was developed as a modified version of the Dynamic Gait Index (DGI).[6] The DGI has been used for several years to

Department of Neurology, Beijing Tiantan Hospital, Capital Medical University, Beijing, China.

* Correspondence: Yongjun Wang, No. 6 Tiantanxili, Dongcheng District, Beijing 100050, China (e-mail: yongjunwang1962@sina.com).

assess postural stability during gait tasks in adults over 60 years of age.[7–10] However, some researchers discovered that the DGI was not sensitive when used for patients with walking impairments.[11,12] Additionally, the instructions for the DGI are ambiguous for several items, leading to difficultly in scoring by raters. To address these problems, Wrisley et al[5] created the FGA by modifying the DGI. The FGA is a 10-item gait test that includes 7 of the 8 items from the original DGI (item 7, "walking around obstacles," was of insufficient difficulty to be included) and adds 3 new items. The 10 items of the FGA are as follows: gait on a level surface; change in gait speed; gait with horizontal head turns; gait with vertical head turns; gait and pivot turn; step over obstacle; gait with narrow base of support; gait with eyes closed; ambulating backwards; and steps. Each item is scored on a 4-pointordinal scale with scores of 0, 1, 2, and 3. The maximum total score is 30; higher scores represent better balance and gait ability.[4,5]

The reliability of a new scale should be tested in different people. Scale reliability refers to the degree of approximation of repeated measurements under the same conditions. It is used to evaluate scale stability and consistency, and may vary over time and between subjects. Reliability encompasses both external and internal reliability. The former includes intrarater and inter-rater reliability, whereas the latter refers to the consistency of the elements in the scale's items.

Since the FGA has been published, several researchers have analyzed its reliability in patients with different diseases and disorders. Wrisley et al[5] tested the reliability of the FGA in patients with vestibular disorders. The inter-rater and intrarater reliability of the total FGA score were found to be high, with intraclass correlation coefficients (ICCs) of 0.84 and 0.83, respectively.[5] The internal consistency of the FGA was determined to be 0.79.[5] Walker et al[13] published reference-group data for the FGA by stratifying people aged 40 to 89 years

into decade cohorts. The ICC for inter-rater reliability was 0.93 in their study. The mean FGA scores systematically decreased with increasing age, and ranged from 29/30 for adults in their 40s to 21/30 for adults in their 80s. At the same time, there was an increase in the standard deviation of the total scores with each decade, demonstrating that the variability of the performance on the FGA increased with age, which is in agreement with the findings of other research studies on gait and balance in old adults.[14–16] Thieme et al[17] tested the reliability and validity of the FGA in patients with subacute stroke; the intrarater and inter-rater reliability of total FGA scores were found to be excellent, with ICCs of 0.97 and 0.94, respectively. Leddy et al[18] compared the Berg Balance Scale (BBS), FGA, and a newly developed Balance Evaluation Systems Test (BESTest) in community-dwelling patients with PD. The inter-rater reliability was excellent for all 3 tests, with ICCs greater than 0.93.

Since the FGA was developed, its reliability has been assessed in subjects with vestibular dysfunction, community-dwelling elders, and stroke patients.[5,13,17] Leddy et al[18] reported the reliability of the FGA in community-dwelling patients with PD. The purpose of the present study was to evaluate balance and gait disorders in hospitalized patients with PD to further verify the reliability of the FGA. We predicted that the FGA would be a reliable means of assessing balance and gait in patients with PD, based on previous research. This study may provide tools for the clinical assessment and rehabilitation of patients with PD, and facilitate their motor-function training.

2. Materials and methods

2.1. Participants

All inpatients with PD who were hospitalized in the Movement Impairment Ward of the Department of Neurology at Beijing Tiantan Hospital between March 2011 and December 2011 were screened. In all, 121 in-patients (82 males and 39 females), representing 28.3% of the 428 patients who were screened, were enrolled in this study. The primary reason for hospitalization was to optimize PD treatment (medication adjustments). The anti-Parkinson medicines administered included compound levodopa (levodopa with benserazide or carbidopa), dopamine receptor agonists (piribedil and pramipexole), catechol-O-methyltransfer-ase inhibitor (entacapone), amantadine, and antimuscarinic drugs (Artane).

All participants met the following inclusion criteria: diagnosed with idiopathic PD according to the diagnostic criteria of the UK Parkinson Disease Society Brain Bank,[19] able to stand still without support for at least 1 minute, and Mini-Mental State Examination score ≥24. Subjects with the following criteria were excluded: diagnosed with secondary Parkinson syndrome or Parkinson plus syndrome, inability to walk at least 10 m without physical assistance or walking aids (no participant was allowed the use of a walking assist device), or presence of a comorbidity affecting motor function (such as stroke, amputation, or visual impairment).[4]

2.2. Ethics statement

This study was approved by the Medical Ethics Committee of Beijing Tiantan Hospital, in compliance with the Declaration of Helsinki. All participants or their legal representatives signed informed consent forms from the Medical Ethics Committee of Beijing Tiantan Hospital.

2.3. Procedure

Patient characteristics were recorded upon admission. The following variables were extracted from the participants' medical records: age, sex, duration of PD (in years), medical and surgical history, current medication regimen, and modified Hoehn and Yahr (HY) scale.[20] The HY scale is used to evaluate disease severity and duration (higher scores indicating worse impairment).[20] FGA was conducted in the rehabilitation room of the Department of Neurology at Beijing Tiantan Hospital. All participants underwent FGAs performed by 2 designated licensed physical therapists (rater A and rater B) at the same time. Both raters had received FGA training and practiced the FGA on 2 healthy adults and 2 patients with PD. The assessments were completed within 1 day, 24 to 48 hours after hospital admission. The FGA was conducted in the ON medication phase (approximately 1 hour after taking anti-PD medications). The evaluation time was 10 to 30 minutes.[4]

The 2 raters scored the performance on the FGA by directly observing the patients. They were instructed not to discuss the grading criteria or the test with each other during the entire sequence, and were blinded to each other's results. Additionally, the FGA procedure was video-recorded. Rater A assessed the performances of the same FGA test at separate time points via review of the videotaped recording. There was a minimum interval of 4 weeks between the 2 ratings to minimize possible memory effects.

2.4. Data analysis

All analyses were performed using SPSS version 17.0 (SPSS Inc., Chicago, IL). For sample characteristics, descriptive statistics were used. The inter-rater and intrarater reliability of the total FGA score were statistically evaluated using ICCs and a 2-way random-effects model. Inter-rater reliability was calculated using the 2 raters' scores obtained after direct observation of the FGA test. Intrarater reliability was calculated using the 2 ratings of rater A obtained after direct and videotaped observation. The reliability of single items was calculated using the weighted kappa statistic. Agreement strengths for kappa values were classified as follows: <0, poor; 0.00 to 0.20, slight; 0.21 to 0.40, fair; 0.41 to 0.60, moderate; 0.61 to 0.80, substantial; and 0.81 to 1.00, almost perfect.[21] Internal consistency or the homogeneity of the FGA was determined using Cronbach alpha. A Cronbach alpha value ≥0.80 indicated good internal consistency.[22] The correlation coefficient of 1 item with the total score of the remaining items was denoted by the "corrected item-total correlation." A coefficient less than 0.40 indicated a low degree of correlation between the item and the remaining items.[22]

3. Results

A total of 121 inpatients (82 males and 39 females) completed the study. The baseline patient characteristics, FGA scores, and HY stages are shown in Table 1.

3.1. Inter-rater and intrarater reliability

The results for inter-rater and intrarater reliability are shown in Table 2. Inter-rater reliability was found to be excellent, with an ICC for the total score of 0.99 (95% confidence interval [CI] 0.99–1.00). Kappa values for single items ranged from 0.49 (item 3) to 0.98 (item 8). Kappa values for items 3 and 4 were fair at

Sex (n)	
Men	82
Women	39
Age, y	
Men	
Range	41–81
Mean ± SD	61.8 ± 11.5
Women	
Range	43–79
Mean ± SD	62.1 ± 9.2
Course of disease, y	
Range	0.5–39
Median (25th, 75th percentile)	3.0 (2.0, 5.0)
FGA	
Range	0–30
Median (25th, 75th percentile)	22 (15, 26)
HY (n)	
1.0	28
1.5	33
2.0	16
2.5	13
3.0	18
4.0	13

Report mean ± standard deviations (SDs) when data are normally distributed; report median (25th, 75th percentile) when data are not normally distributed.

FGA = functional gait assessment, HY = modified Hoehn and Yahr scale.

0.49 and 0.60, respectively, and those for the other items were substantial or almost perfect.[21]

Excellent intrarater reliability was observed for the total FGA score, with an ICC of 0.99 (95% CI 0.99–1.00). For single items, kappa values ranged from 0.91 (item 4) to 0.99 (items 6–10). Intrarater reliability for all items was almost perfect.[21]

3.2. Internal consistency

The Cronbach alpha value for the total FGA score was 0.94. The corrected item-total correlations ranged from 0.65 to 0.80. Each item of the FGA was omitted in turn, and the alpha value of the remaining items was calculated; the alpha value thus calculated was found to be 0.93 in all cases (Table 3).

4. Discussion

Intrarater and inter-rater reliability represents the external reliability of a scale, and refers to the consistency of measurements obtained at different times or by different evaluators, respectively. When the FGA was used to assess hospitalized patients with PD, the inter-rater reliability of the whole scale was found to be excellent, with an ICC of 0.99. This value is higher than those previously reported for FGA reliability.[5,13,17,18] In the study on patients with vestibular dysfunction by Wrisley et al,[5] the raters were given just 10 minues to familiarize themselves with the FGA items, scoring rules, and guidance, and the inter-rater reliability was 0.84. In several other studies, the raters received prior training, and the inter-rater reliability was markedly higher (0.93–0.94), and was similar to the results of the present study.[13,17,18] These observations suggest that adequate training of raters on how to use the FGA scale improves evaluation reliability.

In the present study, the inter-rater reliability of a single item on the FGA ranged from 0.49 to 0.98. Low inter-rater reliability was observed for items 3 (gait with horizontal head turns) and 4 (gait with vertical head turns), with ICCs of 0.49 and 0.60, respectively. The other items showed high inter-rater reliability. The inter-rater differences for items 3 and 4 were mainly focused on scores 3 and 2, which referred to "Performs head turns smoothly with no change in gait" and "Performs head turns smoothly with slight change in gait velocity (e.g., minor disruption to smooth gait path)," respectively. It is likely that bradykinesia and walking slowly were the main clinical manifestations in PD patients, some of whom may have exhibited freezing behaviors. Therefore, slow walking and walking pauses were common manifestations. The difficulty in clearly defining the boundaries between no change and slight change in gait velocity likely led to the scoring discrepancies.

In the present study, the whole scale and individual FGA items had excellent intrarater reliability. The ICC of the full FGA was 0.99, and the kappa values of individual items ranged from 0.91 to 0.99, which are higher than the values previously reported in patients with vestibular dysfunction, stroke, and PD.[5,17,18] In the vestibular dysfunction study, intrarater reliability with a 2-hour interval was just 0.83,[5] whereas secondary evaluation of video data from the first evaluation of stroke patients yielded an intrarater reliability of 0.97,[17] which is similar to that in the

	Inter-rater reliability		Intrarater reliability	
	ICC/kappa	95% CI	ICC/kappa	95% CI
FGA total	0.99	0.99–1.00	0.99	0.99–1.00
FGA-item 1	0.89		0.97	
FGA-item 2	0.75		0.95	
FGA-item 3	0.49		0.94	
FGA-item 4	0.60		0.91	
FGA-item 5	0.75		0.93	
FGA-item 6	0.87		0.99	
FGA-item 7	0.87		0.99	
FGA-item 8	0.98		0.99	
FGA-item 9	0.97		0.99	
FGA-item 10	0.94		0.99	

Reliability of total score was statistically evaluated by ICC; reliability of single items was calculated by kappa statistic.

CI = confidence interval, FGA = functional gait assessment, ICC = intraclass correlation coefficient.

Table 3

Internal consistency of FGA.

Corrected item-total correlations		Alpha values if item deleted
FGA-item 1	0.73	0.93
FGA-item 2	0.76	0.93
FGA-item 3	0.80	0.93
FGA-item 4	0.76	0.93
FGA-item 5	0.65	0.93
FGA-item 6	0.77	0.93
FGA-item 7	0.79	0.93
FGA-item 8	0.76	0.93
FGA-item 9	0.71	0.93
FGA-item 10	0.75	0.93

FGA = functional gait assessment.

present study. In the study by Leddy et al[18] on patients with PD, 2 evaluations were performed 2 weeks apart. Even though the evaluations were performed at the same time of day to try to ensure that the patients were in similar states, intrarater reliability showed considerable fluctuation.[18] The functional status of PD patients is affected by multiple factors, including anti-PD drugs, physiological function, mood, and diet, and patients may show significant differences in balance and gait between the "ON" and "OFF" periods of drugs. In the present study, videotaped data were used to overcome the abovementioned disadvantages. The above findings suggest that appropriate method selection is important for FGA intrarater reliability.

Internal reliability refers to the consistency of the elements in a scale's items, and is usually expressed as the Cronbach alpha value, with higher values signifying greater internal reliability. "Corrected item-total correlation" is an indicator used to judge if the homogeneity of the latent trait of the items to be measured (characteristics of balance and gait in the present study) is identical to that of the remaining items. After a certain item is omitted, the alpha of the total score of the remaining items should become smaller than the Cronbach alpha of the total scale. If the alpha of the scale becomes larger after a certain item is deleted, then the homogeneity of the latent trait of the item to be measured is not identical to that of the other items.

In the present study, the Cronbach alpha for the total FGA score was 0.94, which was higher than that reported by Wrisley et al[5] in patients with vestibular dysfunction. This indicates that the internal consistency of the FGA was ideal when applied to patients with PD. The "corrected item-total correlation" ranged from 0.65 to 0.80, which is >0.40, suggesting that there was good correlation between each item and the remaining items, and that the elements of the measurements were consistent. Of these, item 5 (gait and pivot turn) had the lowest correlation with the other items ($r = 0.65$). This item required the patients to stand and turn around while they walked, and the score was based on the time necessary to perform this action. Due to marked bradykinesia, patients with PD have difficulty turning over, standing, sitting, and turning around, leading to the phenomenon of freezing gait. Item 5 assessed a major gait problem experienced by patients with PD. In contrast, the other items focused on common actions related to balance and walking; therefore, they might not reflect the gait disorder, and also item 5 in PD patients. Consequently, item 5 weakly correlated with the other items. All alpha values calculated after the omission of a single item were 0.93. Because this value is <0.94, it indicates that the same concept was measured by every single item and the remaining items, namely, balance and gait disorder.

4.1. Limitations

The present study population comprised PD patients who were hospitalized at a single center. It is uncertain whether the same conclusions would be reached if the FGA was applied to other PD populations, such as community-dwelling individuals. The use of walking-assist devices is allowed in the FGA, but no participant was allowed to use them in this study. This means that more patients with relatively serious conditions were not included in this study, which could affect the scale reliability. In the present study, the 2 raters were female physical therapists who were working in the Department of Neurology, had graduated from medical school within the past 2 years, and had similar educational backgrounds and practical experience. It is possible that the reliability may change if the raters have different backgrounds.

5. Conclusions

The FGA has high external and internal reliability with good scale stability for evaluating balance and gait disorders in patients with PD. To improve its clinical utility, further studies should assess the concurrent validity, discriminative validity, and predictive validity of falls, and also the responsiveness of the FGA in patients with PD.

Acknowledgment

References

[1] Muslimovic D, Post B, Speelman JD, et al. Determinants of disability and quality of life in mild to moderate Parkinson disease. Neurology 2008;70:2241–7.

[2] Dibble LE, Addison O, Papa E. The effects of exercise on balance in persons with Parkinson's disease: a systematic review across the disability spectrum. J Neurol Phys Ther 2009;33:14–26.

[3] Ashburn A, Stack E, Ballinger C, et al. The circumstances of falls among people with Parkinson's disease and the use of Falls Diaries to facilitate reporting. Disabil Rehabil 2008;30:1205–12.

[4] Yang YQ, Wang YJ, Zhou YN, et al. Validity of the functional gait assessment in patients with Parkinson disease: construct, concurrent, and predictive validity. Phys Ther 2014;94:392–400.

[5] Wrisley DM, Marchetti GF, Kuharsky DK, et al. Reliability, internal consistency, and validity of data obtained with the functional gait assessment. Phys Ther 2004;84:906–18.

[6] Shumway-Cook A, Woollacott MH. Motor Control: Theory and Practical Applications. Baltimore, MD:Lippincott Williams & Wilkins; 1995.

[7] Galgon AK. Berg balance score, Dynamic Gait Index, and walking speed as outcome measures for individuals after stroke in an outpatient physical therapy setting [abstract]. Neurol Rep 2003;26:196–7.

[8] Gill-Body KM, Murphy MP, Sullivan PE. Concurrent validity of the Dynamic Gait Index in community dwelling elderly [abstract]. Neurol Rep 2003;26:196–7.

[9] Vereek L, Truijen S, Wuyts FL, et al. The Dizziness Handicap Inventory and its relationship with functional balance performance. Otol Neurol 2007;28:87–93.

[10] Legters K, Whitney SL, Porter R, et al. The relationship between the activities specific balance confidence scale and the dynamic gait index in peripheral vestibular dysfunction. Physiother Res Int 2005;10:10–22.

[11] Wrisley DM, Walker ML, Echternach JL, et al. Reliability of the Dynamic Gait Index in people with vestibular disorders. Arch Phys Med Rehabil 2003;84:1528–33.

[12] Whitney SL, Wrisley DM, Brown KE, et al. Is perception of handicap related to functional performance in persons with vestibular dysfunction? Otol Neurotol 2004;25:139–45.

[13] Walker ML, Austin AG, Banke GM, et al. Reference group data for the functional gait assessment. Phys Ther 2007;87:1468–77.

[14] Lusardi MM, Pellecchia GL, Schulman M. Functional performance in community living older adults. J Geriatric Phys Ther 2003;26:14–22.

[15] Fregly AR, Smith MJ, Graybiel A. Revised normative standards of performance of men on a quantitative ataxia test battery. Acta Otolaryngol 1973;75:10–6.

[16] Whitney SL, Wrisley DM, Marchetti GF, et al. The effect of age on vestibular rehabilitation outcomes. Laryngoscope 2002;112:1785–90.

[17] Thieme H, Ritschel C, Zange C. Reliability and validity of the functional gait assessment (German Version) in subacute stroke patients. Arch Phys Med Rehabil 2009;90:1565–70.

[18] Leddy AL, Crowner BE, Earhart GM. Functional gait assessment and Balance Evaluation System Test: reliability, validity, sensitivity, and specificity for identifying individuals with Parkinson disease who fall. Phys Ther 2011;91:102–13.

[19] Meara J, Bhownick BK, Hobson P. Accuracy of diagnosis in patients with presumed Parkinson disease. Age Ageing 1999;28: 99–102.

[20] Goetz CG, Poewe W, Rascol O, et al. Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations. Mov Disord 2004;19:1020–8.

[21] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.

[22] Henson RK. Understanding internal consistency reliability estimates: a conceptual primer on coefficient alpha. Measurement Eval Counsel Dev 2001;34:177–89.