Utility of the Mini-BESTest, BESTest, and BESTest Sections for Balance Assessments in Individuals with Parkinson Disease

Abigail L. Leddy, BS,
Washington University School of Medicine

Beth E. Crowner, PT, DPT, NCS, MPPA, and
Washington University School of Medicine

Gammon M. Earhart, PT, PhD
Washington University School of Medicine

Abstract

Background and Purpose—The Balance Evaluation Systems Test (BESTest) has been shown to be a reliable and valid measure of balance in individuals with Parkinson disease (PD). A less time-consuming assessment may increase clinical utility. We compared the discriminative fall risk ability of the Mini-BESTest to that of the BESTest, and determined the reliability and normal distribution of scores for each section of the BESTest and the Mini-BESTest in individuals with PD.

Methods—Eighty individuals with idiopathic PD were assessed using the BESTest and Mini-BESTest. A faller was defined as an individual with 2 or more falls in the prior 6-months. Subsets of individuals were used to determine inter-rater (n=15) and test-retest reliability (n=24).

Results—The Mini-BESTest, total BESTest score, and all sections of the BESTest, showed a significant difference between the average scores of fallers and non-fallers. For both the Mini-BESTest and BESTest, inter-rater (intraclass correlation ICC≥0.91) and test-retest (ICC≥0.88) reliability was high. The Mini-BESTest and BESTest were highly correlated (r=0.955). Accuracy of identifying a faller was comparable for the Mini-BESTest and BESTest (area under the ROC plots =0.86 and =0.84, respectively).

Discussion—No specific section of the BESTest captured the primary balance deficit for individuals with PD. The post-test probabilities for discriminating fallers versus non-fallers were comparable-to-slightly stronger when using the Mini-BESTest.

Conclusion—Although the Mini-BESTest has fewer than half of the items in the BESTest and takes only 15 minutes to complete, it is as reliable as the BESTest and has comparable-to-slightly greater discriminative properties for identifying fallers in individuals with Parkinson Disease.

INTRODUCTION

Individuals with Parkinson disease (PD) have balance impairments and postural instability. This leads to an increased risk of falling and an increase in both fractures and soft tissue injury.[1–5] Decreased postural stability is also associated with decreased quality of life,[6] and

individuals with PD tend to limit their activity levels due to a fear of falling, contributing to inactivity and further compromising balance.[2,4,7] Some studies have shown that over 60% of individuals with PD fall at least once in a 12-month period.[8,9] Due to the negative impact of balance impairments on individuals with PD, it is important to be able to assess who has a balance deficit and is at increased risk of falling. Identification of those at risk is essential in order to intervene, as physical therapy and exercise have been shown to improve balance.[10] Currently, many different balance outcome measures and fall risk assessments are being used in individuals with PD.[11–15] Many of these measures have limitations exist including low sensitivity and/or specificity,[11,16] ceiling effects,[17,18] and inclusion of items that physical therapy intervention cannot address (such as prior number of falls).[14,19,20] Some research has shown that a battery of tests is necessary to fully assess balance, however, a consensus on which tests to include and interpretation of results from multiple tests has not been reached.[12,16,21]

Balance is a complex construct with many contributing physiological systems. To help differentiate between possible causes of imbalance, the Balance Evaluation Systems Test (BESTest) was developed.[22] The BESTest is a balance assessment for mixed populations that includes 36-items divided into six sections, each of which is designed to test a theoretical control system for balance: Section I-Biomechanical Constraints, Section II-Stability Limits and Verticality, Section III-Anticipatory Postural Adjustments, Section IV-Postural Responses, Section V-Sensory Orientation, and Section VI-Stability in Gait.[22] Although there is some overlap between the sections, ideally the scores across sections allow a physical therapist to choose treatments that will focus on the primary deficits causing the balance impairment. Each section has also been proposed to be used separately if needed.[22]

The BESTest total score has been shown to have high inter-rater reliability and moderate concurrent validity with an individual's self-perceived balance in a mixed population including individuals with PD, vestibular dysfunction, hip arthroplasty, peripheral neuropathy, and healthy controls.[22] In this mixed population, each section of the BESTest had high inter-rater reliability. Another study showed that the BESTest is a valid measure of balance for individuals with PD, having high correlations with the Berg Balance Scale, disease severity, and the Activities Specific Balance Confidence Scale (ABC). The BESTest was more sensitive and specific at identifying who was a faller and a non-faller, more accurate overall at discriminating a faller, and had more beneficial post-test probabilities than the Berg.[23]

The length of BESTest, taking 30–35 minutes to perform, decreases its clinical utility and feasibility. To address the limitations associated with this lengthy time requirement, and to improve the psychometric properties and unidimensional structure of the BESTest, the Mini-BESTest was developed. The Mini-BESTest is a shortened version that includes only 16 of the original 36 BESTest items, is believed to measure dynamic balance, and takes only 10–15 minutes to perform.[24] Dynamic balance is balance that is associated with movement during transfers and gait, as well as external perturbations and cognitive dual task performance.

This study aimed to the compare the discriminative capability of the Mini-BESTest to the BESTest, as well as determine the reliability and utility of the Mini-BESTest, and of each section of the BESTest in individuals with PD. We hypothesized that the Mini-BESTest would be reliable and perform comparably to the BESTest in identifying fallers. We also hypothesized that individuals with PD might have lower scores on Section I-Biomechanical Constraints and Section IV-Postural Responses compared to other sections of the BESTest. The reliability and discriminative properties of the full BESTest have been presented

elsewhere.[25] This study further explored the reliability of each section of the full BESTest, and compared scores between fallers and non-fallers for each section.

METHODS

Participants

Eighty individuals with PD completed the study, with fallers defined as those who had 2 or more self-reported falls in the previous 6 months. Participants were recruited using a database from the Washington University in St. Louis Movement Disorders Center database. Individuals were called, using a random number generator, from a stratified list based on Hoehn and Yahr (H&Y) staging to ensure participants across stages of disease severity were recruited. Inclusion criteria included a diagnosis of idiopathic PD, greater than 40 years old, H&Y stages I–IV, community-dwelling, and the ability to follow commands and give informed consent. Exclusion criteria included a diagnosis of atypical PD and a prior surgical treatment of PD. Individuals from the St. Louis area who heard about the study from other participants or through the Volunteers for Health database also participated. Eighty-two individuals met the inclusion criteria and were tested; however, two participants were unable to complete the study due to unrelated illness or exclusion criteria. The study was approved by the Human Research Protection Office, and all participants provided informed consent prior to participating.

Testing Procedure

Evaluations were performed at Washington University in St. Louis in a laboratory setting. Participants provided demographic and recent fall information. Disease severity was measured using the Movement Disorders Society Unified Parkinson Disease Rating Scale (MDS-UPDRS) and modified H&Y staging.[26,27] The BESTest with slight modifications and Mini-BESTest were performed along with a larger battery of balance and gait tests which took a total of two hours. The full battery included the Berg Balance Scale, Functional Gait Assessment, 6-minute walk test, and 10-meter walk test, as well as a series of questionnaires. Fall status was assessed via questionnaire, so raters were blinded to fall status until after completion of testing. No specific definition of a fall was described. Participants were tested on medication, as they were instructed to take their medication according to their normal regimen.

The full, 36-item BESTest with slight modifications was administered with shoes off, unless discomfort was expressed by the participant. The balance test is scored using a 4-point scale, 0 to 3, with a higher score indicating better balance. The maximum score for the full BESTest is 108. All sections of the test were performed as described by Horak et al.,[22] except slight alterations in Section IV and V. In Section IV items 16–18, two trials were performed with the second trial always being rated. This was to allow consistency, as many individuals would not lean appropriately beyond their stability limits into the tester's hands during the first trial to adequately assess compensatory stepping. In Section V, only one trial was used for each item due to time constraints. During item 27, the dual-task timed-up-and-go, option b (listing random numbers) was always used. The score for each section and the total BESTest score were converted into percentage scores, as is designated in the BESTest instructions.[22]

The Mini-BESTest is rated on a 3-point scale from 0 to 2. A perfect score is 32. The final score was reported in both crude score and a percentage score for comparison with the BESTest. Although all items of the Mini-BESTest are included in the BESTest, the grading criteria are different. During the performance of the full BESTest, the items were graded

according to the criteria for both the BESTest and the Mini-BESTestso that no items had to be repeated.

Reliability

Inter-rater reliability of each Section of the BESTest and the Mini-BESTest was determined using a subgroup of participants (n=15: MDS-UPDRS=74.2± SD18.6, disease duration=6.8±SD3.26; H&Y stage 1 had 2 participants, stage 2 had 7 participants, stage 2.5 had 3 participants, stage 3 had 2 participants, stage 4 had 1 participant; 20% (n=3) were fallers with fallers at H&Y stages 2.5, 3, and 4). Scoring was performed by three raters (2 physical therapists and 1 physical therapy student); the physical therapists had 13 and 21 years of experience, and the student had finished 2 years of a Doctorate of Physical Therapy program. All raters watched the BESTest training video (available for purchase online)[28] and read the testing procedures. The raters had one training trial, testing one individual without PD prior to testing participants. The purpose of this trial was to familiarize raters with the order in which the items would be presented. All raters concurrently scored each evaluation, while one of the raters (author AL) administered the test. If an item was missed by any rater, the item was repeated and all raters scored the second trial. Scoring was not discussed.

Test-retest reliability was determined using a subgroup of study participants (n=24: MDS-UPDRS=71±SD21.9, disease duration=6.9±SD3.38; H&Y stage 1 had 2 participants, stage 2 had 11 participants, stage 2.5 had 6 participants, stage 3 had 3 participants, and stage 4 had 2 participants; 21% (n=5) were fallers with fallers at H&Y stages 2.5, 3, 3, 4, and 4). Participants were evaluated twice, with two weeks between evaluations (range=11–16 days). Testing was performed on medication at the same time of day for both test sessions to prevent possible balance fluctuations due to medication. After the initial 24 participants were tested and the reliability was determined to be acceptable, the physical therapy student (author AL) continued administering the evaluation for the remainder of the participants to reach a total of 80 participants.

Statistical Analysis

Statistics were calculated using SPSS for Windows (version 17.0, SPSS Inc., Chicago, IL). The distributions of scores were assessed using the Kolmogorov-Smirnov (K-S) test. Independent sample t-tests were used for comparisons between fallers and non-fallers for age, MDS-UPDRS total score, duration of disease, total BESTest score, and Mini-BESTest scores. Mann Whitney U tests were used for H&Y staging and for each section of the BESTest. Bonferroni correction was used for multiple comparisons, with p<0.00625 being statistically significant for comparisons between fallers and non-fallers in the balance testing and p<0.0125 for demographic and disease severity information. Pearson's rank correlation coefficient was used to compare the Mini-BESTest to the BESTest total score. Although the grading scale is ordinal for both the BESTest and the Mini-BESTest, we considered the score of each section of the BESTest and the total score of the BESTest and Mini-BESTest to be continuous data. Intraclass correlations (ICC 2,1), 2-way random, absolute agreement were used to assess the both inter-rater and test-retest reliability of the BESTest sections and the Mini-BESTest.

A receiver operating characteristic (ROC) plot was used to compare overall accuracy of the BESTest to the Mini-BESTest, as well as to determine appropriate cutoff scores for identifying a faller versus a non-faller. Cutoff scores were reported using two methods. The first method maximized both sensitivity and specificity, compromising between avoiding false negatives and false positives. The second method maximized sensitivity and minimized the negative likelihood ratio (LR−) for being a faller, which selected the score most likely to

identify as many fallers as possible without regard to increased false positives. A positive likelihood ratio (LR+) is the probability of someone who is a faller having a positive test divided by the probability of someone without the disease having a positive test (scoring below the cutoff score), and a LR-ratio is the probability of someone who is a faller having a negative test (scoring above the cutoff score) divided by the probability of someone without the disease having a negative test.[29] Bayes Theorem was used to calculate post-test probabilities of being a faller. Bayes Theorem quantifies the odds of someone being a faller after performing the balance test by multiplying the pre-test odds of being a faller by the likelihood ratio.[30] These odds can then be converted into the percent chance of being a faller.

Sample size calculations were based on α=0.05, with a power of 0.80. For reliability, a null ICC was 0.5, and an acceptable reliability was 0.8. For inter-rater reliability between 3 raters, a sample of 15 participants was required. For test-retest reliability using 2 test sessions, 22 participants were required. With an estimated fall rate of 30%, CI width of 0.20, and a 95% confidence interval, 81 participants were needed for sensitivity, specificity, and ROC plots.

RESULTS

Participant Demographics

Twenty-five of the 80 participants were fallers (31.3%). Participants had an average age of 68.2 (SD 9.3) years with an average H&Yscore stage of 2.45 (SD 0.64) and UPDRS score of 72.6 (SD 25.1). Disease duration, MDS-UPDRS score, and H&Y stage were all significantly different between fallers and non-fallers (p<0.001), but age was not (p=0.681). Demographic information for the overall sample, as well as the fallers and non-fallers separately is presented in Table 1.

Mini-BESTest

The Mini-BESTest scores were normally distributed (K-S, p=0.16). There was a statistically significant difference between fallers and non-fallers, with an average difference of 27% between the groups (Table 1). Mini-BESTest had high inter-rater and test-retest reliability (ICC= 0.91 and =0.92, respectively) (Tables 2 and 3).

BESTest and Sections

The total BESTest scores were normally distributed (K-S, p=0.47). When assessing individual sections of the BESTest, only Sections IV and V were not normally distributed (K-S, p=0.038 and p<0.001, respectively) and had unequal variance between groups. T-tests reported for these two sections are based on unequal variance assumptions. On average, the overall study sample scored lowest on Section I (Biomechanical Constraints) at 59%, and highest on Section II (Stability Limits /Verticality) at 80%.

For the BESTest total score and all sections of the BESTest, there was a statistically significant difference between fallers and non-fallers (Table 1). Section II showed the least difference between fallers and non-fallers, with an average difference of only 9%. All other sections showed 17–24% difference in average scores between fallers and non-fallers, while the overall BESTest showed a 19% difference.

Inter-rater reliability was lowest for Sections I and II with ICC=0.81 and =0.79, respectively. All other sections of the BESTest and the overall BESTest had high inter-rater reliability (ICC≥0.86) (Table 2). Test-retest reliability varied between sections of the BESTest, but it was high for the overall BESTest (ICC=0.88) (Table 3).

Correlation and Accuracy

The Mini-BESTest demonstrated a strong relationship with the BESTest total score (r=0.955). The Mini-BESTest overall accuracy for identifying who was a faller was comparable to the BESTest, with an area under the curve (AUC) of 0.86 (95% CI 0.76–0.95) for the Mini-BESTest and 0.84 (0.75–0.93) for the BESTest.

Cutoff Scores

When both sensitivity and specificity are maximized, a cutoff score of 20/32 (63%) was identified for the Mini-BESTest (sensitivity=0.88, specificity=0.78) and 69% was identified for the BESTest (sensitivity=0.84, specificity=0.76). When maximizing sensitivity and LR-, a cutoff score of 23/32 (72%) was identified for the Mini-BESTest (sensitivity=0.96, specificity=0.47) and 84% for the BESTest (sensitivity=1.0, specificity=0.39) (Figure 1 and Table 4).

DISCUSSION

This study expands upon previously reported data describing the reliability, validity, and discriminatory ability of the BESTest in individuals with PD.[25] This previous work demonstrated that the BESTest was superior to the Berg in discriminating between fallers and non-fallers. The present work shows a significant difference between fallers and non-fallers for all individual sections of the BESTest, and perhaps more importantly, that the Mini-BESTest is a reliable measure of balance in individuals with PD. The Mini-BESTest is comparable to the BESTest in its ability to discriminate between fallers and non-fallers, and is more feasible for clinical use than the full BESTest since it can be administered in a shorter time (estimated 15 minutes and 35 minutes, respectively).

BESTest Sections

Individuals with PD scored lowest in Section I (Biomechanical Constraints, 59%), as hypothesized, since postural abnormalities are known to be common in individuals with PD.[3] Average scores were highest on Section II (Verticality/Stability Limits, 80%), which is interesting as functional reach laterally and forward account for a over a third of this section and spinal flexibility is thought to be impaired in individuals with PD.[31] The remaining items in Section II, however, included leaning and midline orientation while in sitting, which might not be as affected in individuals with PD. Scores on Sections III-VI ranged from 68% to 75%. The inter-rater reliability for each section was very similar to what was reported in a mixed population.[22] Test-retest reliability was variable for the sections of the BESTest. This variability could be a result of the decreased heterogeneity of the smaller sections of the BESTest as well as the non-normal distribution of Sections IV and V, as ICC calculations inherently compare the measurement error to the total variability seen in the group.[32]

All sections of the BESTest were statistically different between fallers and non-fallers. There was not a specific section, or balance control system, that showed a more prominent change in those who are fallers. The fact that a specific section is not affected consistently could be due to the fact that there is overlap between the sections, and the 6 theoretical control systems have not been truly isolated. Alternatively, it may be that there are many underlying balance deficits that contribute to the increased risk for falling, including postural reflex, sensory, strength, motor control, postural adjustment, and attention allocation deficits in individuals with PD.[3,9,33] Horak et al.[22] reported that the three individuals with PD scored lower (50%) on Section IV (Postural Responses) than on other sections of the test (71–78%). This was not true of the present large sample of individuals with PD (n=80), as Section IV (68%) was not more affected on average than other sections. As such, one section of the

BESTest alone cannot be recommended to quantify the main balance deficit for individuals with PD. It is unknown whether the sections of the BESTest could be used to guide treatment on an individual basis, as different sections might be more or less affected for specific individuals and might allow for a more individualized treatment.

Mini-BEST and BESTest

Both the Mini-BESTest and the BESTest had high inter-rater and test-retest reliability. The BESTest has been shown to be a valid measure of balance in individuals with PD.[25] The Mini-BESTest was strongly correlated with the BESTest (r=0.955), signifying that the Mini-BESTest is also an appropriate measure of balance for individuals with PD.

When the BESTest was shortened to create the Mini-BESTest, no items from Section I or Section II were included. Section I of the BESTest includes items assessing foot deformities, center of mass alignment, ankle and hip range and strength, and a floor to stand transfer. Section II includes functional reach forward, functional reach laterally, and ability to lean and return to vertical while in sitting. According to Franchignoni,[24] the items in these sections did not truly measure dynamic balance and were therefore removed. The improved unidimensionality of the Mini-BESTest compared to the BESTest is supported by the increased difference between average scores of fallers and non-fallers, 27% difference for the Mini-BESTest versus 19% for the BESTest. Also, BESTest Sections I and II had the lowest inter-rater reliability and Section II had the least variability and smallest difference between fallers and non-fallers. Removal of these sections seemed to have improved the Mini-BESTest for individuals with PD, while decreasing the time to perform the test to 10–15 minutes.

The Mini-BESTest has components of multiple individual balance assessments including items from the Berg,[34] the Dynamic Gait Index,[35] single-limb stance test, functional reach[36], timed up and go,[37] and modified Clinical Test of Sensory Interaction on Balance[38] as well as a dual-task item and postural response items. Since balance is a complex construct to measure and those with PD have multiple systems affected, studies have suggested the use of a combination of existing balance assessments to accurately identify those at risk for falls.[12,39,40] Since the Mini-BESTest has been created as a combination of tests, it might allow one score to be used as opposed to having multiple separate scores for independent tests. Future work comparing the Mini-BESTest to a battery of other tests may be warranted.

Many studies have found, as in this study, that increased disease severity based on the H&Y scale is related to increased risk of falling, yet the H&Y scale has not been found to adequately predict falls and is not sensitive to smaller changes in balance due to having only 5 categories.[8,12,14]. The Berg Balance Scale, which is similar in time requirement to the Mini-BESTest, is commonly used for individuals with PD, yet it has been shown to have a ceiling effect that,[17,23,41] based on the present study, does not seem to be present in the Mini-BESTest.

Clinical Utility of Balance Assessment Tools

It is important that, when assessing fall risk, the balance measure provides more information post-test than was available pre-test, and that it is capable of accurately identifying those at increased risk for falls.[30] The overall accuracy of the Mini-BESTest was similar to the BESTest as demonstrated by similar AUC. Sensitivity and specificity are both measures of correct identification (as a faller or non-faller) when using a specific cutoff score. The preferred cutoff values for the Mini-BESTest and BESTest, chosen by minimizing both false positives and false negatives, were 20/32 (63%) and 69%, respectively. Cutoff scores maximizing sensitivity and decreasing LR− are also reported for comparison with other

literature; however, this method greatly decreases specificity thereby increasing the false positive rate, or rate of people who are falsely identified as fallers.

A cutoff score is not a score above which no one falls and below which everyone falls; it only indicates the level of balance impairment that is associated with falling. As balance deficits increase, balance scores decrease, and there is a progressively increasing risk for falling. The pre-test probability and post-test probabilities, which can be calculated using likelihood ratios (Table 4), are useful to allow the clinician to understand the risk of falling for the individual after performing the assessment. When an individual scores 20/32 on the Mini-BESTest (LR+=4.03), for example, the probability of that individual being a faller is 65% (65 out of 100 individuals identified as at risk will be a faller). When an individual scores above a 20/32, which is considered less risk of being a faller (LR−=0.15), there is a 7% chance that the individual will be a faller. This can be compared to the pre-test probability of being a faller, which was 31% in this sample. The result of this test, either positive (≤ cutoff score) or negative (> cutoff score), could impact a clinician's decision about the needfortreatment. When the cutoff scores were chosen by maximizing sensitivity, the post-test probabilities were much less informative, with a positive test showing little change in the actually probability of an individual being a faller (Table 4). The cutoff scores for the Mini-BESTest, by maximizing both sensitivity and specificity, allowed for comparable-to-slightly better LRs and post-test probabilities than the BESTest.

To further establish the Mini-BESTest as a useful balance outcome measure, future studies should evaluate these cutoff scores prospectively to see if these balance tests are predictive of the occurrence of falls. Future studies should also assess the sensitivity of the Mini-BESTest to clinically significant changes in dynamic balance across time.

Limitations

One limitation of the study is that the Mini-BESTest was not performed separately due to time constraints. The grading criteria were used for the Mini-BESTest as the participant performed the full BESTest. Since the scores from the Mini-BESTest and the BESTest were from the same 80 individuals during the same instance of testing, future studies are necessary to evaluate the Mini-BESTest when performed completely separate to fully validate the Mini-BESTest. Another factor that might affect generalizability of these results to clinical use was the administration of the evaluation by only one rater while the other raters concurrently observed. Although the tests are standardized, the effect of a single administrator of the assessment is unknown. The raters were blinded only to fall status during testing, not disease severity, which might have allowed some bias during testing. There were also slight alterations to Sections IV and V. Although we were hesitant to alter the sections, the raters felt that it was necessary to allow two trials in Section IV so the participant would adequately lean and perform the test as intended. Section V utilized only one trial of each item due to time constraints. This alteration potentially alters the psychometric properties reported by Horak et al.[22] A retrospective fall report was also utilized, which offers the possibility of recall bias. Also, no definition of a fall was specified, so individuals may have had different criteria of what constituted a fall when giving their fall report. The risk of selection bias is slightly increased as 8 individuals were recruited through means other than the electronic database (word of mouth/PD organizations). Since this is a small portion of the total sample, the effect is likely small. Finally, a statistical limitation should be noted when interpreting ICCs. Having increased variability in the study population increases the ICC.[32] Conversely, less variability, as seen in Section II, can explain a lower ICC value.

CONCLUSION

There was a significant difference in all sections of the BESTest between fallers and non-fallers. However, there was no specific section of the BESTest that captured the primary balance deficit for individuals with PD. Although the Mini-BESTest has fewer than half of the items in the BESTest and takes only 15 minutes to complete, it is as reliable and has comparable to slightly increased discriminative properties for identify fallers.

Acknowledgments

References

1. Melton LJ 3rd, Leibson CL, Achenbach SJ, et al. Fracture risk after the diagnosis of Parkinson's disease: Influence of concomitant dementia. Mov Disord. 2006; 21(9):1361–1367. [PubMed: 16703587]

2. Bloem BR, Grimbergen YA, Cramer M, Willemsen M, Zwinderman AH. Prospective assessment of falls in Parkinson's disease. J Neurol. 2001; 248(11):950–958. [PubMed: 11757958]

3. Benatru I, Vaugoyeau M, Azulay JP. Postural disorders in Parkinson's disease. Neurophysiol Clin. 2008; 38(6):459–465. [PubMed: 19026965]

4. Boonstra TA, van der Kooij H, Munneke M, Bloem BR. Gait disorders and balance disturbances in Parkinson's disease: clinical update and pathophysiology. Curr Opin Neurol. 2008; 21(4):461–471. [PubMed: 18607208]

5. Ashburn A, Stack E, Ballinger C, Fazakarley L, Fitton C. The circumstances of falls among people with Parkinson's disease and the use of Falls Diaries to facilitate reporting. Disabil Rehabil. 2008; 30(16):1205–1212. [PubMed: 18608387]

6. Muslimovic D, Post B, Speelman JD, Schmand B, de Haan RJ. CARPA Study Group. Determinants of disability and quality of life in mild to moderate Parkinson disease. Neurology. 2008; 70(23): 2241–2247. [PubMed: 18519873]

7. Mak MK, Pang MY. Parkinsonian single fallers versus recurrent fallers: different fall characteristics and clinical features. J Neurol. 2010

8. Wood BH, Bilclough JA, Bowron A, Walker RW. Incidence and prediction of falls in Parkinson's disease: a prospective multidisciplinary study. J Neurol Neurosurg Psychiatry. 2002; 72(6):721–725. [PubMed: 12023412]

9. Allcock LM, Rowan EN, Steen IN, Wesnes K, Kenny RA, Burn DJ. Impaired attention predicts falling in Parkinson's disease. Parkinsonism Relat Disord. 2009; 15(2):110–115. [PubMed: 18487069]

10. 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(1):14–26. [PubMed: 19265767]

11. Landers MR, Backlund A, Davenport J, Fortune J, Schuerman S, Altenburger P. Postural instability in idiopathic Parkinson's disease: discriminating fallers from nonfallers based on standardized clinical measures. J Neurol Phys Ther. 2008; 32(2):56–61. [PubMed: 18645292]

12. Mak MK, Pang MY. Fear of falling is independently associated with recurrent falls in patients with Parkinson's disease: a 1-year prospective study. J Neurol. 2009

13. Mak MK, Pang MY. Balance confidence and functional mobility are independently associated with falls in people with Parkinson's disease. J Neurol. 2009; 256(5):742–749. [PubMed: 19240961]

14. Pickering RM, Grimbergen YA, Rigney U, et al. A meta-analysis of six prospective studies of falling in Parkinson's disease. Mov Disord. 2007; 22(13):1892–1900. [PubMed: 17588236]

15. Ashburn A, Fazakarley L, Ballinger C, Pickering R, McLellan LD, Fitton C. A randomised controlled trial of a home based exercise programme to reduce the risk of falling among people with Parkinson's disease. J Neurol Neurosurg Psychiatry. 2007; 78(7):678–684. [PubMed: 17119004]

16. Dibble LE, Lange M. Predicting fallsin individuals with Parkinson disease: a reconsideration of clinical balance measures. J Neurol Phys Ther. 2006; 30(2):60–67. [PubMed: 16796770]

17. Steffen T, Seney M. Test-retest reliability and minimal detectable change on balance and ambulation tests, the 36-item short-form health. Phys Ther. 2008; 88(6):733–746. [PubMed: 18356292]

18. Franzén E, Paquette C, Gurfinkel VS, Cordo PJ, Nutt JG, Horak FB. Reduced performance in balance, walking and turning tasks is associated with increased neck tone in Parkinson's disease. Exp Neurol. 2009; 219(2):430–438. [PubMed: 19573528]

19. Ashburn A, Stack E, Pickering RM, Ward CD. Predicting fallers in a community-based sample of people with Parkinson's disease. Gerontology. 2001; 47(5):277–281. [PubMed: 11490147]

20. Balash Y, Peretz C, Leibovich G, Herman T, Hausdorff JM, Giladi N. Falls in outpatients with Parkinson's disease: frequency, impact and identifying factors. J Neurol. 2005; 252(11):1310–1315. [PubMed: 15895303]

21. Kerr GK, Worringham CJ, Cole MH, Lacherez PF, Wood JM, Silburn PA. Predictors of future falls in Parkinson disease. Neurology. 2010; 75(2):116–124. [PubMed: 20574039]

22. Horak FB, Wrisley DM, Frank J. The Balance Evaluation Systems Test (BESTest) to differentiate balance deficits. Phys Ther. 2009; 89(5):484–498. [PubMed: 19329772]

23. Leddy AL, Crowner BE, Earhart GM. Functional gait assessment and balance evaluation systemtest: reliability, validity, sensitivity, and specificity for identifying individuals with Parkinson disease who fall. Phys Ther. 2011; 91(1):102–113. [PubMed: 21071506]

24. Franchignoni F, Horak F, Godi M, Nardone A, Giordano A. Using psychometric techniques to improve the Balance Evaluation Systems Test: the mini-BESTest. J Rehabil Med. 2010; 42(4): 323–331. [PubMed: 20461334]

25. Leddy AL, Crowner BE, Earhart GM. Functional Gait Assessment and Balance Evaluation System Test: reliability, validity, and sensitivity and specificity for identifying individuals with Parkinson disease who fall. Phys Ther.

26. Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. Mov Disord. 2008; 23(15):2129–2170. [PubMed: 19025984]

27. 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(9):1020–1028. [PubMed: 15372591]

28. [Accessed March 2, 2011] Balance Evaluation Systems Test (BESTest). Web site. Available at: http://www.bestest.us/purchasing.html

29. Akobeng AK. Understanding diagnostic tests 2: likelihood ratios, pre-and post-test probabilities and their use in clinical practice. Acta Paediatrica. 2007; 96(4):487–491. [PubMed: 17306009]

30. Deeks JJ. Using evaluations of diagnostic tests: understanding their limitations and making the most of available evidence. Ann Oncol. 1999; 10(7):761–768. [PubMed: 10470421]

31. Morris ME, Martin CL, Schenkman ML. Striding out with Parkinson disease: evidence-based physical therapy for gait disorders. Phys Ther. 2010; 90(2):280–288. [PubMed: 20022998]

32. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005; 19(1):231–240. [PubMed: 15705040]

33. Robinson K, Dennison A, Roalf D, et al. Falling risk factors in Parkinson's disease. NeuroRehabilitation. 2005; 20(3):169–182. [PubMed: 16340098]

34. Berg KO, Wood-Dauphinee SL, Williams JI, Maki B. Measuring balancein the elderly: validation of an instrument. Can J Public Health. 1992; 83 (Suppl 2):S7–11. [PubMed: 1468055]

35. Wrisley DM, Walker ML, Echternach JL, Strasnick B. Reliability of the dynamic gait index in people with vestibular disorders. Arch Phys Med Rehabil. 2003; 84(10):1528–1533. [PubMed: 14586922]

36. Duncan PW, Weiner DK, Chandler J, Studenski S. Functional reach: a new clinical measure of balance. J Gerontol. 1990; 45(6):M192–7. [PubMed: 2229941]

37. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc. 1991; 39(2):142–148. [PubMed: 1991946]

38. Shumway-Cook A, Horak FB. Assessing the influence of sensory interaction of balance. Suggestion from the field. Phys Ther. 1986; 66(10):1548–1550. [PubMed: 3763708]

39. Lim LI, van Wegen EE, de Goede CJ, et al. Measuring gait and gait-related activities in Parkinson's patients own home environment: a reliability, responsiveness and feasibility study. Parkinsonism Relat Disord. 2005; 11(1):19–24. [PubMed: 15619458]

40. Dibble LE, Christensen J, Ballard DJ, Foreman KB. Diagnosis of fall risk in Parkinson disease: an analysis of individual and collective clinical balance test interpretation. Phys Ther. 2008; 88(3): 323–332. [PubMed: 18187494]

41. Tanji H, Gruber-Baldini AL, Anderson KE, et al. A comparative study of physical performance measures in Parkinson's disease. Mov Disord. 2008; 23(13):1897–1905. [PubMed: 18709681]
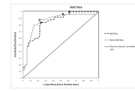
Figure 1.
Receiver operating characteristic plot of the Mini-BESTest and the BESTest to compare overall accuracy of identifying who is a faller and who is a non-faller. The boxes represent cutoff scores chosen by maximizing both sensitivity and specificity. X's represent cutoff scores chosen by maximizing sensitivity and minimizing LR-.

Table 1

Comparison of overall subjects, fallers, and non-fallers.

	Overall (n=80) Mean ± SD	Fallers (n=25) Mean ± SD	Non-Fallers (n=55) Mean ± SD
% male	59%	64%	56%
Age	68.2 ± 9.3	68.8 ± 7.8	67.9 ± 10.0
Disease duration	8.5 ±.54	11.4 ± 5.5[*]	7.15 ± 3.81
MDS-UPDRS score	72.6 ± 25.1	93.8 ± 23.1[*]	62.9 ± 19.0
H&Y stage-mean (SD)/median	2.45 ± 0.64/2.5	2.9 ± 0.71/3.0[*]	2.3 ± 0.50/2.0
Stage 1	4	1	3
Stage 2	27	1	26
Stage 2.5	30	10	20
Stage 3	13	8	5
Stage 4	6	5	1
Mini-BESTest	20.2 ± 7.0	14.3 ± 6.2[*]	22.9 ± 5.5
Mini-BESTest %	63.2 ± 21.8	44.8 ± 19.4[*]	71.6 ± 17.3
BESTest Total %	70.3 ± 16.7	57.1 ± 15.4 [*]	76.4 ± 13.6
Section I: Biomechanical Constraints %	59.2 ± 22.3	45.3 ± 19.4[*]	65.6 ± 20.7
Section II: Stability Limits/Verticality %	80.1 ± 12.6	73.9 ± 11.8[†]	82.9 ± 12.1
Section III : Anticipatory Postural Adjustments %	69.5 ± 20.4	57.6 ± 18.4[*]	74.9 ± 18.9
Section IV: Postural Responses %	68.5 ± 22.9	52.2 ± 25.6[*]	75.9 ± 18.0
Section V: Sensory Orientation %	75.3 ± 22.4	58.9 ± 24.9[*]	82.7 ± 16.8
Section VI: Stability in Gait %	68.0 ± 21.5	51.4 ± 20.6[*]	75.5 ± 17.4

Statistically significant difference between fallers and non-fallers

[*] (p<0.001) and

[†](p<0.004).

Table 2

Inter-rater reliability

n=15	ICC (2,1)	95% CI
Mini-BESTest	0.91	0.75, 0.97
BESTest Total	0.96	0.89, 0.99
Section I: Biomechanical Constraints	0.81	0.61, 0.92
Section II: Stability Limits/Verticality	0.79	0.58, 0.92
Section III: Anticipatory Postural Adjustments	0.91	0.81, 0.97
Section IV: Postural Responses	0.91	0.81, 0.97
Section V: Sensory Orientation	0.96	0.91, 0.99
Section VI: Stability in Gait	0.86	0.62, 0.95

ICC=Intraclass correlation, CI= confidence interval

*All p-values <0.01

Table 3

Test-Retest reliability

n=24	ICC (2,1)	95% CI	p-value
Mini-BESTest (n=23)	0.92	0.82, 0.96	*p<0.001
BESTest Total (n=23)	0.88	0.72, 0.95	*p<0.001
Section I: Biomechanical Constraints	0.69	0.41, 0.85	0.075
Section II: Stability Limits/Verticality	0.63	0.31, 0.82	0.180
Section III: Anticipatory Postural Adjustments	0.83	0.45, 0.94	*0.018
Section IV: Postural Responses	0.87	0.68, 0.95	*<0.001
Section V: Sensory Orientation (n=23)	0.72	0.44, 0.87	0.053
Section VI: Stability in Gait	0.72	0.45, 0.87	*0.047

ICC=Intraclass correlation, CI= confidence interval

Table 4

Discriminative properties

Cutoff scores, sensitivity, specificity, likelihood ratios, and post-test probability of being a faller

	Cutoff Score	Sensitivity	Specificity	LR+ (95% CI)	LR− (95%CI)	Post-test probability with test ≤ cutoff value	Post-test probability with test > cutoff value
Mini-BESTest	≤20/32* (63%)	0.88	0.78	4.03 (2.40–6.79)	0.15 (0.05–0.45)	64.7%	6.5%
	≤23/32 (72%)	0.96	0.47	1.82 (1.40–2.37)	0.08 (0.01–0.59)	45.3%	3.7%
BESTest	≤69%*	0.84	0.76	3.49 (2.11–5.77)	0.21 (0.09–.52)	61.3%	8.7%
	≤ 84%	1.00	0.39	1.64 (1.32–2.02)	0.00 (Unable to calculate)	42.7%	0.0%

Pre-test probability for being a faller was 31.3%. The first cutoff value was chosen to *maximize both sensitivity and specificity.

The second cutoff value was chosen by maximizing sensitivity and LR−.

LR+=positive likelihood ratio; LR−=negative likelihood ratio