3-2013

Four square step test performance in people with Parkinson disease

Ryan P. Duncan
Washington University School of Medicine in St. Louis

Gammon M. Earhart
Washington University School of Medicine in St. Louis

Four Square Step Test Performance in People with Parkinson Disease

Ryan P. Duncan, PT, MPT[1] and Gammon M. Earhart, PhD, PT[1-3]

Running Head: 4SST in PD

[1] Washington University in St. Louis School of Medicine, Program in Physical Therapy

[2] Washington University in St. Louis School of Medicine, Department of Anatomy & Neurobiology

[3] Washington University in St. Louis School of Medicine, Department of Neurology

Corresponding Author:

Gammon M. Earhart, PhD, PT
Washington University School of Medicine
Program in Physical Therapy
Campus Box 8502
4444 Forest Park Blvd.
St. Louis, MO 63108
314-286-1425
earhartg@wusm.wustl.edu

Abstract word count: 248
Text word count: 3,315
Conflicts of Interest and Sources of Funding: No conflicts of interest declared for either author. This work was supported by grants from the Parkinson's Disease Foundation and the Davis Phinney Foundation.

A portion of this work was presented as a platform presentation at the APTA – Combined Sections Meeting 2012 in Chicago, IL.

Abstract

Background and Purpose: The Four Square Step Test (4SST), a quick and simple test of multi-directional stepping, may be useful in predicting falls in people with Parkinson disease (PD). We studied the reliability of the 4SST and its ability to discriminate between freezers and non-freezers, fallers and non-fallers, and factors predictive of 4SST performance in people with PD.

Methods: Fifty-three individuals with idiopathic PD completed the full protocol including the 4SST as well as measures of balance, walking, and disease severity on anti-PD medication.

Results: Inter-rater (ICC = .99) and test-retest reliability were high (ICC = .78). The median 4SST performance was 9.52 seconds. There was a significant difference between 4SST time on medication versus off (p=.03), while differences between fallers/non-fallers (p=.06) and freezers/non-freezers (p=.08) did not reach significance. All outcome measures were significantly related to 4SST time. In an exploratory, simultaneous regression analysis 56% of the variance in 4SST performance could be accounted for by three measures: Mini-BESTest, Five Time Sit to Stand, and Nine Hole Peg Test. The 4SST cutoff score for distinguishing fallers from non-fallers was 9.68 seconds (AUC = .65, sensitivity = .73, specificity = .57). The post-test probability of an individual with a score above the cutoff being a faller was 31% (pre-test probability = 21%).

Discussion and Conclusion: The 4SST is a reliable, quick test that can distinguish between on and off medication conditions in PD, but is not as good as other tests (e.g. Mini-BESTest) for distinguishing between fallers and non-fallers.

Key words: Four Square Step Test, Parkinson disease, fall risk, balance

Introduction

The progressive nature of Parkinson disease (PD) leads to significant movement-related impairments.[1] Gait difficulty, including freezing of gait, and postural instability are common among people with PD and are highly associated with falls.[2,3] Falls occur in up to 70% of individuals with PD, with potential sequelae including mobility restrictions, fractures, and mortality.[4-7] Injuries related to falls significantly increase medical costs for those with PD.[8] Given the devastating effects of falls, it is critical that individuals at risk for falls be accurately identified so that they can receive interventions directed at improving gait and balance and reducing fall risk.

Previous research shows that the best predictor of falls is a prior history of falls,[9] but this relies on accurate self-report of past events. It is preferable for clinicians to be able to directly measure a person's performance in order to evaluate fall risk. Recent work suggests that some clinical balance tests such as the Balance Evaluation Systems Test (BESTest) and Mini-BESTest may be useful for identifying those who will fall in the next six months.[10] However, these and other clinical balance tests are often time consuming and therefore may be impractical to use on a regular basis. As such, investigation of the clinical utility and predictive value of quick and easy-to-administer clinical tests is warranted.

The Four Square Step Test (4SST), first described by Dite and Temple, is a quick and simple clinical test that requires an individual to rapidly step over obstacles in the forward, backward, and sideways directions.[11] We hypothesized that the 4SST may be a useful balance screen in people with PD for the following reasons: a) it demands complex anticipatory postural adjustments for gait initiation, known to be impaired in people with PD,[12-15] b) the requirements

for stepping backward and laterally may be particularly challenging for people with PD,[16,17] and c) the task requires execution of a complex multi-step movement sequence, another area of known difficulty in PD.[18-20] To our knowledge, no study to date has evaluated 4SST performance in people with PD. However, 4SST performance has been studied in older adult fallers and non-fallers, individuals with stroke, and individuals with a vestibular disorder.[11,21,22]

The purpose of this study was to evaluate 4SST performance in individuals with PD. We aimed to:1) establish inter-rater and test-retest reliability of the 4SST in PD, 2) determine the effectiveness of the 4SST in distinguishing between on and off anti-PD medication conditions, fallers and non-fallers, and freezers and non-freezers, and 3) describe factors related to and predictive of 4SST performance in PD. We hypothesized that the 4SST would possess high inter-rater and test-retest reliability, correctly distinguish between on and off anti-PD medication conditions, fallers and non-fallers, and freezers and non-freezers, and be related to balance, bradykinesia, and freezing of gait in people with PD.

Methods

Participants

Individuals over age forty with idiopathic PD who were already participating in two separate, ongoing studies were consecutively recruited.[23,24] Participants included in the study must have been diagnosed with "definite" PD, as defined by Racette and colleagues.[25] Potential participants were screened and excluded if found to have any of the following: 1) a serious medical condition, 2) history or presence of a neurological condition other than PD, 3) previous surgical management of PD (e.g. deep brain stimulation), or 4) a musculoskeletal injury that

considerably inhibited movement. All participants provided written informed consent to participate in the protocol as approved by the Human Research Protection Office of XXXXXX.

Fifty-three individuals (58% male) with idiopathic PD were included in the study. Each participant reported fall history over the past six months and based on this was classified as a faller if he reported more than one fall in that period of time. Participants reporting no falls or only one fall in the previous six months were classified as non-fallers. Participants also completed the Freezing of Gait Questionnaire (FOG-Q), a self-assessment of walking performance during their self-perceived worst state, how gait impairments impact functional independence, whether or not freezing is experienced, and characteristics of freezing episodes.[26] A participant was classified as a freezer if he or she gave a rating greater than one on item three of the FOG-Q, indicating FOG episodes occurring at least once per week. One participant was excluded due to an inability to step in the backward direction secondary to freezing. Demographic information is presented in Table 1. Twenty-eight participants (54% male), a subgroup of the original sample, were chosen to complete the 4SST on and off anti-PD medication to determine if the 4SST possessed the ability to distinguish between medication conditions.

INSERT TABLE 1 HERE

Outcome Measures

4SST

Performance of this test was carried out as originally described by Dite and Temple.[11] Meter sticks, set up to appear like a crosshair, were used as the obstacles over which participants would step (Figure 1). These meter sticks were not fixed to the surface. Participants faced the

same direction when stepping to the different squares, and both feet had to touch each square before the next movement. Stepping was completed in the following sequence: forward, right, backward, left, right, forward, left, backward. Timing began when the first foot hit the second square and stopped when the last foot returned to the final square. Successful completion of a trial was a trial in which the meter sticks were not touched during performance of the test. The 4SST has been shown to have high inter-rater and test-retest reliability among the elderly and individuals with vestibular disorders.[11,22]

Movement Disorders Society - Unified Parkinson Disease Rating Scale – III (MDS-UPRDS-III)

Motor symptom severity was assessed using the MDS-UPDRS-III.[27] From this assessment, we also determined the H&Y stage of each participant. The MDS-UPDRS-III was administered by a physical therapist (RPD) who was trained using the official MDS-UPDRS training video.[28]

Mini-Balance Evaluations Systems Test (Mini-BESTest)

The Mini-BESTest is an instrument used to assess balance through evaluation of postural stability during performance of 14 dynamic tasks.[29] Investigators have previously demonstrated high inter-rater and test-retest reliability when using the Mini-BESTest to evaluate balance in individuals with PD.[30]

Five Times Sit to Stand (FTSTS) Test

Because there was potential that the 4SST and FTSTS test may be measuring similar constructs, the FTSTS was included in the assessment battery. The FTSTS test measures how quickly one can go from sitting to standing five consecutive times without using the upper

extremities. Balance and bradykinesia are predictive of FTSTS performance in PD.[31] This test has also been shown to have high inter-rater and test-retest reliability in PD.[31]

Six Minute Walk Test (6MWT)

Endurance was measured using the 6MWT, which has been shown to be highly reliable when assessing people with PD.[32] Participants were asked to cover as much ground as they could in six minutes while walking, and if necessary, assistive devices were used. Distance covered was measured to the nearest meter.

Nine Hole Peg Test (9HPT)

The 9HPT was used as a measure of bradykinesia.[33,34] Participants were instructed to place nine pegs into nine holes, one at a time, as quickly as possible. The non-dominant hand was used to stabilize the peg board if necessary. Two trials were completed using the dominant hand, and the mean of these two trials was calculated. This test is highly reliable when examining upper extremity function in people with PD.[33]

Procedures

All outcome measures were assessed while participants were on anti-PD medication. In addition, a subset (n=28) of this group who were those recruited for one of the two studies mentioned previously was also tested while off anti-PD medication. Participants were considered off medication if the last medication dose was administered greater than 12 hours prior to evaluation. A physical therapist instructed participants in correct performance of the 4SST. Following instruction, one untimed practice trial and two timed trials of the 4SST were completed. Two raters timed the trials. Inter-rater reliability was determined comparing the times

recorded by the two raters for the first trial performed, while test-retest reliability was determined comparing the times of the first and second trials collected by the primary rater. Because it was our aim to determine if 4SST performance could be distinguished between off and on medication states, only 4SST times were obtained in the off medication state. After completing off medication testing, participants were allowed to take their normal dose of anti-PD medication. For participants on medication, we administered the measures in the following order: MDS-UPDRS-III, 4SST, Mini-BESTest, FTSTS test, 9HPT, FOG-Q, and 6MWT. All outcome measures were administered by the same rater.

Data Analysis

Descriptive statistics were calculated using only the first trial of the 4SST because this was deemed most clinically relevant by the authors. Intraclass correlation coefficients were used to describe inter-rater (ICC 1,1) and test-retest (ICC 2,1) reliability of the 4SST. For all analyses beyond reliability and the off/on medication comparison, only the first timed trial of the 4SST performed on medication was used. Because 4SST times were not normally distributed we used nonparametric statistics to examine differences between conditions or groups. To determine differences between off and on medication 4SST times, we used a Wilcoxon Signed Rank test. To determine if there were differences in 4SST performance between fallers and non-fallers and freezers and non-freezers we used Mann Whitney U tests. Spearman correlation coefficients were used to describe relationships between the 4SST and all other outcome measures. Statistical significance for analyses was set at $\alpha \leq .05$. Outcome measures that were significantly correlated with the 4SST were entered into an exploratory, simultaneous regression model to determine factors predictive of 4SST performance. In cases of multicolinearity, the variable with the highest correlation with the 4SST was retained and the collinear variable was removed from the

analysis. In the final model, only those factors that significantly contributed ($p<0.05$) were retained. Finally, a cutoff score for 4SST performance in fallers versus non-fallers was determined using ROC curves. This cutoff score was chosen based on the minimum value of: $(1\text{-sensitivity})^2 + (1\text{-specificity})^2$.[35] From this cutoff score, positive (LR+) and negative (LR-) likelihood ratios, pre-test and post-test probabilities were calculated. The ROC curve of the 4SST was compared to that of the Mini-BESTest for this sample. Data analysis was conducted using NCSS software version 7.1.19 (NCSS Software, Kaysville, UT).

Results

Inter-rater reliability, tested only on medication, was high (ICC = .99). Test-retest reliability was high when testing participants both off (ICC = .90) and on (ICC = .78) medication.

The median 4SST performance time on medication was 9.52 (95% LCL-UCL: 9.03 – 10.54) seconds (Table 2). Figure 2 shows the frequency distribution for 4SST scores on medication for the full sample, with 4SST times binned into one second intervals. 4SST performance was significantly different off compared to on medication ($p = .03$). There were no significant differences in 4SST performance between freezers and non-freezers ($p = .08$) and between fallers and non-fallers ($p = .06$) (Table 2). All outcome measures were significantly related to 4SST performance on medication (Table 3), most notably the Mini-BESTest, MDS-UPDRS-III, age, 9HPT, and FTSTS.

INSERT FIGURE 1 HERE

INSERT TABLE 2 HERE

INSERT TABLE 3 HERE

Age, MDS-UPDRS-III, Mini-BESTest, FTSTS, 9HPT, FOGQ, and 6MWT were all entered into a simultaneous regression model, as all of these factors were significantly correlated with 4SST performance. However, due to issues with multicolinearity with the Mini-BESTest, age and MDS-UPDRS-III were removed from the model. Of the remaining five variables entered, only three contributed significantly to the model. Thus, the final model included only the Mini-BESTest, 9HPT and FTSTS as shown in Table 4. This model explained 56% of the variance in 4SST performance.

INSERT TABLE 4 HERE

Figure 3 displays the ROC curves of the 4SST and the Mini-BESTest for discriminating between fallers and non-fallers. The AUC for the 4SST was 0.65 (95% CI: 0.43-0.80) compared to an AUC of 0.80 for the Mini-BESTest. Maximizing sensitivity and specificity, the cutoff score for the 4SST was determined to be 9.68 seconds (sensitivity = .73, specificity = .57). Forty-seven percent of our sample scored above this cutoff. The positive and negative likelihood ratios for those scoring above the cutoff were 1.7 (95% CI: 1.00-2.73) and 0.48 (95% CI: 0.17-1.27), respectively. The post-test probability of a fall for those with scores above the cutoff was 31%, compared to a 21% pre-test probability.

INSERT FIGURE 2 HERE

Discussion

To our knowledge, this is the first study to describe 4SST performance in people with PD. According to our results, there is high inter-rater and test-retest reliability for the 4SST when

testing individuals with PD on and off anti-PD medication. Secondly, people with PD perform the 4SST more slowly off medication than they do when on medication; however, there was no difference in 4SST performance for freezers/non-freezers and fallers/non-fallers. Finally, all measures employed were significantly correlated with 4SST performance. Measures reflective of balance and bradykinesia combined to explain 56% of the variance in 4SST performance in people with PD.

Inter-rater reliability (ICC = .99), when testing individuals with PD on anti-PD medication, matches that noted by Dite and Temple in the original description of the 4SST.[11] The test-retest reliability when testing people on medication (ICC = .78) was lower than expected. While the designs for assessing test-retest reliability were similar, Whitney and colleagues reported much higher reliability (ICC = .93) when testing individuals with a vestibular disorder. In our study, it is possible that a practice effect could have led to consecutive improvements in 4SST performance times from the practice trial and then to the first and second trials, lowering agreement between 4SST times. Because we were limited in our ability to have the participants return for a follow-up visit, we could not capture test-retest reliability with more than one or two minutes between trials.

On medication, the median time taken by our full sample to complete the 4SST was 9.52 seconds, which is faster than that of healthy older individuals who had at least one fall in the previous six months, individuals post-stroke, and individuals with a vestibular disorder.[11,21,22] It is particularly interesting that on average our group with PD was faster than healthy older individuals who had at least one fall in the previous six months. This difference could be related to the fact that this sample with a history of at least one fall was heterogeneous due to inclusion of individuals with varied histories such as stroke, cardiovascular disorders, and neurologic

disorders.[11] 4SST performance in PD may also be enhanced by the visual cues provided by the apparatus used for the test. Investigators have shown that visual cues improve gait initiation in the forward direction.[36] We speculate that these cues may also translate to improved stepping in the backward and lateral directions and as such the grid used for the 4SST may provide visual cues that enhance performance of stepping during the test. However, it is important to note that Dite and Temple reported that a group of 27 healthy older adults, a homogenous sample without co-morbidities or history of falls, had a mean 4SST time of 8.70 seconds.[11] This suggests that while the meter sticks may have improved performance for those with PD, there is still evidence of impairment relative to healthy controls.

Foreman and colleagues stated that there is a need for clinical balance tests that are "responsive to changes in performance within individuals (medication status) and responsive to differences between individuals (e.g. fallers versus non-fallers) (p169)."[37] We studied the responsiveness of the 4SST to changes in medication status in those with PD, in addition to the responsiveness to differences between freezers and non-freezers and fallers and non-fallers with PD. The results of this study also indicate that people with PD demonstrate different 4SST performance when off anti-PD medication compared to on. We found improved 4SST performance while on medication, a finding in accordance with Morris and colleagues who found improved Timed Up and Go performance on medication versus off.[38] In our sample, the 4SST did not distinguish between fallers and non-fallers and freezers or non-freezers. This may be related to our relatively small sample sizes of fallers and freezers which reduces our power to detect differences.

Our results demonstrated that all measures employed in this study were significantly related to 4SST performance in people with PD. Most notably, the Mini-BESTest score

demonstrated a significant moderate correlation to 4SST performance. This is not surprising as good balance is necessary to complete the multi-directional stepping test in a timely fashion. The other correlation worth noting is the significant relationship between the 9HPT and 4SST. Earhart and colleagues noted that bradykinesia was a significant predictor of 9HPT performance time in people with PD.[33] Due to the relationship between the 9HPT and 4SST, we suspect that bradykinesia might also be influencing 4SST performance in this population. We also noted that the Mini-BESTest, 9HPT and FTSTS combined to explain 56% of the variance in 4SST performance. This is very similar to results reported by Duncan et al., who studied factors predictive of FTSTS performance in people with PD.[31] Because both the 4SST and FTSTS are timed tests of mobility, the results noted by Duncan et al. and those reported in this study suggest that balance and bradykinesia play significant roles in timed tests of mobility in people with PD.

Regarding the detection of fallers, the 4SST demonstrated moderate levels of sensitivity (.78) and specificity (.57) with a cutoff time of 9.68 seconds. A 4SST time above the cutoff does not immediately suggest that the person with PD is at significant risk for falls. This notion is supported by the fact that when an individual with PD completes the 4SST in more than 9.68 seconds (LR+ = 1.7), the post-test probability of being a faller is 31% compared to a pre-test probability of 21%. A 4SST completion time less than the cutoff (LR- = .48) yields a post-test probability of 11%. Based on our analysis, it is evident that the Mini-BESTest outperforms the 4SST in identifying fallers and non-fallers with PD. Leddy and colleagues demonstrated that the Mini-BESTest (sensitivity = .88, specificity = .78) and BESTest (sensitivity = .84, specificity = .76) had better predictive ability than that of the 4SST in the present study.[30] These findings were echoed in a prospective study of fall risk in PD in which the Mini-BESTest (sensitivity = .86, specificity = .78) and BESTest (sensitivity = .93, specificity = .84) outperformed other balance

measures when identifying fallers six months after assessment.[10] While the 4SST is similar in its predictive abilities to other quick tests of motor function, the 4SST should not replace more extensive balance testing when attempting to determine fall risk in people with PD.[31,37]

The results of the present study should be interpreted in light of several limitations. First, this pilot study included a small sample size of people with mild to moderate PD among which there were small numbers of fallers and freezers. The results demonstrating that the fallers and freezers performed the 4SST more slowly than non-fallers and non-freezer, respectively, should be interpreted with caution. Second, all analyses other than test-retest reliability utilized our analyses utilized only one timed trial of the 4SST. This was done in order to reflect the clinical practice of rehabilitation professionals where it is likely that only one trial of the 4SST would be collected. However, somewhat different results may have been obtained if the results of several trials had been averaged. Finally, we measured test-retest reliability with only minutes between 4SST trials and recommend that investigators measure test-retest reliability of the 4SST in people with PD with perhaps days or a week between administrations of the test. In the future, investigators should study the 4SST's responsiveness to change over time or after intervention and may also consider examining the combined ability of the 4SST and other quick tests of motor function in predicting who will fall with PD.

Conclusion

The 4SST, when used in people with PD, is able to distinguish between off and on anti-PD medication states, but not between freezers and non-freezers and fallers and non-fallers. Mini-BESTest, 9HPT, and FTSTS scores were most predictive of 4SST performance. Because of

its limited ability in accurately predicting falls in those with PD, we do not recommend use of the 4SST in lieu of other balance measures such as the Mini-BESTest.

Acknowledgements

This work was supported by grants from the Parkinson's Disease Foundation and the Davis Phinney Foundation. The authors thank XXXXX for her assistance with timing of the 4SST.

References

1. Speelman AD, van de Warrenburg BP, van Nimwegen M, Petzinger GM, Munneke M, Bloem BR. How might physical activity benefit patients with Parkinson disease? *Nat Rev Neurol*. 2011;7(9):528-34.

2. Bloem BR, Grimbergen YA, Cramer M, Willemsen M, Zwinderman AH. Prospective assessment of falls in Parkinson's disease. *J Neurol*. 2001;248(11):950-8.

3. Bloem BR, Hausdorff JM, Visser JE, Giladi N. Falls and freezing of gait in Parkinson's disease: a review of two interconnected, episodic phenomena. *Mov Disord*. 2004;19(8):871-84.

4. Bloem BR, Steijns JA, Smits-Engelsman BC. An update on falls. *Curr Opin Neurol*. 2003;16(1):15-26.

5. Snijders AH, Nonnekes J, Bloem BR. Recent advances in the assessment and treatment of falls in Parkinson's disease. *F1000 Med Rep*. 2010;2:76.

6. Sato Y, Kaji M, Tsuru T, Oizumi K. Risk factors for hip fracture among elderly patients with Parkinson's disease. *J Neurol Sci*. 2001;182(2):89-93.

7. Wenning GK, Ebersbach G, Verny M, Chaudhuri KR, Jellinger K, McKee A, et al. Progression of falls in postmortem-confirmed parkinsonian disorders. *Mov Disord*. 1999;14(6):947-50.

8. Pressley JC, Louis ED, Tang MX, Cote L, Cohen PD, Glied S, et al. The impact of comorbid disease and injuries on resource use and expenditures in parkinsonism. *Neurology*. 2003;60(1):87-93.

9. Pickering RM, Grimbergen YA, Rigney U, Ashburn A, Mazibrada G, Wood B, et al. A meta-analysis of six prospective studies of falling in Parkinson's disease. *Mov Disord*. 2007;22(13):1892-900.

10. Duncan RP, Leddy AL, Cavanaugh JT, Dibble LE, Ellis TD, Ford MP, et al. Accuracy of fall prediction in Parkinson disease: six-month and 12-month prospective analyses. *Parkinsons Dis*. 2012;2012:237673.

11. Dite W, Temple VA. A clinical test of stepping and change of direction to identify multiple falling older adults. *Arch Phys Med Rehabil*. 2002;83(11):1566-71.

12. Burleigh-Jacobs A, Horak FB, Nutt JG, Obeso JA. Step initiation in Parkinson's disease: influence of levodopa and external sensory triggers. *Mov Disord*. 1997;12(2):206-15.

13. Hass CJ, Waddell DE, Fleming RP, Juncos JL, Gregor RJ. Gait initiation and dynamic balance control in Parkinson's disease. *Arch Phys Med Rehabil*. 2005;86(11):2172-6.

14. Rocchi L, Chiari L, Mancini M, Carlson-Kuhta P, Gross A, Horak FB. Step initiation in Parkinson's disease: influence of initial stance conditions. *Neurosci Lett*. 2006;406(1-2):128-32.

15. Rosin R, Topka H, Dichgans J. Gait initiation in Parkinson's disease. *Mov Disord*. 1997;12(5):682-90.

16. Gantchev N, Viallet F, Aurenty R, Massion J. Forward versus backward oriented stepping movements in Parkinsonian patients. *Motor Control*. 2000;4(4):453-68.

17. King LA, Horak FB. Lateral stepping for postural correction in Parkinson's disease. *Arch Phys Med Rehabil*. 2008;89(3):492-9.

18. Benecke R, Rothwell JC, Dick JP, Day BL, Marsden CD. Disturbance of sequential movements in patients with Parkinson's disease. *Brain*. 1987;110 (Pt 2):361-79.

19. Weiss P, Stelmach GE, Hefter H. Programming of a movement sequence in Parkinson's disease. *Brain*. 1997;120 (Pt 1):91-102.

20. Yaguez L, Lange HW, Homberg V. Differential effect of Huntington's and Parkinson's diseases in programming motor sequences of varied lengths. *J Neurol*. 2006;253(2):186-93.

21. Blennerhassett JM, Jayalath VM. The Four Square Step Test is a feasible and valid clinical test of dynamic standing balance for use in ambulant people poststroke. *Arch Phys Med Rehabil*. 2008;89(11):2156-61.

22. Whitney SL, Marchetti GF, Morris LO, Sparto PJ. The reliability and validity of the Four Square Step Test for people with balance deficits secondary to a vestibular disorder. *Arch Phys Med Rehabil*. 2007;88(1):99-104.

23. Dibble LE, Cavanaugh JT, Earhart GM, Ellis TD, Ford MP, Foreman KB. Charting the progression of disability in Parkinson disease: study protocol for a prospective longitudinal cohort study. *BMC Neurol*. 2010;10:110.

24. Duncan RP, Earhart GM. Randomized Controlled Trial of Community-Based Dancing to Modify Disease Progression in Parkinson Disease. *Neurorehabil Neural Repair*. 2012;26(2):132-43.

25. Racette BA, Rundle M, Parsian A, Perlmutter JS. Evaluation of a screening questionnaire for genetic studies of Parkinson's disease. *Am J Med Genet*. 1999;88(5):539-43.

26. Giladi N, Tal J, Azulay T, Rascol O, Brooks DJ, Melamed E, et al. Validation of the freezing of gait questionnaire in patients with Parkinson's disease. *Mov Disord*. 2009;24(5):655-61.

27. Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, 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):2129-70.

28. Goetz CG, Stebbins, G.T., Chmura, S., Fahn, S., Klawans, H., Marsden, C.D. Unified Parkinson's Disease Rating Scale (UPDRS) Training Program. 2012 [June 11, 2012]; Available from: http://www.movementdisorders.org/updrs/.

29. 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):323-31.

30. Leddy AL, Crowner BE, Earhart GM. Utility of the Mini-BESTest, BESTest, and BESTest sections for balance assessments in individuals with Parkinson disease. *J Neurol Phys Ther*. 2011;35(2):90-7.

31. Duncan RP, Leddy AL, Earhart GM. Five times sit-to-stand test performance in Parkinson's disease. *Arch Phys Med Rehabil*. 2011;92(9):1431-6.

32. Steffen T, Seney M. Test-retest reliability and minimal detectable change on balance and ambulation tests, the 36-item short-form health survey, and the unified Parkinson disease rating scale in people with parkinsonism. *Phys Ther*. 2008;88(6):733-46.

33. Earhart GM, Cavanaugh JT, Ellis T, Ford MP, Foreman KB, Dibble L. The 9-hole PEG test of upper extremity function: average values, test-retest reliability, and factors contributing to performance in people with Parkinson disease. *J Neurol Phys Ther*. 2011;35(4):157-63.

34. Oxford Grice K, Vogel KA, Le V, Mitchell A, Muniz S, Vollmer MA. Adult norms for a commercially available Nine Hole Peg Test for finger dexterity. *Am J Occup Ther*. 2003;57(5):570-3.

35. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol*. 2006;163(7):670-5.

36. Jiang Y, Norman KE. Effects of visual and auditory cues on gait initiation in people with Parkinson's disease. *Clin Rehabil*. 2006;20(1):36-45.

37. Foreman KB, Addison O, Kim HS, Dibble LE. Testing balance and fall risk in persons with Parkinson disease, an argument for ecologically valid testing. *Parkinsonism Relat Disord*. 2011;17(3):166-71.

38. Morris S, Morris ME, Iansek R. Reliability of measurements obtained with the Timed "Up & Go" test in people with Parkinson disease. *Phys Ther*. 2001;81(2):810-8.

Table 1. Demographic information of participants.

Table 1. Demographic Information	
Off Medication Group (n=28)	
Age (yrs)	70 ± 7.4
Gender	M:15, F:13
Hoehn & Yahr Stage	I (0), II (8), II.5 (15), III (5), IV (0)
Fallers	3
Freezers	5
On Medication Group (n=53)	
Age (yrs)	68 ± 8.5
Gender	M:31, F:22
Hoehn & Yahr Stage	I (2), II (28), II.5 (16), III (2), IV (5)
Fallers	11
Freezers	17

Table 2. 4SST performance times (group comparisons)

Table 2. 4SST Performance Times (Group Comparisons)						
	10th Percentile	25th Percentile	50th Percentile (95% LCL-UCL)	75th Percentile	90th Percentile	
On Medication (n=28)	7.37	8.56	9.60 (8.73-10.62)	11.60	17.07	P = .03
Off Medication (n=28)	7.60	9.33	11.02 (9.42-12.56)	13.28	16.87	
Freezer (n=17)	7.68	8.73	10.39 (8.78-13.06)	14.18	17.36	P = .08
Non-Freezer (n=36)	7.21	8.37	9.22 (8.85-10.45)	10.89	15.66	
Faller (n=11)	8.10	9.19	10.39 (8.68-15.30)	15.30	20.97	P= .06
Non-Faller (n=42)	7.32	8.29	9.27 (8.85-10.54)	11.32	15.46	

Table 3. Spearman correlations of outcome measures with 4SST.

Table 3. Spearman Correlations of Outcome Measures with FSST		
Outcome Measure	**4SST**	***p* Value**
Age	0.59	< 0.001
Gender	0.22	0.11
MDS-UPDRS-III	0.61	< 0.001
Mini-BESTest	-0.65	< 0.001
FTSTS	0.58	< 0.001
6MWT	-0.52	< 0.001
9HPT	0.65	< 0.001
FOG-Q	0.44	0.001

Table 4. Final Regression Model for 4SST

4SST	B	SE B	β	p	R^2 When Fit Alone	Partial R^2	Cumulative R^2
FTSTS	0.07	0.03	0.23	0.04	0.26	0.09	0.26
Mini-BESTest	-0.31	0.08	-0.46	0.0001	0.45	0.26	0.51
9HPT	0.09	0.04	0.25	0.02	0.26	0.10	0.56

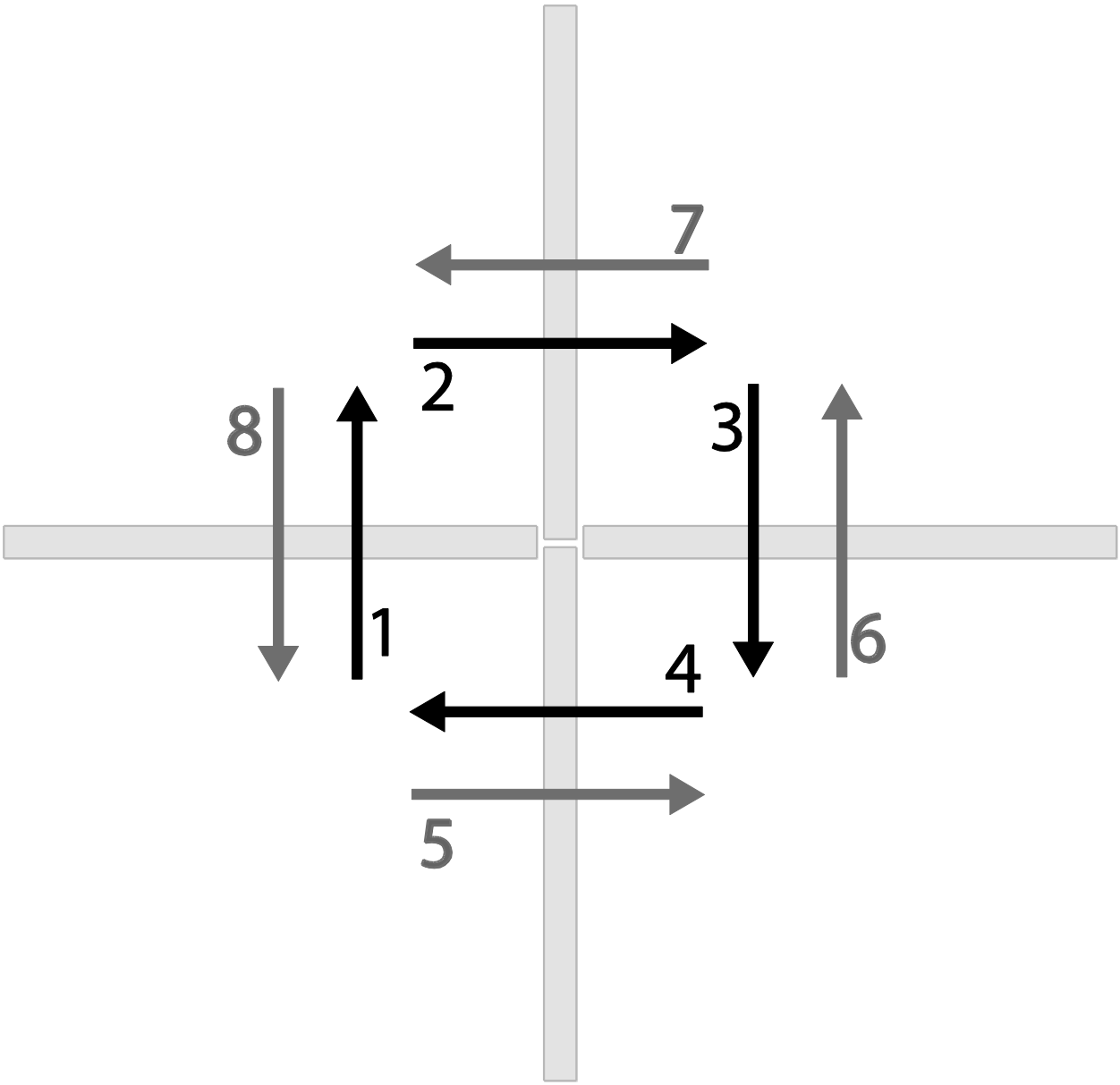

Figure 1. Diagram illustrating the setup utilized for the 4SST. Four separate meter sticks (light gray) were arranged on the floor as shown. Arrows illustrate directions of stepping forward, lateral, and bacward to complete a full circuit first in the clockwise (black, numbers 1-4) and then in the counterclockwise (dark gray, numbers 5-8) direction.

Figure 2. Frequency histogram showing distribution of 4SST times for the full sample when tested on medication. Times are grouped into one second bins.

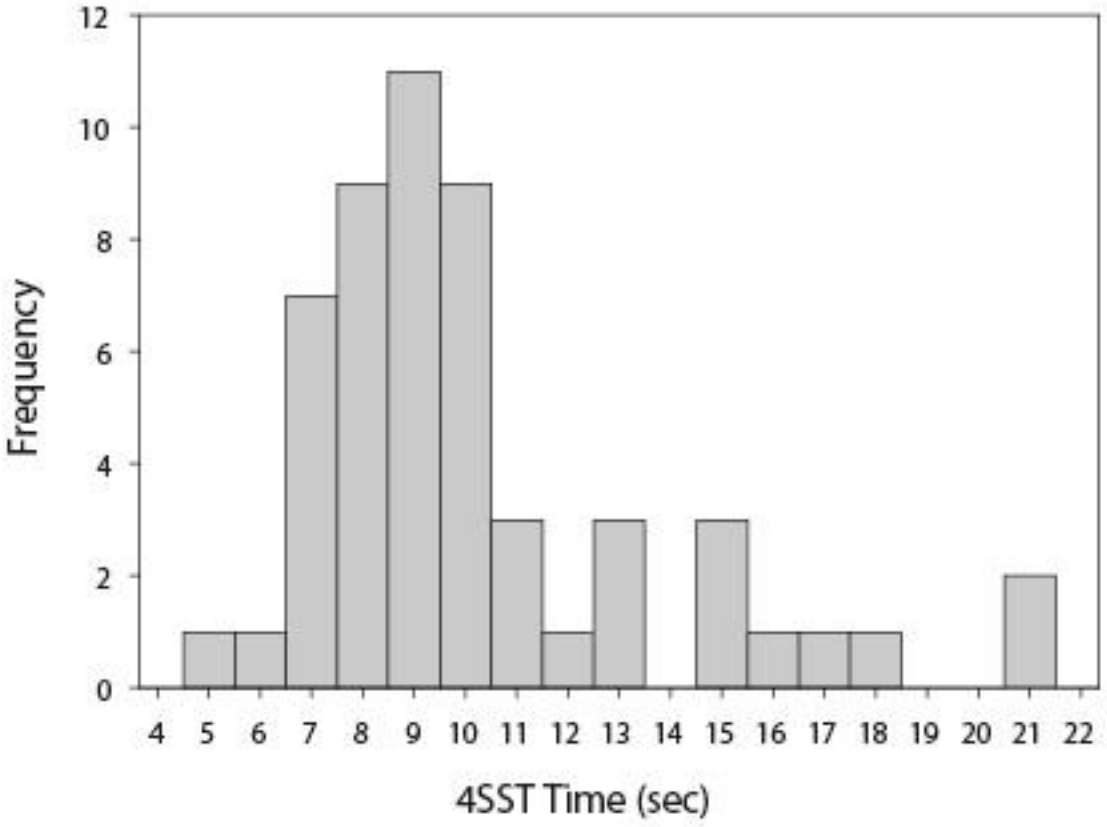

Figure 3. ROC curves of the 4SST and Mini-BESTest.