Psychometric Properties of 3 Functional Mobility Tests for People With Parkinson Disease

Geert Verheyden, Caroline S. Kampshoff, Malcolm E. Burnett, Jessica Cashell, Louis Martinelli, Alice Nicholas, Emma L. Stack, Ann Ashburn

G. Verheyden, PhD, Neuromotor Research Group, Department of Rehabilitation Sciences, KU Leuven, Tervuursevest 101, Bus 1501, 3001 Leuven, Belgium, and Faculty of Health Sciences, University of Southampton, Southampton, United Kingdom. Address all correspondence to Dr Verheyden at: geert.verheyden@faber.kuleuven.be.

C.S. Kampshoff, MSc, Faculty of Health Sciences, University of Southampton.

M.E. Burnett, MSc, Faculty of Health Sciences, University of Southampton.

J. Cashell, MSc, Faculty of Health Sciences, University of Southampton.

L. Martinelli, MSc, Faculty of Health Sciences, University of Southampton.

A. Nicholas, MSc, Faculty of Health Sciences, University of Southampton.

E.L. Stack, PhD, Faculty of Medicine, University of Southampton.

A. Ashburn, PhD, Faculty of Health Sciences, University of Southampton.

Post a Rapid Response to this article at:
ptjournal.apta.org

Background. Standardized outcome measures with high clinical utility are of paramount importance for clinical practice.

Objective. The purpose of this study was to examine interrater and intrarater reliability, construct validity, discriminant ability, and smallest detectable differences of the sit-to-stand test (STS), Timed "Up & Go" Test (TUG), and bed mobility test for people with Parkinson disease (PD).

Design. A cross-sectional, psychometric evaluation study was conducted.

Methods. A group of individuals with PD (PD group) and a group of individuals who were healthy (control group) were recruited through local PD groups and assessed in a movement laboratory in their "on" phase. Measurements of time to perform one STS, TUG, and bed mobility test were collected based on video recordings of that single performance.

Results. Thirty-eight individuals with PD (Hoehn and Yahr stages I–IV) and 19 age-matched control participants were recruited. Intraclass correlation coefficients for interrater and intrarater reliability for the PD group ranged from .95 to .99. Bland-Altman plots showed mean differences close to zero and narrow confidence intervals. Construct validity was established by means of moderate to good Spearman rho correlation coefficients with part III of the Unified Parkinson's Disease Rating Scale and the Hoehn and Yahr stage (range=.51–.63). Timings of all tests discriminated participants in the PD group from those in the control group and participants in the PD group in Hoehn and Yahr stages I and II from those in Hoehn and Yahr stages III and IV but did not discriminate "nonfallers" or those with single falls from repeat "fallers" or "nonfreezers" from "freezers." Applicable smallest detectable differences were established.

Limitations. The results are not generalizable to people in the late stage of PD (Hoehn and Yahr stage IV: n=3).

Conclusions. Timings of video recordings of 3 functional mobility tests with high clinical utility showed good psychometric properties for community-dwelling, ambulatory people with PD.

The use of standardized outcome measures in rehabilitation research and in clinical practice is important in order to assess patients, evaluate the effect of treatment, and communicate among colleagues. For people with Parkinson disease (PD), standardized measures, such as the Hoehn and Yahr stages[1] and the Unified Parkinson's Disease Rating Scale (UPDRS),[2] are available. Both scales provide important information with regard to disease severity, but using these scales in clinical practice to evaluate the effect of treatment is not encouraged due to the subjective nature of the Hoehn and Yahr stages and the time needed to complete the UPDRS. The emphasis in clinical practice should be on standardized assessments that require little time, limited cost, and no specialist equipment and training and that can be performed in the patient's own environment.[3]

People with PD experience a range of mobility difficulties, but those more commonly observed in the clinical setting are deficits related to sit-to-stand (STS) performance, walking and turning, and bed mobility (BM). To our knowledge, psychometric properties for a single STS test have not been reported for people with PD. Suteerawattananon and Protas[4] reported moderate test-retest reliability (intraclass correlation coefficient [ICC]=.76) for the Five-Times Sit-to-Stand Test in people with PD, but only 10 participants completed the test, thus reducing the generalizability of their results.

With regard to walking and turning, the TUG is a standardized measure for the elderly population, but limited psychometric literature is available, specifically for people with PD. Morris et al[5] showed the TUG to be highly reliable and valid for people with PD, with ICC values of .99 for interrater and intrarater reliability. Podsiadlo and Richardson[6] investi-

gated the psychometric properties of the TUG in elderly people, including participants with PD, but did not report their results separately for those with PD. Discriminant ability of the TUG for people with PD was evaluated in 3 studies. Morris et al[5] showed that average TUG times differed between the "on" and "off" phases, and in the study by Thompson and Medley,[7] the TUG was able to detect differences between participants with PD and a control group of individuals who were healthy and across Hoehn and Yahr stages. However, the sample size in the latter study was small (N=19) and warrants further investigation to confirm the conclusions. Finally, Nocera and colleagues[8] reviewed the records of 2,097 people with PD to investigate the ability of the TUG to identify people with PD who are at risk for falling. Their results indicated that 74% of the participants with PD were correctly classified as "fallers" or "nonfallers" based on their TUG time, and they reported a cutoff score of 11.5 seconds for discrimination between those who did and did not fall.[8]

Bed mobility (ie, turning, sitting, and then standing up from the bed) also has been investigated in people with PD. A 3-point ordinal rating scale method showed acceptable interrater and intrarater reliability (Cohen kappa=.64–.79).[9] Other studies timed the movement, grading it on a 4-point scale, and reported excellent interrater reliability (Cohen kappa= 1.00)[10] and ICCs ranging from .77 to .84.[11,12] Nevertheless, there is some research that is less supportive of the intrarater reliability for the BM test. Existing tools such as the Physical Performance Test and Physical Activity Rating Scale have implemented timed functional tests to evaluate functional measures in people with PD.[4] Out of all of the functional measures reported in the study, only the BM test (recorded by timing the lie-to-sit movement) produced a

low ICC of .50, suggesting the need for further investigation of reliability. Validity of the BM test has been examined by comparing the grades of a 4-point ordinal scale with UPDRS scores, producing significant correlations (−.67 and −.63, P<.001), but in this study,[10] the BM test was part of a larger assessment: the Lindop Parkinson's Disease Mobility Assessment.

In summary, for some of our functional mobility tests, information is available concerning the reliability and validity for people with PD and the ability to discriminate between people with PD and controls, but no study to date has comprehensively investigated the psychometric properties of these 3 tests together and evaluated the discriminant ability between "nonfallers" and "fallers" and between "nonfreezers" and "freezers." Therefore, with limited and sometimes conflicting information available concerning psychometric properties for the STS, TUG, and BM tests, it was the aim of this study to examine interrater and intrarater reliability, construct validity, discriminant ability, and the smallest detectable difference (SDD) of these 3 functional mobility tests for people with PD.

Method
A cross-sectional, observational study design was used.

Participants
People with PD were recruited through local PD groups via an invitation letter. Inclusion criteria were confirmed diagnosis of PD by a consultant, independent mobility when assessed in our movement laboratory, and community dwelling. Partners of individuals with PD were invited to participate as controls. People were excluded if they had dizziness or vestibular dysfunction, visual impairments that could not be corrected with glasses, any other

Procedure

People with PD were visited at home for the initial screening. Hoehn and Yahr stage[1] and part III of the UPDRS[2] were administered, plus participants were screened for falls and freezing of gait (FOG). We asked people about the frequency and circumstances of falls in the previous 12 months and defined a *fall* as "an event that results in a person coming to rest unintentionally on the ground or other lower level, not as a result of a major intrinsic event or overwhelming hazard."[14(p118)] We identified a faller as someone who experienced 2 or more fall events in the previous 12 months.[14] *Freezing of gait* was defined as an episodic involuntary inability to generate or maintain walking at least once a week (freezer score >1 on question 3 of the FOG questionnaire).[15] Before the study, participants were provided with a participant information sheet, and informed consent was obtained.

Measurements

Participants attended the movement laboratory of our faculty at Southampton General Hospital for testing. People with PD continued to take their normal dose of medication and were tested in the "on" phase. The STS test required the participant to be seated with his or her back against a standard size chair with armrests. The participant was asked to stand up from the chair in his or her usual way. The STS test commenced with the word "start" and finished when the participant was standing and vertical movement had ceased. The TUG[6] involved the participant starting in the same seated position as in the STS test. On "start," each participant stood up and walked 3 m at his or her self-selected pace, turned 180 degrees after cross-

ing a white line, and walked back to finish in the seated position in the chair. Timing ceased once the individual was seated (buttocks touching the chair). Participants were allowed to push up from the chair, but doing so did not result in a time penalty. Our sample did not use walking aids. The BM task[10] involved the person starting in a supine position on a bed. On "start," the participant got up to a seated position on the edge of the bed. Timing ceased when the participant was seated at the edge of the bed in an upright sitting position and movement had stopped. One practice trial was allowed to acquaint to all tasks. Finally, participants performed one STS, TUG, and BM task that was video recorded and used for analysis.

Tasks were video recorded by a research assistant (C.S.K.), and afterward the single performances were timed with handheld stopwatches by 3 coauthors (J.C, L.M, and A.N.) independently of each other to establish interrater reliability. The 3 raters were free to pause the video, rewind, and view it more than once if needed. Intrarater reliability was established by comparing timings of a randomly chosen coauthor (J.C., L.M., or A.N.) who repeated the measurement on a separate occasion approximately 1 week later. A pilot study using the same method and including participants who were healthy as well as people with PD not included in this study standardized and acquainted all 3 raters. We used one angle (ie, a lateral view) to video record the STS, TUG, and BM tests. We used a tripod and a wide-angle, fixed view for all recordings. We video recorded the lateral side of the participant when performing the STS and TUG tasks. For the BM task, the lateral side was recorded at the beginning when the participant was positioned supine; thus, the participant ended with facing the camera

when finishing performing the BM task.

Data Analysis

Descriptive statistics were calculated for the demographic and clinical data of our sample. Reliability was examined by calculating ICCs and 95% confidence intervals. Interrater reliability assessed agreement between the timings of the 3 independent raters. Intrarater reliability investigated agreement between the first and second timings of the same rater. Intraclass correlation coefficients above .80 would establish good reliability.[16]

Furthermore, we constructed Bland-Altman plots for interrater and intrarater agreement for each mobility task. Mean timings were plotted on the x-axis, and differences between timings were plotted on the y-axis. For interrater agreement, mean timings were calculated as the mean between the first timing of the rater who rated the task on 2 occasions and the mean of the timings of the other 2 raters. The difference was calculated as the difference between the first timing of the rater who rated the task on 2 occasions and the mean of the timings of the other 2 raters. For intrarater agreement, mean timings were calculated between the 2 timings of the rater who rated the participants on 2 occasions. The difference was calculated between these 2 timings. We plotted the mean difference and 95% limits of agreement of the mean; we calculated the latter as mean ± 1.96 × standard deviation.

Construct validity was investigated by calculating Spearman rho correlation coefficients and corresponding *P* values among the 3 mobility tasks and the score on part III of the UPDRS and the Hoehn and Yahr stage. Correlation coefficients below .40 would be classified as poor, .41 to .60 as moderate, .61 to .80 as

good, and above .80 as very good.[17] Additionally, we assessed discriminant ability of the timings of the mobility tasks by means of Mann-Whitney U tests between the PD and control groups, between participants in the PD group who were in Hoehn and Yahr stages I and II and those in Hoehn and Yahr stages III and IV, between participants in the PD group who were nonfallers or single fallers and repeat fallers, and between participants in the PD group who were nonfreezers and those who were freezers. Finally, we calculated the SDD (SDD=standard error of measurement [SEM] $\times$ 1.96[$\sqrt{2}$]) based on the SEM (SEM = SD $\times$ [$\sqrt{(1-ICC)}$]) for interrater and intrarater agreement.[18]

Level of significance was set at $P<.05$. Analyses were conducted with SPSS version 17 (SPSS Inc, Chicago, Illinois).

Role of the Funding Source
This work was supported by a research grant from Parkinson's UK (grant no. G-0802).

Results
Our study included 38 people with PD and 19 individuals who were healthy and served as a control group. The participants in the PD group (23 men and 15 women; mean age=69 years, SD=8, range=47–88) were distributed over Hoehn and Yahr stages I to IV: 12 in stage I, 10 in stage II, 13 in stage III, and 3 in stage IV. Thus, the PD group included 22 people in Hoehn and Yahr stages I and II and 16 people in Hoehn and Yahr stages III and IV. We recruited 23 nonfallers or single fallers, 15 repeat fallers (2 or more falls in the previous 12 months), and 20 nonfreezers and 18 freezers (>1 on item 3 of the FOG questionnaire). Mean disease duration was 7 years (SD=4, range=1–18), and the mean UPDRS part III score was 17 (SD=6, range 4–35). The participants in

Table 1.
Intraclass Correlation Coefficients (95% Confidence Interval) for Interrater and Intrarater Reliability of the 3 Functional Mobility Tests for People With Parkinson Disease

Measure	Sit-to-Stand Test	Timed "Up & Go" Test	Bed Mobility Test
Interrater reliability	.95 (.91–.97)	.99 (.99–.99)	.97 (.95–.98)
Intrarater reliability	.98 (.96–.99)	.99 (.99–1.00)	.99 (.99–.99)

the control group (6 men and 13 women) had a mean age of 68 years (SD=0, range=52–85).

Interrater and Intrarater Reliability
For all 3 functional tests, good reliability was established, with values for interrater reliability ranging between .95 and .99 and values for intrarater reliability ranging between .98 and .99 (Tab. 1).

Figures 1, 2, and 3 show the Bland-Altman plots for interrater and intrarater agreement and indicate mean differences close to zero as well as narrow confidence intervals, with the exception of interrater agreement for the BM test (Fig. 3A), where the mean difference was slightly above zero and the confidence interval was wider in comparison with the other plots.

Construct Validity
Spearman rho correlation coefficients between the STS test and part III of the UPDRS and the Hoehn and Yahr stage were moderate and significant ($P\leq.001$), being .53 and .58, respectively. The correlation coefficients for the TUG with part III of the UPDRS and with the Hoehn and Yahr stage were good (.61, $P\leq.001$) and moderate (.51, $P\leq.001$), respectively. Finally, for the BM test, correlation coefficients with part III of the UPDRS and the Hoehn and Yahr scale also were good (.63, $P\leq.001$) and moderate (.54, $P\leq.001$), respectively.

Discriminant Ability and SDD
All 3 functional mobility tests discriminated between the PD and control groups (Tab. 2). The PD group took significantly longer to perform the STS, TUG, and BM tests. On average, the control group performed the STS, TUG, and BM tests 38%, 31%, and 39% faster, respectively, in comparison with the PD group. The 3 tests also discriminated people in Hoehn and Yahr stages I and II from people in Hoehn and Yahr stages III and IV; people in Hoehn & Yahr stages III and IV took significantly longer to perform the 3 tests (Tab. 2). On average, people in Hoehn and Yahr stages I and II performed the STS, TUG, and BM tests 22%, 16%, and 32% faster, respectively, in comparison with people in Hoehn and Yahr stages III and IV. The 3 functional mobility tests did not discriminate between nonfallers or single fallers and repeat fallers or between nonfreezers and freezers (Tab. 2).

The SDD values for interrater agreement for the STS, TUG, and BM tests were 0.61, 2.13, and 1.88 seconds, respectively. For intrarater agreement, we found SDD values of 0.39, 2.13, and 1.08 seconds for the STS, TUG, and BM tests, respectively.
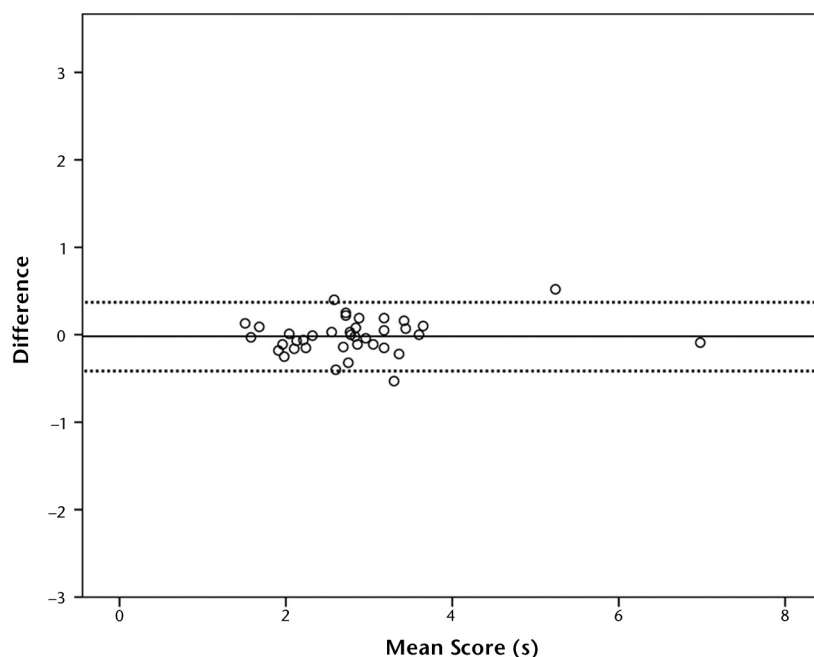
Discussion
The results of our study indicate established interrater and intrarater reliability, construct validity, and the ability to discriminate between people with PD and people who are healthy and between people in

A
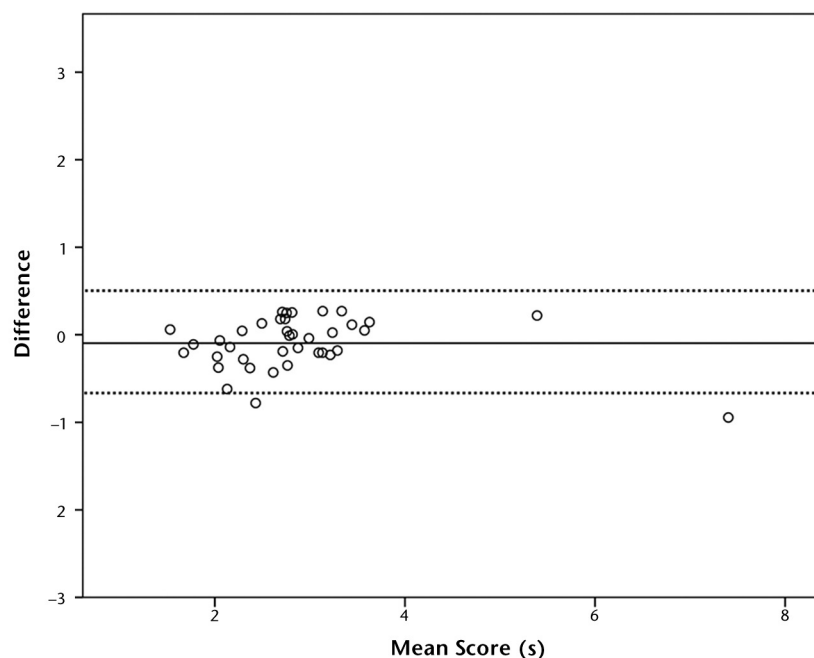
B

Figure 1.
Bland-Altman plots for interrater (A) and intrarater (B) agreement of the sit-to-stand test in people with Parkinson disease. Plots show mean values on x-axis and differences on y-axis, together with mean difference (solid line) and upper and lower limits of 95% confidence interval of the mean difference (dotted lines).

Hoehn and Yahr stages I and II and those in Hoehn and Yahr stages III and IV for the STS, TUG, and BM tests in people with PD comparable to the participants included in our sample. Furthermore, applicable SDDs were reported for all 3 tests. To our knowledge, this is the first study that evaluated a range of psychometric properties for these 3 functional mobility tests that are quick and easy to assess.

Reliability

The ICC values for interrater and intrarater reliability of all 3 tests were .95 or higher, and all had narrow confidence intervals. In this study, we opted for a single STS test. A previous study investigated test-retest reliability of the Five-Times Sit-to-Stand Test and reported a lower value (ICC=.76).[4] The authors also did not explain who rated the test and how it was timed exactly, making reproducibility and integration into clinical practice difficult. We opted for a single STS test because the literature suggests that the multiple STS test is more a test of endurance, which was not the aim of our measurement.[19] Furthermore, differences in ICC results could be explained by the fact that Suteerawattananon and Protas[4] examined test-retest reliability, whereas our study evaluated intrarater reliability. Patients redoing a test can be a source of variability, leading to possibly lower ICC values. Reliability values obtained for the TUG and BM tests in this study were comparable to values obtained in the study by Morris and colleagues.[5]

We noted with interest the slightly increased mean difference and widened 95% confidence interval in Figure 3A, presenting the Bland-Altman plot for interrater agreement of the BM test. A possible explanation for this finding is that although we used a thorough standardized protocol, our agreed-upon definition for the

end of movement could be further improved, as there were occasions when recording the endpoint of the BM test was difficult to assess due to participants shuffling on the edge of the bed. This subjective element of timing could have caused the discrepancies among raters. Future research could further standardize the endpoint of the BM test with an additional definition or cue such as resting with hands on the lap and coming to a complete stop.

Validity

We found moderate correlations between the STS test and part III of the UPDRS and Hoehn and Yahr stage. Although comparable for the Hoehn and Yahr stage, correlations were slightly higher and classified as "good" for the TUG and BM tests and part III of the UPDRS. We believe that these correlation coefficients indicate construct validity for our 3 functional mobility tests, as all measures have a similar underlying construct, and thus positive (significant) correlations indicate that longer timings to perform the tests are related to higher UPDRS part III scores and Hoehn and Yahr stage, with both latter measures giving higher scores to worse motor and mobility performances. Because we were unable to find studies relating the STS test with the UPDRS and Hoehn and Yahr stage, we are unable to compare our results with the previous literature.

Our results for the TUG confirm the findings in previous studies, which also showed a positive correlation between TUG timings and the UPDRS score.[20,21] Our literature search did not result in a previous study exploring the relationship between BM and UPDRS score or Hoehn and Yahr stage.

Discriminant Ability and SDD

All 3 tests discriminated significantly between the PD and control groups and between people in Hoehn and
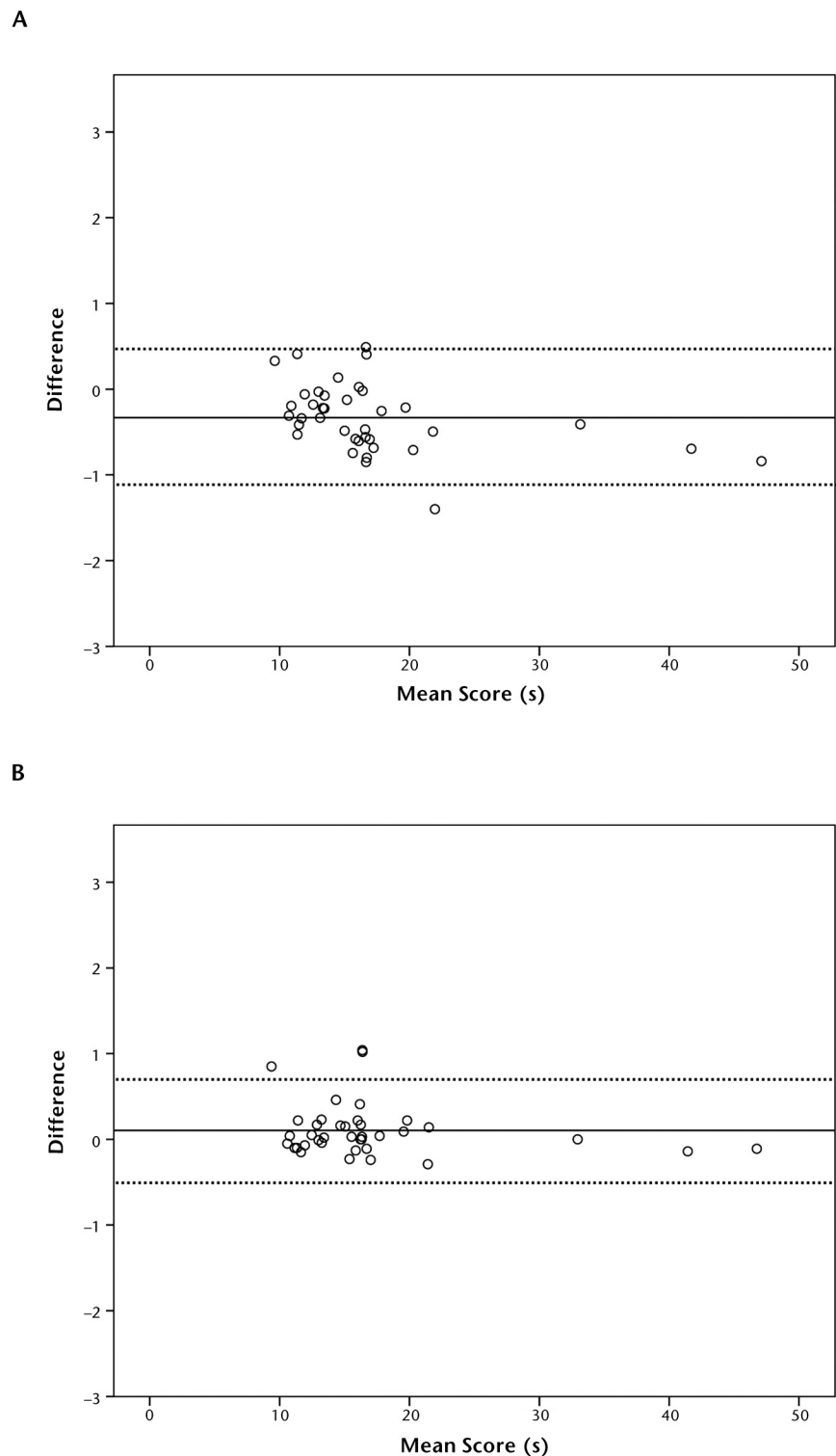
Figure 2.
Bland-Altman plots for interrater (A) and intrarater (B) agreement of the Timed "Up & Go" Test in people with Parkinson disease. Plots show mean values on x-axis and differences on y-axis, together with mean difference (solid line) and upper and lower limits of 95% confidence interval of the mean difference (dotted lines).

A

B

Figure 3.
Bland-Altman plots for interrater (A) and intrarater (B) agreement of the bed mobility test in people with Parkinson disease. Plots show mean values on x-axis and differences on y-axis, together with mean difference (solid line) and upper and lower limits of 95% confidence interval of the mean difference (dotted lines).
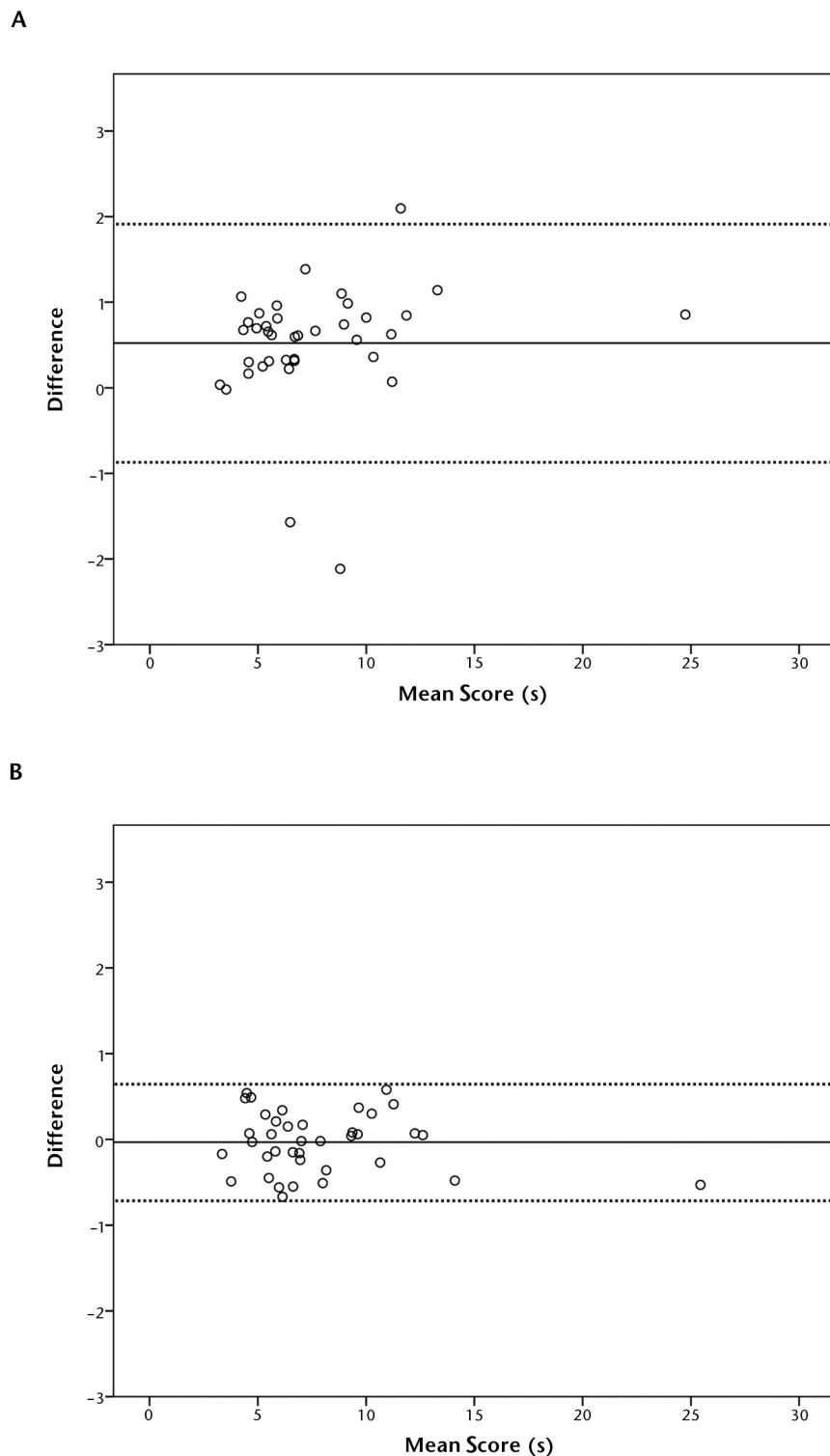
Yahr stages I and II and those in Hoehn and Yahr stages III and IV but not between nonfallers or single fallers and repeat fallers or between nonfreezers and freezers. Our results confirm the findings of previous biomechanical studies that people with PD are slower getting up from a chair compared with people who are healthy.[22-27] Our sample of people with PD was 38% slower than their healthy counterparts, which is slightly lower than but comparable to the 50% reported by Mak and Hui-Chan.[22] Some authors have argued that reduced hip flexion, difficulties switching from flexion to extension, and lower hip torques related to reduced hip strength correlate with difficulty rising from a chair in people with PD.[22-27] Thompson and Medley[7] investigated discriminant ability of the TUG for the different Hoehn and Yahr stages. They concluded that the test was able to detect differences between people in Hoehn and Yahr stages I, II, and III and people who were are healthy; thus, our results are in line with their findings. We opted to combine Hoehn and Yahr stages I and II and stages III and IV based on the relatively low number of participants in each stage, especially stage IV (n=3). Recently, Nocera et al[8] suggested discriminative ability for the TUG in people with PD, but this suggestion was based on a different methodological approach. Their results indicated that, overall, 74% of their sample could be correctly classified, but classification of fallers (sensitivity=54%) was lower compared with classification of nonfallers (specificity=85%).[8] Our literature search did not result in a previous study investigating discriminant ability of a BM test.

Of interest was the fact that although the timings of the 3 functional mobility tests were all relatively slower for repeat fallers in comparison with nonfallers or single fallers and for

Table 2.
Discriminant Ability Analysis of the 3 Functional Mobility Tests for Participants With Parkinson Disease (PD Group) and Participants Who Were Healthy (Control Group)[a]

Measure	Sit-to-Stand Test	Timed "Up & Go" Test	Bed Mobility Test
PD group	2.76 (2.19–3.18)	15.7 (12.77–16.78)	6.78 (5.49–9.62)
Control group	1.71 (1.63–2.03)	10.9 (10.34–11.81)	4.11 (3.54–4.74)
P	≤.001	≤.001	≤.001
Hoehn & Yahr stage I–II	2.28 (1.98–2.96)	13.79 (11.39–16.32)	5.91 (4.71–6.99)
Hoehn & Yahr stage III–IV	2.91 (2.76–3.42)	16.48 (14.77–21.01)	8.74 (6.73–11.19)
P	≤.001	.01	≤.001
Nonfallers/single fallers	2.72 (2.1–3.18)	14.68 (11.94–16.38)	6.39 (4.74–9.66)
Repeat fallers	2.84 (2.58–3.3)	16.23 (14.34–17.01)	7.01 (5.84–9.61)
P	.34	.26	.41
Nonfreezers	2.57 (2.01–3.22)	13.79 (11.72–16.36)	6.27 (4.7–9.76)
Freezers	2.82 (2.67–3.18)	16.21 (14.89–18.23)	7.95 (5.95–9.62)
P	.22	.07	.18

[a] Values are expressed as median time in seconds (interquartile ratio) and corresponding P value for the Mann-Whitney U test.

freezers in comparison with non-freezers, the measurements did not reach significance, and thus no discriminant ability could be established for these subgroups. The fact that the STS, TUG, and BM tests are specific functional mobility measures and falls are multifactorial in nature may explain these results.[14] Similarly, freezing is typically a problem that occurs within gait,[28] thus limiting a possible relationship with the STS and BM tests. The TUG does include walking (and turning), and indeed the results in Table 2 show the relative difference between nonfreezers and freezers is larger for the TUG in comparison with the STS and BM tests and the P value for the TUG comparison shows a trend toward significance (P=.07).

For all 3 functional mobility tests, we established applicable SDDs, which can be used in clinical practice as well as rehabilitation research but, due to the lack of examining test-retest reliability, are limited to timing of single performances in groups of patients having similar variance to our sample. Scientists as well as clinicians should continue to be aware

of the difference between statistical and clinical difference. In order to consider these differences, properties such as the SDD are useful. When looking at the results in Table 2, differences in timings for all 3 tests were statistically significant between the PD and control groups and between people in Hoehn and Yahr stages I and II and those in stages III and IV, but the differences also exceed the SDD values that we reported in this study, thus further validating the discriminant ability we found in our sample. To our knowledge, this is the first study that provides such a comprehensive overview of psychometric properties for functional mobility measures in people with PD. These measures can now be used with confidence in larger-scale studies. Schenkman et al[29] reported previously on functional limitations and task performance in a sample of 339 people with early- and middle-stage PD. With evidence of reliability and validity for measures in an elderly population, validation is necessary in more specific populations as well, such as in people with PD, as they display specific motor deficits that

might influence assessment of performance. Further results from our study, especially the SDD data, may assist in interpreting the result of clinical trials aimed at improving functional mobility in people with PD.

As mentioned in the introduction, besides established psychometric properties, outcome measures for clinical practice should have good clinical utility; they should require little time, limited cost, and no specialist equipment and training, and they can be taken into the patient's own environment.[3] We believe these requirements for clinical utility are all true for the functional mobility tests we evaluated. It does not take long to ask a patient to perform a single STS, TUG, or BM test. The only equipment needed is a video camera, which also can be used for other outcome measures or for filming a home exercise program for a patient, so it is not solely of use for these tests. Furthermore, when using a video camera, the tests should only be performed once; the timing can happen later. A video camera these days is not what we would call spe-

cialist equipment, and for timing the tests, no specialist training is necessary, although we must emphasize the point that training and the use of a standardized protocol should be encouraged, even for measuring with a stopwatch the timings of a functional mobility test. Finally, these tests can easily be performed at home, which we would argue would highlight the ecological validity of these tests, as the patient can use his or her own chair or bed.

There are some limitations of our study that have to be considered. Participants were recruited through convenience; therefore, our sample may not be representative of the PD population or the elderly population. Most of our sample of people with PD was classified into the first 3 Hoehn and Yahr stages, and all of these participants were in the "on" phase of medication. Our sample did not include those who are more severely affected by PD or people in the "off" phase; thus, psychometric properties of this subgroup and this phase cannot be presented.

One should also consider our sample of men and women across our groups. The healthy group was female dominated, whereas the PD group was male dominated. Differences in build and strength (eg, leg length) may affect overall speed and could have accounted for differences between the groups. Furthermore, although the raters were blinded for study group (PD or control) when observing the videos, motor problems displayed by the PD group were easily recognizable; therefore, bias could have been introduced when observing the videos of the 3 functional mobility tests. Furthermore, one has to consider the fact that we evaluated interrater and intrarater reliability and did not incorporate test-retest agreement. The reliability tested in this study addressed measurement between

and among investigators, rather than between assessments.

One must also be aware that reliability based on a single measurement is a limitation of our study; intra-participant variability is possibly more important in determining whether a participant changed with an intervention than is the variability in measurement when viewing performance from a video recording. Generalizability of our reliability values to non–video-recorded performance also should be considered with caution. Finally, one could suggest that video recordings are less practical in a clinical setting compared with direct observations, but we believe that video recordings are more beneficial because they can be stored for comparison between performances over time or to discuss performance together with the patient. Furthermore, from a standardization point of view, the same performance can be viewed repeatedly, if necessary, which again is more advantageous than asking the patient to perform the test again.

Dr Verheyden and Dr Ashburn provided concept/idea/research design. Dr Verheyden, Ms Kampshoff, Mr Burnett, Ms Cashell, and Dr Ashburn provided writing. Ms Kampshoff, Mr Burnett, Ms Cashell, Mr Martinelli, and Ms Nicholas provided data collection. Dr Verheyden, Ms Kampshoff, Ms Cashell, Mr Martinelli, and Dr Ashburn provided data analysis. Dr Verheyden, Ms Kampshoff, and Dr Ashburn provided project management. Dr Verheyden provided fund procurement. Mr Burnett provided facilities/equipment. Ms Kampshoff, Ms Cashell, Ms Nicholas, Dr Stack, and Dr Ashburn provided consultation (including review of manuscript before submission).

References

1 Hoehn MM, Yahr MD. Parkinsonism: onset, progression, and mortality. *Neurology.* 1967;17:427–442.

2 Martínez-Martín P, Gil-Nagel A, Gracia LM, et al; the Cooperative Multicentric Group. Unified Parkinson's Disease Rating Scale: characteristics and structure. *Mov Disord.* 1994;9:76–83.

3 Tyson F, Connell L. The psychometric properties and clinical utility of measures of walking and mobility in neurological conditions: a systematic review. *Clin Rehabil.* 2009;23:1018–1033.

4 Suteerawattananon M, Protas EJ. Reliability of outcome measures in individuals with Parkinson's disease. *Physiother Theory Pract.* 2000;16:211–218.

5 Morris S, Morris E, Iansek R. Reliability of measurements obtained with the Timed "Up and Go" Test in people with Parkinson's disease. *Phys Ther.* 2001;81:810–817.

6 Podsiadlo D, Richardson S. The Timed Up and Go: a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc.* 1991;39:142–148.

7 Thompson M, Medley A. Performance of individuals with Parkinson's disease on the Timed Up & Go. *J Neurol Phys Ther.* 1998;22:16–21.

8 Nocera JR, Stegemöller EL, Malaty IA, et al. Using the timed up & go test in a clinical setting to predict falling in Parkinson's disease. *Arch Phys Med Rehabil.* 2013;94:1300–1305.

9 Kamsma YPT, Brouwer WH, Lakke JPWF. Training of compensational strategies for impaired gross motor skills in Parkinson's disease. *Physiother Theory Pract.* 1995;11:209–229.

10 Pearson MJT, Lindop FA, Mockett SP, et al. Validity and inter-rater reliability of the Lindop Parkinson's Disease Mobility Assessment: a preliminary study. *Physiotherapy.* 2009;95:126–133.

11 Shenkman M, Cutson TM, Kuchibhatla M, et al. Reliability of impairment and physical performance measures for persons with Parkinson's disease. *Phys Ther.* 1997;77:19–27.

12 Purser JL, Pieper CF, Duncan PW, et al. Reliability of physical performance tests in four different randomized clinical trials. *Arch Phys Med Rehabil.* 1999;80:557–561.

13 Folstein MF, Folstein SE, McHugh PR. "Mini-Mental State": a practical method for grading the cognitive state of patients for the clinician. *J Psych Res.* 1975;12:189–198.

14 Clark RD, Lord SR, Webster IW. Clinical parameters associated with falls in an elderly population. *Gerontology.* 1993;39:117–123.

15 Giladi N, Shabtai H, Simon E, et al. Construction of freezing of gait questionnaire for patients with Parkinsonism. *Parkinsonism Relat Disord.* 2000;6:165–170.

16 Koch G. Intraclass correlation coefficient. In: Kotz S, Johnson NL, eds. *Encyclopedia of Statistical Sciences.* New York, NY: John Wiley & Sons Inc; 1982:213–217.

17 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159–174.

18 Portney LG, Watkins MP. *Foundations of Clinical Research: Application to Practice.* 3rd ed. Upper Saddle River, NJ: Pearson-Prentice Hall, 2009.

19 Netz Y, Ayalon M, Dunsky A, et al. "The multiple sit-to-stand" field test for older adults: what does it measure? *Gerontology.* 2004;50:121–126.

20 Brusse KJ, Zimdars S, Zalewski KR, et al. Testing functional performance in people with Parkinson disease. *Phys Ther.* 2005;85:134–141.

21 Matinolli M, Korpelainen JT, Korpelainen R, et al. Mobility and balance in Parkinson's disease: a population-based study. *Eur J Neurol.* 2009;16:105–111.

22 Mak MK, Hui-Chan CW. The speed of sit-to-stand can be modulated in Parkinson's disease. *Clin Neurophysiol.* 2005;116:780–789.

23 Mak MK, Hui-Chan CW. Audiovisual cues can enhance sit-to-stand in patients with Parkinson's disease. *Mov Disord.* 2004;19:1012–1019.

24 Pääsuke M, Ereline J, Gapeyeva H, et al. Leg-extension strength and chair-rise performance in elderly women with Parkinson's disease. *J Aging Phys Act.* 2004;12:511–524.

25 Ramsey VK, Miszko TA, Horvat M. Muscle activation and force production in Parkinson's disease during sit to stand transfers. *Clin Biomech (Bristol, Avon).* 2004;19:377–384.

26 Inkster LM, Eng JJ, MacIntyre DL, et al. Leg muscle strength is reduced in Parkinson's disease and relates to the ability to rise from a chair. *Mov Disord.* 2003;18:157–162.

27 Inkster LM, Eng JJ. Postural control during a sit-to-stand task in individuals with mild Parkinson's disease. *Exp Brain Res.* 2004;154:33–38.

28 Morris ME. Movement disorders in people with Parkinson's disease: a model for physical therapy. *Phys Ther.* 2000;80:578–597.

29 Schenkman M, Ellis T, Christiansen C, et al. Profile of functional limitations and task performance among people with early- and middle-stage Parkinson disease. *Phys Ther.* 2011;91:1339–1354.