HSS Journal®

Validation of Physical Performance Tests in Individuals with Advanced Knee Osteoarthritis

Saurabh P. Mehta, PT, PhD ⓘ · Nathan Morelli, PT, DPT · Caleb Prevatte, PT, DPT · Derrick White, PT, DPT · Ali Oliashirazi, MD

Abstract *Background:* Individuals with advanced osteoarthritis (OA) of the knee experience significant impairments in balance and in essential physical functions such as walking and rising from a chair. There is limited evidence on valid outcome measures to capture these impairments. *Questions/Purposes:* We sought to examine the construct validity of three physical performance measures in patients with advanced knee OA: a gait speed (GS) test, the Short Physical Performance Battery (SPPB), and the Timed Up and Go (TUG) test. *Methods:* We designed a cross-sectional clinical measurement study in which patients with advanced knee OA completed two self-reported measures: the Knee Injury and Osteoarthritis Outcome Score—Physical Function Shortform (KOOS-PS) and a four-part numeric pain rating scale (Q-NPRS). They were also administered the GS test, TUG test, and SPPB. Convergent and divergent construct validity were assessed by examining relationships between the GS test, the SPPB, the TUG test, the KOOS-PS, and the Q-NPRS and calculating Pearson correlation coefficients (*r*). The scores for the GS, TUG test, and SPPB were compared with established normative values for age-matched healthy controls. *Results:* Forty-four subjects (mean age, 66.9 ± 8.1 years) participated in the study. The GS test showed low concordance with the SPPB component tests and the TUG test. The relationships between the physical performance measures and the self-reported measures were low. The scores for the GS test, TUG test, and SPPB in our sample were significantly worse when compared with age-matched normative values, indicating impairments in physical performance. *Conclusion:* These results advance the understanding of the validity of the GS test, TUG test, and SPPB in demonstrating the impairments in physical performance that patients with advanced knee OA experience in walking, balancing, and rising from a chair. Future research should examine the reproducibility and responsiveness of the GS test, TUG test, and SPPB in patients with advanced knee OA, in order to facilitate the integration of these measures into clinical practice.

Keywords knee osteoarthritis · gait speed · physical functioning · balance · clinical measurement study

S. P. Mehta, PT, PhD (✉)
School of Physical Therapy, Marshall University,
SOPT Rm 129, 2874 5th Ave.,
Huntington, WV 25702, USA
e-mail: mehtas@marshall.edu

S. P. Mehta, PT, PhD · A. Oliashirazi, MD
Deptartment of Orthopedic Surgery, Joan C. Edwards School of Medicine, Marshall University,
Huntington, WV, USA

N. Morelli, PT, DPT
Department of Rehabilitation Sciences, College of Health Sciences, University of Kentucky,
Lexington, KY, USA

C. Prevatte, PT, DPT
Department of Physical Therapy, Roper St. Francis Mount Pleasant Hospital,
Mount Pleasant, SC, USA

D. White, PT, DPT
Huntington Physical Therapy,
Huntington, WV, USA

Introduction

Aside from causing pain and limiting movement, advanced osteoarthritis (OA) of the knee can result in impaired physical function and overall restrictions in mobility [12, 35]. In particular, those with advanced knee OA demonstrate markedly worse impairments in postural sway and proprioception, which elevate the risk of falling [23, 42, 43]. Such multidimensional impairments cascade into more-global

problems, such as reduced participation in physical activity and community engagement [14]. It is not surprising that reaching optimal physical activity levels and participating in different life roles are key endpoints for individuals with advanced knee OA and those undergoing total knee replacement (TKR) [3, 7, 32].

Although self-reported measures are recommended for the assessment of pain and functional limitations in patients with advanced knee OA [8, 19], self-reporting may not provide an accurate and comprehensive assessment of a patient's ability to function within his or her environment [38, 39]. In particular, self-reported measures result in overestimates of functional ability in patients with advanced knee OA and those who have undergone TKR [39, 40]. In 2013, the Osteoarthritis Research Society International (OARSI) released consensus-based recommendations regarding which performance-based measures (PBMs) to use to obtain a clearer assessment of physical functioning in individuals with advanced knee OA or TKR [11]. In essence, these PBMs are used to assess such aspects of everyday physical functioning as walking shorter or longer distances, changing directions while ambulating, negotiating stairs, and sitting down and standing up [11]. Despite the OARSI recommendations, the use of PBMs to assess physical functioning in patients with advanced knee OA remains limited in rehabilitation practice. One possible reason is that the evidence supporting the measurement properties of these PBMs in people with advanced knee OA is still emerging. There is preliminary evidence to support the use of the Timed Up and Go (TUG) and chair stand tests in advanced knee OA [9, 10]. Takacs et al. showed that the TUG and gait speed (GS) tests both have adequate concurrent and discriminant construct validity in the assessment of patients with knee OA [41]. Nonetheless, most participants in their study had mild-to-moderate knee OA and were highly functional. The evidence supporting the measurement properties of the common PBMs in assessing physical function in individuals with advanced knee OA is limited [20, 44, 47].

The Short Physical Performance Battery (SPPB) is a commonly used PBM in the assessment of physical functions as they relate to lower-extremity impairments [17]. The SPPB uses balance tests, a 4-m GS test, and the five-repetition sit-to-stand test to assess the performance of functional tasks [17]. Research has consistently demonstrated that when used in older adults with different pathologies, the SPPB has high interrater and test–retest reliability, with an intraclass correlation coefficient (ICC) exceeding the "acceptable" benchmark of 0.75 [24, 28, 29]. The SPPB has been found to be valid and provides a fair prediction of the development of functional disability, as well as the risk of institutionalization, in individuals with lower-extremity impairments [16, 17]. The advantage of using the SPPB is that it provides a reliable and valid estimate of important domains of physical functioning but carries a low administrative burden. Importantly, two of the areas in which it captures impairments—the ability to walk for short distances and the ability to stand up and sit down—have been deemed crucial by the OARSI in the assessment of

physical functioning in patients with knee OA [10]. Research has also found the SPPB to be clinically useful in assessing physical functioning in patients with knee OA [45]. There is also evidence suggesting that SPPB scores have good concurrent and discriminant construct validity in the assessment of individuals with symptomatic knee OA [34]. Nonetheless, the evidence regarding the measurement properties of the SPPB in individuals with advanced knee OA remains limited and needs further empirical validation.

With an objective of advancing the knowledge and evidence surrounding the measurement properties of commonly used measures of physical performance in advanced knee OA population, this study aimed to answer three specific research questions: (1) Do the GS test, the TUG test, and the SPPB have acceptable construct validity in assessing physical performance in patients with advanced knee OA when being evaluated by an orthopedic surgeon for suitability for TKR? (2) Do GS test, TUG test, and SPPB scores adequately discriminate between healthy subjects and those with symptomatic advanced knee OA? (3) Do the GS test scores obtained using two different test variants agree with each other?

Methods

This was a cross-sectional clinical measurement study with one collection session to obtain the required data. The data were extracted from two separate cross-sectional studies (unpublished data), both of which had a common overarching goal of validating balance and physical performance measures in individuals with advanced knee OA seeking consultation regarding TKR.

Patients

Patients with advanced knee OA who were consulting an orthopedic surgeon at a hospital-based outpatient tertiary care clinic were approached to participate in both of the cross-sectional studies. The orthopedic surgeon (A.O.) established the severity of the knee OA using radiographs and clinical presentation, although no specific objective criteria were used for this determination. A set of common inclusion and exclusion criteria guided the selection of participants for the cross-sectional studies. Native English speakers with advanced knee OA were deemed eligible. Excluded from the study were non–English speakers and patients with a history of TKR or total hip replacement on either side, lower-extremity injury in the previous 6 months, or other neuromuscular impairments that could account for reduced physical function. Participants who needed a walker or rollator for ambulation were also excluded. Details of the protocols for both studies were discussed with patients who met eligibility criteria, and consent was obtained from all participants. The current study was approved by the institutional review board at our university.

GS Testing

There are several variants for assessing GS in community-dwelling adults or those with defined pathologies [37]. We chose to test the GS along a 10-m path; in the test, the first 2 m and the last 2 m were considered the areas of acceleration and deceleration, and the middle 6 m was considered the testing distance [36]. Our rationales for using this version were to ensure that the distance was short enough to minimize fatigue, which could have affected subjects' performance, and to use a long enough walk to provide a reliable estimate of walking speed; assessments of GS of less than 6 m have not been considered to have sufficient validity [31]. Procedures for obtaining GS have been described in detail elsewhere [31]. In summary, each subject underwent three trials, the first being a test trial; the average of the subsequent two trials was recorded as the GS. There was a rest break of 30 s between each trial. The time needed to complete the 6-m walking distance was recorded to the nearest 100th of a second. The average time of the two trials was then divided by 6 to determine the GS in meters per second.

The TUG Test

The TUG test is commonly used to assess functional mobility in patients with balance or strength impairments. Each participant was seated on a normal-height chair with their back against the backrest. On the word "go," the participant stood up from the chair, walked to a target 3 m (9.8 ft) away, turned around, walked back to the chair, and sat down. The time (in seconds) from the command "go" to the moment the participant's buttocks touched the chair was recorded. Participants were encouraged to wear regular footwear, use gait devices as necessary, and walk at a comfortable and safe speed during the test. Participants were given one practice trial, and the average of the subsequent two trial times was recorded. Participants had a rest break of 30 s between each trial. The interrater and intrarater reliability of the TUG test in individuals with OA of the knee has been found to be excellent, with an ICC of greater than 0.9 [1]. A recent study also suggested that the TUG test had fair interrater and intrarater reliability, with an ICC greater than or equal to 0.75, but raised concerns over the stability of this point estimate because of a suboptimal lower bound of confidence interval for the ICC value (as low as 0.54) [9].

The SPPB

The SPPB involves the assessment of balance (side-by-side, semitandem, and tandem stand), assessment of GS (over 4 m), and the five-repetition sit-to-stand test. Testing and scoring procedures are described in detail elsewhere (http://hdcs.fullerton.edu/csa/Research/documents/SPPBInstructions_ScoreSheet.pdf). Participants were allowed to use a gait device if they needed to during the GS component of the SPPB. Each of the three components of the SPPB is scored between 0 and 4, where 0 indicates the worst and 4 indicates the best performance in that component. Therefore, the summary score on the SPPB can range from 0 to 12, with 12 indicating the best physical performance. The summary score for the SPPB has been shown to have acceptable reliability (ICC > 0.75) in subjects with pathologies other than advanced knee OA [24, 28, 29].

The Knee Injury and Osteoarthritis Outcome Score—Physical Function Shortform

The seven-item Knee Injury and Osteoarthritis Outcome Score—Physical Function Shortform (KOOS-PS) is a self-reported measure that was developed from the full-length KOOS with the aim of reducing administrative burden and item redundancy while preserving sound measurement properties in assessing physical functions in patients with advanced knee OA [30]. Each of the seven items of the KOOS-PS is scored on a scale of 0 to 4, with 0 indicating no difficulty and 4 indicating extreme difficulty in completing the functional task. The total score across the seven items is converted into an adjusted score of 0 to 100, with 0 indicating complete functional impairment and 100 indicating no impairment. The KOOS-PS has been validated as a tool for examining physical functions in multiple linguistic and cultural contexts [25].

Numeric Pain Rating Scale

The 10-point numeric pain rating scale (NPRS) is arguably the most common tool used by clinicians to assess the intensity of pain. Patients indicate pain intensity by selecting a number between 0 and 10, with 0 indicating no pain and 10 indicating the worst pain. For the purposes of our study, we administered a quadruple NPRS (Q-NPRS) that determined pain intensity using four questions: pain at present, average pain over 24 h, pain when it is at its worst, pain when it is at its best. Each question had the same scaling structure, 0 to 10, used in the single-question NPRS. The responses across the four questions were averaged to obtain the Q-NPRS score. The NPRS is deemed to be reliable and valid in individuals with advanced knee OA [2].

Examiners

Both studies from which the current data were extracted were cross-sectional in nature: the study data were collected over a single session by one of the students enrolled in an entry-level physical therapy program. The three student physical therapists (SPTs) had been exposed to the outcome measures used in this study as part of their curriculum, which ensured standardization to some extent. They also underwent a standardization session before initiating data collection for the study. The purposes of the standardization session were to familiarize the SPTs with the study protocol and ensure consistency in the instructions, the administration of the tests, and the scoring of the outcome measures. The standardization session also involved collecting GS test, TUG test, and SPPB pilot data on 12 healthy subjects to examine interrater reliability among the three SPTs. The analysis revealed that the SPTs had excellent interrater

reliability (ICC > 0.90) [22] in administering the GS test, TUG test, and SPPB.

Protocol

Once participants provided consent, one of the SPTs obtained demographic and health-related information. The demographic variables collected were age, sex, and height and weight. The health-related variables collected were the duration of advanced knee OA pain (in months) and the side affected. Subsequently, participants completed the paper versions of the KOOS-PS and Q-NPRS. Finally, participants were administered the GS test, TUG test, and SPPB in a random order determined a priori. They were allowed to wear the footwear they had worn during their visit. This was to avoid an order effect on test results. Participants were given a 1-min break between each of the tests. The SPPB data were collected in only one of the two cross-sectional studies; therefore, scores for the SPPB were not available for all of the participants in the current study.

Data Analysis

Demographic and health-related variables were summarized using descriptive statistics, including mean and standard deviations (SD) for continuous variables and frequency count (percent) for categorical variables. The assumptions of normality were verified using the Shapiro–Wilk test [13]. The concurrent construct validity of the physical performance measures was examined by assessing convergent and divergent relationships between the measures. Pearson correlation coefficients (r) were calculated to assess these relationships (greater than 0.7 was considered high, 0.50 to 0.7 was moderate, and less than 0.5 was low). We hypothesized that because the GS test, TUG test, and SPPB assess overlapping (albeit diverse) aspects of physical functioning, they would demonstrate moderate-to-high convergent validity with one another, with r values of greater than or equal to 0.50 (hypothesis 1) [27]. We also hypothesized that the GS test, TUG test, and SPPB, being PBMs, would demonstrate poor convergent validity with the self-reported measures (the KOOS-PS and Q-NPRS), with r values of less than 0.50 (hypothesis 2) [27].

In order to examine the discriminant validity, we ran an independent sample t test between the scores of the GS test, TUG test, and SPPB obtained in our sample with the normative values established (in the literature) for community-dwelling, age-matched healthy controls for the GS test [6], the TUG test [21], and the SPPB [4]. p values of less than 0.05 were considered significant and indicative of sufficient discriminant validity.

Last, we extracted the GS test scores from the SPPB (expressed as meters per second over 4 m) to determine whether they were comparable to those obtained in the separate, longer (10 m) version. If the GS test scores obtained from the SPPB have reliability comparable to that of the 10-m version, an argument could be made that the GS component of the SPPB can serve a dual purpose, being factored in the SPPB summary score and being an independent estimate of GS (to be used in related inferences from that score). ICCs were used to examine the reliability between these two GS scores under the hypothesis that both GS scores would have good reliability (an ICC of greater than 0.75) [22]. In addition, we examined whether there were any systematic differences between these two GS scores using a Bland–Altman plot: the differences in the GS scores were plotted on the y-axis, and the average (mean) GS scores for the sample were plotted on the x-axis [5], with limits of agreement (LOAs) shown as the mean difference in GS test score ($\pm$ 2 SD). The graph shows the extent of agreement between the two GS test scores across the sample and whether participants scored consistently higher on one than on the other.

The data analyses were completed using IBM SPSS software, version 22.0 (Armonk, NY, USA).

Results

A total of 44 participants (age, 66.9 ± 8.1 years; 27 women, 17 men) consented to participate and completed the testing (Table 1). Given that both of the studies were cross-sectional in nature, we had no missing data or dropout. The data for the SPPB were available for 32 participants. The data for all the continuous variables met the assumptions of normality, with the exception of duration of knee pain.

The concurrent construct validity of the physical performance measures was variable. The relationships of the TUG test with the SPPB and its component tests were moderate to high ($r \geq 0.50$), which was consistent with hypothesis 1. However, the relationships of the GS test with other measures, including the GS component of the SPPB, were often low ($r < 0.50$), which was in contrast to hypothesis 1. We also examined the concurrent validity of the 10-m GS test with the scores extracted from the SPPB. This relationship was clearly convergent, with r values of 0.86 ($p < 0.0001$) (data not shown). Consistent with hypothesis 2, the physical performance measures showed poor convergent validity ($r < 0.50$), with self-reported measures of the KOOS-PS and Q-NPRS. Pearson correlation coefficients demonstrating the concurrent relationships between the KOOS-PS, the Q-NPRS, the GS test, the TUG test, and SPPB tasks are shown in Table 2.

The physical performance measures were found to have discriminant validity (Table 3). The mean scores of the physical performance measures obtained in the current study were compared with pre-established normative values for age-matched older adults. Because the normative scores for the GS test [6] and the TUG test [21] were given for women and men separately without a pooled value for adults between 60 and 69 years, we considered the worse score on these tests, to minimize the possibility of a type 1 error. The results suggest that the subjects with advanced knee OA in our study had worse GS test, TUG test, and SPPB scores, as compared with the age-matched norms, indicating worse physical performance.

The results of the 10-m GS test and the GS value extracted from the SPPB (4-m distance) were reliably

Table 1 Characteristics of participants (N = 44)

Variable	
Age, years (± SD)	66.9 ± 8.1
Sex	
No. women (%)	27 (61.3%)
No. men (%)	17 (38.7%)
Side of involvement	
No. right (%)	21 (47.7%)
No. left (%)	17 (38.7%)
No. both (%)	6 (13.6%)
Duration of knee pain in months (± SD)	18.9 ± 21.5
BMI, mean (± SD)	33 ± 6.1
KOOS-PS score, mean (± SD)	54.6 ± 15.6
Q-NPRS, mean score (± SD)	4.7 ± 2
TUG test time in seconds, mean (± SD)	12.8 ± 3.8
GS score, meters per second (± SD)	1.08 ± 0.32
SPPB total score (n = 32), mean (± SD)	8.94 ± 2.26
Balance, mean (± SD)	3.44 ± 0.95
Gait speed in seconds, mean (± SD)	3.69 ± 0.64
Chair stand in seconds, mean (± SD)	1.81 ± 1.23

SD standard deviation, *BMI* body mass index (weight in kilograms divided by the square of the height in meters), *KOOS-PS* Knee Injury and Osteoarthritis Outcome Score—Physical Function Shortform, *Q-NPRS* quadruple numeric pain rating scale, *TUG* Timed Up and Go, *GS* gait speed, *SPPB* Short Physical Performance Battery

similar; the ICC values were 0.81 (95% confidence interval, 0.64 to 0.90). This validated our hypothesis that GS values obtained over the shorter distance would provide results comparable to those obtained in the 10-m version. In the Bland–Altman plot, showing the agreement between the two GS tests (Fig. 1), the mean score for the GS extracted from the SPPB was higher by 0.06 m per second, as compared with that of the GS obtained from the 10-m version. In addition, there were outliers—not contained within the LOAs—on both sides. This suggests that there is no pattern in which one version of the GS test definitively scores higher than the other.

Discussion

Physical performance measures provide a look into measurements of functional status in patients with advanced knee OA that self-reported measures are unable to capture [38],

despite their use being recommended [11]. This study provides a further understanding of the measurement properties of some of the commonly used physical performance measures in advanced knee OA. Importantly, this study substantiates the concurrent validity of the TUG test and GS test, although it provides only preliminary evidence concerning the concurrent validity of the SPPB in this population. The study also demonstrated that patients with advanced knee OA had significantly slower GS and worse scores on the TUG test and SPPB than did age-matched controls. This finding not only indicates deficits in physical performance in those with advanced knee OA but also shows that the GS test, TUG test, and SPPB have good discriminant validity. However, these findings should be viewed in the context of the fact that sample size estimates were not conducted.

A few limitations to our study should be acknowledged. First, we did not calculate the requisite sample size needed for conducting the analyses and testing our study hypotheses. Our study was exploratory in nature, and we recommend further research to substantiate our results using larger sample sizes. Second, we examined selected measurement properties of concurrent and discriminant validity but did not provide an assessment of other properties, such as reproducibility, measurement error, or responsiveness, for these measures. One study is not adequate to indicate integration of any measure into clinical practice, and building evidence across different studies over time facilitates such integration. To this end, we do recommend future research that assesses important attributes such as reproducibility and responsiveness in subjects with advanced knee OA. Finally, our results are specific to those with advanced knee OA and may not be relevant to patients with early knee arthritis or in those who undergo TKR.

Several variants of the GS test have been described in the literature that vary in the participants' walking distance and speed (comfortable gait speed [CGS] or fast gait speed [FGS]). Some of the common variants are the 10-m version [36], a 40-m FGS version recommended by the OARSI [11], and GS assessed over a 3-m and 4-m path [15]. We chose the 10-m version rather than the longer or shorter versions for two reasons. First, normative values have been established for the 10-m version, which will provide context for future assessments [36]. Second, we wanted a balance between high accuracy and pragmatism wherein the administrative

Table 2 Concurrent validity (Pearson correlation coefficients) between outcome measures

	Q-NPRS	TUG test	GS test	SPPB total	SPPB balance	SPPB gait	SPPB chair stand
KOOS-PS	− 0.716[a]	0.025	0.115	0.308	0.128	0.419[b]	0.246
Q-NPRS		0.168	− 0.199	− 0.406[b]	0.286	− 0.432[b]	0.299
TUG test			− 0.324[b]	− 0.710[a]	− 0.536[a]	− 0.688[a]	− 0.529[a]
GS test				0.518[a]	0.324	0.442[b]	0.470[a]
SPPB total					0.782[a]	0.696[a]	0.867[a]
SPPB balance						0.389[b]	0.460[a]
SPPB gait							0.453[a]

Q-NPRS quadruple numeric pain rating scale, *TUG* Timed Up and Go, *GS* gait speed, *SPPB* Short Physical Performance Battery, *KOOS-PS* Knee Injury and Osteoarthritis Outcome Score—Physical Function Shortform
[a] Correlations are significant at the 0.01 level (two-tailed)
[b] Correlations are significant at the 0.05 level (two-tailed)

Table 3 Discriminant validity of the physical performance measures

| Tests	Our sample	Established norms	p value
Gait speed (m/s)	1.08 ± 0.32	1.29 ± 0.21	$p = 0.003$
TUG test (s)	12.8 ± 3.8	11.4 ± 2.4	$p = 0.021$
SPPB	8.94 ± 2.26	10.5 ± 1.6	$p = 0.002$

m/s meters per second, *TUG* Timed Up and Go, *SPPB* Short Physical Performance Battery

burden was optimal for the test. The 10-m GS test has been shown to have superior construct validity over the shorter version of the GS test [31]. In addition, we believe that the 10-m GS test will be more easily and quickly implemented than the 40-m FGS version. Our results show only low-to-moderate correlations of the GS with the component tests of the SPPB, as well with as the TUG test. However, the relationship between the GS test and the TUG test has been shown to be high in individuals with multiple sclerosis ($r = 0.90$) [33]. The validity is context dependent, and although the relationship between the GS and TUG tests was low in our study ($r = -0.42$), the possibility that this result could be a function of the sample recruited cannot be ruled out. We were not able to locate earlier research in which the concurrent validity of the 10-m version of GS test was compared with either the SPPB or the TUG test; it is therefore difficult to contextualize our results. The performance on the SPPB is scored on an ordinal scale of 0 to 4, with 4 being the best performance, rather than the actual performance in actual units, such as meters per second for the GS component or the number of repetitions for the sit-to-stand component. It is possible that the low-to-moderate correlations we found between the 10-m GS test scores and the aggregate scores of the SPPB were a result of the diverse scoring approaches.

This was partly substantiated in our study in that the relationship between the scores of the 10-m GS and the 4-m GS extracted from the SPPB, also scored in meters per second, was high ($r = 0.86$).

The finding that the scores of the GS test, TUG test, and SPPB were significantly different between our sample and the established norms [4, 6, 21] has two important implications. First, it provides support of the discriminant validity of the GS test, TUG test, and SPPB in individuals with advanced knee OA. Second, it is clear that the individuals with advanced knee OA experience poor physical function across important domains of daily activities, such as maintaining a healthy walking speed, maintaining balance during mobility, and rising from a chair. This is consistent with the prevalent understanding that individuals with advanced knee OA demonstrate significantly impaired mobility and balance [18]. In particular, the mean GS test score of 1.08 ± 0.32 m per second observed in our sample clearly suggests that, even when they can walk at a safe speed, people with advanced knee OA are likely to experience significant limitations in the performance of household activities and in essential functions outside the home such as safely crossing the street and carrying groceries [26]. In addition, the mean SPPB score of 8.94 ± 2.26 in our sample was lower than what is recommended (higher than 10) to be able to walk at least 400 m in the community, and patients with advanced knee OA are likely to transition to severe mobility deficits over time [46]. In view of these findings, a preventive approach to managing balance and mobility impairments might be warranted in those who have advanced knee OA.

The scores extracted from the GS component of the SPPB were reproducible when compared with the scores of 10-m GS test (ICC = 0.81). Our findings provide further support to the previously reported result of Peters et al. [31], in which GS
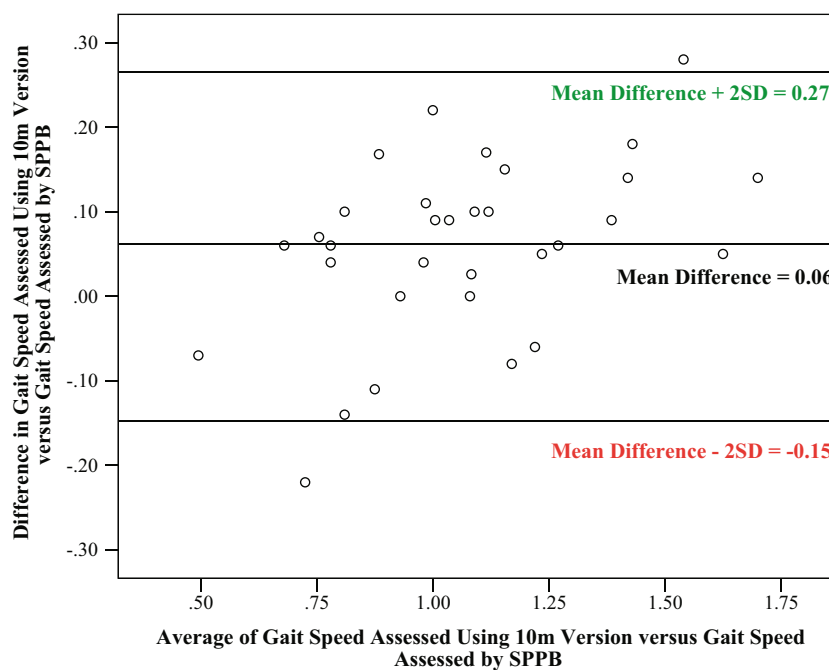
Fig. 1. Bland–Altman plot showing agreement between the scores of 10-m gait speed and 4-m gait speed extracted from the Short Physical Performance Battery (SPPB)

assessed over both distances had comparable reliability matrices, albeit in healthy older adults. However, our results diverge from other findings: the Bland–Altman plot in our study revealed a smaller discrepancy of 0.06 m per second between the two versions of the GS test; in Peters et al., there was a difference of 0.15 to 0.17 m per second [31]. A simple paired t test between the mean values of two the GS test scores in our study also showed no differences ($p = 0.69$). Last, the relationship between the two GS test scores in our study was high ($r = 0.86$). All of these ancillary analyses suggest that in patients with advanced knee OA, the shorter version of the GS test is reproducible and provides a valid estimate of the GS, as compared with that seen with the 10-m version. Measures that have a low administrative burden yet provide a reliable estimate of an important domain, in this instance GS, are preferred by clinicians. Clinicians could extract the scores of a GS test in meters per second when they administer the SPPB without having to administer an additional 10-m GS test, which reduces the burden of testing while still successfully providing a measurement of an important separate domain.

In conclusion, the results of our research advance the understanding of the validity of the GS test, TUG test, and SPPB in patients with advanced knee OA. In addition, they indicate that those with advanced knee OA experience marked impairments in physical performance in important activities such as walking, balancing, and rising from a chair, as compared with their age-matched healthy counterparts. We recommend future research that increases our understanding of reproducibility and responsiveness of the GS, TUG test, and SPPB in patients with advanced knee OA, in order to facilitate the integration of these measures into clinical practice.

Compliance with Ethical Standards

References

1. Alghadir A, Anwer S, Brismee JM. The reliability and minimal detectable change of Timed Up and Go test in individuals with grade 1–3 knee osteoarthritis. *BMC Musculoskelet Disord.* 2015;16:174.

2. Alghadir AH, Anwer S, Iqbal A, Iqbal ZA. Test–retest reliability, validity, and minimum detectable change of visual analog, numerical rating, and verbal rating scales for measurement of osteoarthritic knee pain. *J Pain Res.* 2018;11:851–856.

3. Arnold JB, Walters JL, Ferrar KE. Does physical activity increase after total hip or knee arthroplasty for osteoarthritis? A systematic review. *J Orthop Sports Phys Ther.* 2016;46(6):431–442.

4. Balasubramanian CK. The community balance and mobility scale alleviates the ceiling effects observed in the currently used gait and balance assessments for the community-dwelling older adults. *J Geriatr Phys Ther.* 2015;38(2):78–89.

5. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat.* 2007;17(4):571–582.

6. Bohannon RW. Comfortable and maximum walking speed of adults aged 20–79 years: reference values and determinants. *Age Ageing.* 1997;26(1):15–19.

7. Davis AM, Perruccio AV, Ibrahim S, et al. The trajectory of recovery and the inter-relationships of symptoms, activity and participation in the first year following total hip and knee replacement. *Osteoarthritis Cartilage.* 2011;19(12):1413–1421.

8. Davis AM, Perruccio AV, Canizares M, et al. The development of a short measure of physical function for hip OA HOOS-Physical Function Shortform (HOOS-PS): an OARSI/OMERACT initiative. *Osteoarthritis Cartilage.* 2008;16(5):551–559.

9. Dobson F, Hinman RS, Hall M, et al. Reliability and measurement error of the Osteoarthritis Research Society International (OARSI) recommended performance-based tests of physical function in people with hip and knee osteoarthritis. *Osteoarthritis Cartilage.* 2017;25(11):1792–1796.

10. Dobson F, Hinman RS, Hall M, Terwee CB, Roos EM, Bennell KL. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. *Osteoarthritis Cartilage.* 2012;20(12):1548–1562.

11. Dobson F, Hinman RS, Roos EM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. *Osteoarthritis Cartilage.* 2013;21(8):1042–1052.

12. Dunlop DD, Song J, Semanik PA, Sharma L, Chang RW. Physical activity levels and functional performance in the osteoarthritis initiative: a graded relationship. *Arthritis Rheumatol.* 2011;63(1):127–136.

13. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. *Int J Endocrinol Metabolism.* 2012;10(2):486–489.

14. Gignac MA, Backman CL, Davis AM, et al. Understanding social role participation: what matters to people with arthritis? *J Rheumatol.* 2008;35(8):1655–1663.

15. Graham JE, Ostir GV, Fisher SR, Ottenbacher KJ. Assessing walking speed in clinical research: a systematic review. *J Eval Clin Pract.* 2008;14(4):552–562.

16. Guralnik JM, Ferrucci L, Pieper CF, et al. Lower extremity function and subsequent disability: consistency across studies, predictive models, and value of gait speed alone compared with the short physical performance battery. *J Gerontol A Biol Sci Med Sci.* 2000;55(4):M221-M231.

17. Guralnik JM, Simonsick EM, Ferrucci L, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol.* 1994;49(2):M85–M94.

18. Hatfield GL, Morrison A, Wenman M, Hammond CA, Hunt MA. Clinical tests of standing balance in the knee osteoarthritis population: systematic review and meta-analysis. *Phys Ther.* 2016;96(3):324–337.

19. Hawker GA, Stewart L, French MR, et al. Understanding the pain experience in hip and knee osteoarthritis—an OARSI/OMERACT initiative. *Osteoarthritis Cartilage.* 2008;16(4):415–422.

20. Hiyama Y, Asai T, Wada O, et al. Gait variability before surgery and at discharge in patients who undergo total knee arthroplasty: a cohort study. *PloS One.* 2015;10(1):e0117683.

21. Ibrahim A, Singh DKA, Shahar S. "Timed Up and Go" test: age, gender and cognitive impairment stratified normative values of older adults. *PloS One.* 2017;12(10):e0185641.

22. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15(2):155–163.

23. Marcum ZA, Zhan HL, Perera S, Moore CG, Fitzgerald GK, Weiner DK. Correlates of gait speed in advanced knee osteoarthritis. *Pain Med.* 2014;15(8):1334–1342.

24. Medina-Mirapeix F, Bernabeu-Mora R, Llamazares-Herran E, Sanchez-Martinez MP, Garcia-Vidal JA, Escolar-Reina P. Inter-observer reliability of peripheral muscle strength tests and short physical performance battery in patients with chronic obstructive pulmonary disease: a prospective observational study. *Arch Physical Med Rehabil.* 2016;97(11):2002–2005.

25. Mehta SP, Sankar A, Venkataramanan V, et al. Cross-cultural validation of the ICOAP and physical function short forms of the HOOS and KOOS in a multi-country study of patients with hip and knee osteoarthritis. *Osteoarthritis Cartilage.* 2016;24(12):2077–2081.

26. Middleton A, Fritz SL, Lusardi M. Walking speed: the functional vital sign. *J Aging Phys Act.* 2015;23(2):314–322.

27. Mukaka MM. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. *Malawi Med J.* 2012;24(3):69–71.

28. Olsen CF, Bergland A. Reliability of the Norwegian version of the short physical performance battery in older people with and without dementia. *BMC Geriatr.* 2017;17(1):124.

29. Ortega-Perez de Villar L, Martinez-Olmos FJ, Junque-Jimenez A, et al. Test–retest reliability and minimal detectable change scores for the short physical performance battery, one-legged standing test and timed up and go test in patients undergoing hemodialysis. *PloS One.* 2018;13(8):e0201035.

30. Perruccio AV, Stefan Lohmander L, Canizares M, et al. The development of a short measure of physical function for knee OA KOOS-Physical Function Shortform (KOOS-PS)—an OARSI/OMERACT initiative. *Osteoarthritis Cartilage.* 2008;16(5):542–550.

31. Peters DM, Fritz SL, Krotish DE. Assessing the reliability and validity of a shorter walk test compared with the 10-Meter Walk Test for measurements of gait speed in healthy, older adults. *J Geriatr Phys Ther.* 2013;36(1):24–30.

32. Sankar A, Davis AM, Palaganas MP, Beaton DE, Badley EM, Gignac MA. Return to work and workplace activity limitations following total hip or knee replacement. *Osteoarthritis Cartilage.* 2013;21(10):1485–1493.

33. Sebastião E, Sandroff BM, Learmonth YC, Motl RW. Validity of the Timed Up and Go Test as a measure of functional mobility in persons with multiple sclerosis. *Arch Phys Med Rehabil.* 2016;97(7):1072–1077.

34. Segal NA, Yack HJ, Brubaker M, Torner JC, Wallace R. Association of dynamic joint power with functional limitations in older adults with symptomatic knee osteoarthritis. *Arch Phys Med Rehabil.* 2009;90(11):1821–1828.

35. Sharma L, Cahue S, Song J, Hayes K, Pai YC, Dunlop D. Physical functioning over three years in knee osteoarthritis: role of psychosocial, local mechanical, and neuromuscular factors. *Arthritis Rheumatol.* 2003;48(12):3359–3370.

36. Steffen TM, Hacker TA, Mollinger L. Age- and gender-related test performance in community-dwelling elderly people: Six-Minute Walk Test, Berg Balance Scale, Timed Up & Go Test, and gait speeds. *Phys Ther.* 2002;82(2):128–137.

37. 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):733–746.

38. Stratford PW, Kennedy DM. Performance measures were necessary to obtain a complete picture of osteoarthritic patients. *J Clin Epidemiol.* 2006;59(2):160–167.

39. Stratford PW, Kennedy DM, Maly MR, Macintyre NJ. Quantifying self-report measures' overestimation of mobility scores postarthroplasty. *Phys Ther.* 2010;90(9):1288–1296.

40. Stratford PW, Kennedy DM, Woodhouse LJ. Performance measures provide assessments of pain and function in people with advanced osteoarthritis of the hip or knee. *Phys Ther.* 2006;86(11):1489–1496.

41. Takacs J, Garland SJ, Carpenter MG, Hunt MA. Validity and reliability of the community balance and mobility scale in individuals with knee osteoarthritis. *Phys Ther.* 2014;94(6):866–874.

42. Tasci Bozbas G, Sendur OF, Aydemir AH. Primary knee osteoarthritis increases the risk of falling. *J Back Musculoskelet Rehabil.* 2017;30(4):785–789.

43. Tsonga T, Michalopoulou M, Kapetanakis S, et al. Risk factors for fear of falling in elderly patients with severe knee osteoarthritis before and one year after total knee arthroplasty. *J Orthop Surg (Hong Kong).* 2016;24(3):302–306.

44. Umehara T, Tanaka R. Effective exercise intervention period for improving body function or activity in patients with knee osteoarthritis undergoing total knee arthroplasty: a systematic review and meta-analysis. *Braz J Phys Ther.* 2018;22(4):265–275.

45. van Leeuwen DM, van de Bunt F, de Ruiter CJ, van Schoor NM, Deeg DJH, Emanuel KS. Functioning without cartilage: older people with radiographic knee osteoarthritis who self-report no functional limitations do score lower on a performance battery. *J Aging Phys Act.* 2017;25(4):570–575.

46. Vasunilashorn S, Coppin AK, Patel KV, et al. Use of the Short Physical Performance Battery Score to predict loss of ability to walk 400 meters: analysis from the InCHIANTI study. *J Gerontol A Biol Sci Med Sci.* 2009;64A(2):223–229.

47. Waller B, Munukka M, Rantalainen T, et al. Effects of high intensity resistance aquatic training on body composition and walking speed in women with mild knee osteoarthritis: a 4-month RCT with 12-month follow-up. *Osteoarthritis Cartilage.* 2017;25(8):1238–1246.