Published in final edited form as:
*PM R*. 2020 January ; 12(1): 16–25. doi:10.1002/pmrj.12160.

# Using clinical balance tests to assess fall risk among established unilateral lower limb prosthesis users: cutoff scores and associated validity indices

**Andrew Sawers, PhD, CPO**[1], **Brian Hafner, PhD**[2]

[1]Department of Kinesiology and Nutrition, University of Illinois at Chicago, Chicago, IL 60612

[2]Department of Rehabilitation Medicine, University of Washington, Seattle, WA 98195

## Abstract

**Introduction:** Clinicians are routinely required to make decisions about fall risk among lower limb prosthesis (LLP) user. These decisions can be guided by standardized clinical balance tests, but require population- and test-specific cutoff scores and validity indices to categorize individuals as probable fallers or non-fallers on the basis of test performance. Despite the importance of cutoff scores and validity indices to clinical interpretation of clinical balance test scores, they are rarely reported for LLP users. In their absence, clinicians cannot use results from clinical balance tests to assess the likelihood of a fall by any one patient.

**Objective:** Derive cutoff scores, and associated validity indices, for clinical balance tests administered to established unilateral LLP users.

**Design:** Cross-sectional study.

**Setting:** Outpatient clinic and research laboratory.

**Participants:** Established ambulatory unilateral transtibial and transfemoral prosthesis users (n=40).

**Intervention:** Not applicable.

**Main Outcome Measure(s):** Optimal cutoff scores and related validity indices (i.e., area under the curve, sensitivity, specificity, likelihood ratios) were computed for five balance tests, the Activities-Specific Balance Confidence Scale (ABC), Timed Up and Go (TUG), Four Square Step Test (FSST), Berg Balance Scale (BBS), and Narrowing-Beam Walking Test (NBWT).

**Results:** Cutoff scores were identified for the NBWT (≤.43/1.0), TUG (≥8.17 seconds), FSST (≥8.49 seconds), BBS (≤50.5/56), and ABC (≤80.2/100). Validity indices (i.e., area under the

---

**Corresponding Author:** Andrew Sawers, PhD, Department of Kinesiology and Nutrition, University of Illinois at Chicago, 1919 W Taylor Street, Chicago, IL 60612, Phone: 312-996-1427, Fax: 312-413-0319, asawers@uic.edu.

Declaration of Conflicting Interest
The authors declare that there is no conflict of interest.

**Prior Publication**: The work described in this manuscript has not been previously published, nor is it under consideration for publication elsewhere.

**Financial Disclosure**: The authors declare no competing interest.

**Institutional Review**: All study procedures were reviewed and approved by a XXXX institutional review board.

curve, sensitivity, specificity, and likelihood ratios) for the NBWT, TUG, and FSST had greater diagnostic accuracy and provided more information about the probability of a fall than those for the BBS or ABC.

**Conclusion:** Performance above or below identified cutoff scores for the NBWT, FSST, and TUG provides information about potentially important shifts in the probability of falling among established unilateral LLP users. These results can serve as initial benchmarks to reduce uncertainty surrounding fall risk assessment in established unilateral LLP users, but require future prospective evaluation.

**Keywords**

amputee; accidental falls; patient outcomes assessment; rehabilitation; validity

## Introduction

Clinicians are routinely required to make important decisions about whether a patient may be at risk for an adverse condition or event. Diagnostic decisions such as these are ideally made using psychometrically sound clinical tests with which the probability of an event or condition occurring or existing can be estimated[1]. Establishing population- and test-specific cutoff scores and validity indices is key to applying clinical tests in this manner. Cutoff scores are required to dichotomize continues scales and create benchmarks that categorize individuals as with or without the condition or event on the basis of a "positive" or "negative" test result[2]. Validity indices (e.g., sensitivity, specificity, and likelihood ratios) serve to provide information on the probability of the event or condition. In the absence of cutoff scores and associated validity indices, clinical tests can be administered to evaluate change over time or differences between individuals, but cannot be used to assess the probability with which an adverse event or condition may occur.

Falls remain a frequent event that negatively affects the lives of a substantial portion of lower limb prosthesis (LLP) users[3,4]. Over 50% of LLP users report falling at least once a year, with up to 39% reporting multiple falls a year[4–8]. Falls among LLP users frequently lead to adverse health outcomes including injury[3,5,7,9,10], financial costs[11], reduced mobility[12,13], and diminished quality of life[14]. A major barrier to reducing falls among LLP users has been effective screening of those at risk[15,16]. Central to this barrier is a scarcity of LLP user-specific cutoff scores and validity indices for contemporary clinical balance instruments such as the Berg Balance Scale (BBS)[17], Timed Up and Go (TUG)[18], Four Square Step Test (FSST)[19], and Activities-specific Balance Confidence (ABC) scale[20]. A number of studies have assessed the validity and reliability of these instruments among LLP users[15,16,21–27], but cutoff scores to infer fall risk have only been established for transtibial prosthesis users with limited prosthetic experience (i.e., 6 months of prosthetic use)[21]. Clinicians therefore, lack the necessary information to assess fall risk among individual patients within the larger LLP user population (i.e., established transtibial and transfemoral prosthesis users).

Two different faller classifications, either ≥1 fall[3,16,28,29] or ≥2 falls[3,15,21] have been used to assess the diagnostic accuracy and validity indices of performance-based clinical balance tests. Validity indices for one performance-based balance test, the Berg Balance Scale (BBS), increased when a classification of ≥ 2 falls was used in a prior study[8]. Adopting a ≥2 falls classification inherently creates a “faller” group that is likely to possess worse balance ability than if it also included individuals with 1 fall (i.e., ≥1 fall classification). As a result, it may be easier to identify fallers versus non-fallers, and possibly make it more difficult to identify people with moderate balance impairments who may be at risk for falls. Despite both faller classifications being used in research and clinical settings, no study of LLP users has systematically evaluated the influence of faller classification on validity indices associated with performance-based clinical balance tests commonly administered to established unilateral LLP users.

The objective of this study was therefore to establish cutoff scores, and associated validity indices, for clinical balance tests administered to established unilateral LLP users. In prior studies it was found that single-task tests like the FSST, TUG, or Narrowing Beam Walking Test (NBWT) demonstrated better discriminant validity than multi-item tests like the BBS, ABC, or the Locomotor Capabilities Index[16,21]. Based on this prior research therefore, it was hypothesized that the validity indices (e.g., likelihood ratios) associated with cutoff scores for the FSST, TUG, and the NBWT would exceed those of the BBS or ABC scale. Confirmation of this hypothesis would offer further evidence that clinical tests like the NBWT, FSST, and TUG can be used to assess fall risk among established unilateral transtibial and transfemoral prosthesis users. Further, cutoff scores and related validity indices would provide clinicians with the information needed to appropriately interpret clinical balance test scores to assess fall risk in their patients who use LLPs. Owing to the variety of ways researchers and clinicians can classify individuals as fallers or non-fallers[3,15,16,21], a secondary objective of this study was to determine whether cutoff scores and associated validity indices differed across two common fall classifications (i.e., ≥ 1 fall versus ≥ 2 falls over the past 12 months). It was hypothesized that classifying fallers as ≥ 2 falls would increase validity indices[8].

## Methods

### Study Design

A cross-sectional study was performed from July 2016 to May 2017. The STROBE (Strengthening The Reporting of Observational studies in Epidemiology) Statement guideline[30] was followed during the collection and reporting of study data. All data were stored and managed using a REDCap database hosted at XXXX[31]. Study protocols were reviewed and approved by a XXXX institutional review board. Study participants provided written informed consent prior to participation.

### Participants

Lower limb prosthesis (LLP) users were recruited from local prosthetic clinics. Inclusion criteria included unilateral transtibial or transfemoral amputation due to traumatic, dysvascular, or oncologic causes; 18 years of age or older; one or more years of using a prosthesis (i.e., established users); ability to ambulate at least 10 feet without an upper extremity assistive device (e.g., cane), and use of a comfortable prosthesis (assessed with the

Socket Comfort Scale)[32]. Participants with complications to their contralateral leg (e.g., joint replacement, arthritis, or wounds), amputation of a second limb, or a neurological or cardiovascular condition that limited the ability to complete the study protocol were excluded.

## Procedures

Participants completed demographic and prosthetic-related characterization measures, as well as a retrospective falls survey. Wearing their preferred prosthesis-footwear combination, participants were administered five clinical balance tests. Five-minute rest periods were enforced between each balance test. Cutoff scores and validity indices were derived using recommended methods[1,33–35].

## Measurements

**Participant demographic and characterization measures—**Age, height, weight, and sex were collected from study participants via self-report. Medicare Functional Classification Level (MFCL) (i.e., K-level) was determined by a certified prosthetist via interview and physical evaluation, while amputation-related information (i.e., level, etiology, and time since) were obtained via interview with a study investigator. Perceived mobility was assessed with the Prosthetic Limb Users Survey of Mobility (PLUS-M)[36]. To ascertain the number of falls experienced by each participant over the past 12 months, participants were asked, "In the past year have you had any falls including a slip or a trip in which you inadvertently lost your balance and landed on the ground or lower level?"[37–40] Participants who reported falling in the past year were then asked to recall the number of falls in the past 12 months. To determine whether different faller classifications influence cutoff scores and related validity indices, data was analyzed using two faller classifications: ≥1 fall[3,16,28,29] and ≥2 falls[3,15,21] over the past 12 months.

**Clinical balance tests—**Five clinical balance instruments, the Activities-Specific Balance Confidence (ABC) Scale[20], the Timed Up and Go (TUG)[18], the Four Square Step Test (FSST)[19], the Berg Balance Scale (BBS)[17], and the Narrowing-Beam Walking Test (NBWT)[16] were administered and scored according to the developers' instructions (Appendix 1). Each balance instrument has demonstrated acceptable levels of validity[15,16,21–23] and/or reliability (i.e. ICC range .70 – .99)[15,22,23,25–27,41] among LLP users.

## Statistical Analysis

Clinical balance test scores were compared across participant subgroups defined by their reported fall history (i.e., 0 falls, 1 fall, or ≥ 2 falls), and within commonly adopted fall classifications (i.e., 0 falls versus ≥ 1 fall, or 0–1 falls vs. ≥ 2 falls) using parametric (i.e., ANOVA) or non-parametric equivalents (Kruskal-Wallis Test).

Receiver Operating Characteristic (ROC) curves were used to identify cutoff scores[33,42] and related validity indices (i.e., area under the curve, sensitivity, specificity, and likelihood ratios) for each clinical balance test. ROC curves were obtained by plotting the sensitivity of a test against 1-specificity. The area under the curve (AUC) represents the probability of correctly identifying a faller from a randomly selected pair of lower limb prosthesis users, one being a faller and the other a non-faller[43]. The larger the AUC, the greater the test's general discriminative ability[43]. The AUC was selected as a summary measure of diagnostic accuracy[44]. The AUC can assume any value between 0 and 1, with a value of .5 representing chance[2], and values greater than or equal to .8 recommended as the limit of clinical acceptability[45]. Areas for each test were therefore compared to a threshold of .8 to determine the clinical acceptability of a test. While the AUC provides insight into the overall discriminant ability of a test based on group data, it does not provide clinicians with actionable information when assessing an individual patient. Cutoff scores and validity indices such as likelihood ratios are required to make individual-level decision based on test scores.

Optimal cutoff scores were chosen for each balance test by selecting the point on a test's ROC curve closest to the (0,1) point[46]. This point was identified by choosing the minimal value of the function $(1\text{-sensitivity})^2 + (1\text{-specificity})^2$, a method that maximizes sensitivity and specificity[46,47]. The test score corresponding to this minimal value was therefore selected as the cutoff score that best discriminates between unilateral LLP users with and without a history of falls[2]. Values for sensitivity (i.e., how often a diagnostic test is positive when a condition is present), and specificity (i.e., how often a diagnostic test is negative in the absence of the condition) at that point were recorded.

One issue with sensitivity and specificity indices is that they have limited clinical utility[48]. Sensitivity, for example, does not describe how often patients with positive tests have the condition of interest (i.e., a fall). Sensitivity indices therefore do not indicate the probability of the event or condition occurring[49], information that is likely to be of greatest value to clinicians[1]. Predictive values alternatively offer probabilities of an event occurring, but are dependent on the prevalence of the event in the study sample and thus rarely generalize beyond the study[50]. Likelihood ratios overcome these aforementioned limitations by quantifying how much the obtained test score increases or decreases the probability of an event occurring[1,35], independent of its prevalence in the sample (i.e., they generalize beyond the study). Likelihood ratios are therefore considered more efficient and clinically useful than sensitivity and specificity values or positive and negative predictive values[34,35,51,52]. Unlike the AUC, likelihood ratios provide clinicians with actionable information for individual patients (i.e., the probability a patient will or will not fall based on a positive or negative test outcome). Likelihood ratios were therefore selected as the primary validity index in this study. The likelihood ratio for a positive test (LR+) was computed as sensitivity/1-specificity, while the likelihood ratio for a negative test (LR-) was computed as 1-specificity/sensitivity[49]. Likelihood ratios greater than 5 or less than 0.2 indicate moderate changes in the probability of an event (e.g., fall), while likelihood ratios between 5 and 2, or 0.5 and 0.2 result in small, but potentially important shifts in probability[1]. Likelihood ratios greater than 2 or less than 0.2 were considered the minimum for considering a clinical test able to contribute to a fall risk assessment.

A multivariate logistic regression model was developed for each faller classification (i.e., ≥ 1 fall = Model 1, ≥ 2 falls = Model 2) to determine whether a combination of clinical balance test scores and other fall-related variables could enhance validity indices beyond the score from any single test. Univariate logistic regression was first performed with each potential predictor variable (e.g., FSST time, age, sex, etiology, level of amputation, mobility level) as the independent variable, and faller status as the dependent variable. Independent variables found to have an odds ratio of ≥2 or <0.5, and a p-value of <0.05 were retained for further consideration[53]. To avoid multicollinearity, any of the independent variables retained from the univariate logistic regression that had a strong correlation (i.e., Spearman rho ≥ .75) and a lower odds ratio than the other independent variable with which it was correlated were excluded from the multivariate logistic regression model[53]. A multivariate logistic regression was then performed with all remaining independent variables included. Validity indices described above were computed from the regression output. Cutoff scores and validity indices were computed for each of the faller classifications (i.e., ≥1 fall and ≥ 2 falls). 95% confidence intervals (CI) were computed to determine the precision of each validity index[52]. All statistical analyses were conducted using SPSS v.25 (Chicago, IL).

## Results

The sample included 40 established unilateral lower limb prosthesis (LLP) users (height: 173 ±9.10cm; weight: 78.5 ±14.1kg; SCS: 7.8 ±1.4) (Table 1). All participants were able to perform and compete all tests. Scores for each of the studied tests, stratified by participant fall history and fall classification are presented in Table 2 and Table 3 respectively. Pairwise comparisons revealed that with the exception of the ABC, participants with a history of 2 or more falls had significantly worse test scores than those with a history of either one fall or no falls (p < .027 to .004) (Table 2). There were no statistically significant differences in test scores between participants who reported one and no falls. When a fall classification scheme of 2 or more versus 0–1 falls in the past 12 months was implemented, pairwise comparisons revealed that again with the exception of the ABC, participants reporting 2 or more falls had worse test scores than those reporting 0–1 falls (Table 3). With the exception of the NBWT, these differences in test scores between groups were not observed when a fall classification scheme of 1 or more versus zero falls was implemented (Table 3).

Cutoff scores and validity indices (estimate, 95% CI) for each clinical balance test are presented in Table 4. When fallers were defined as individuals reporting ≥1 fall in the 12 months prior to assessment, the NBWT had the greatest area under the ROC curve (AUC) (.81, .62–.91), as well as the largest specificity (76%, 54–96%) and positive likelihood ratio (3.0, 1.5–6.9). The TUG had the greatest sensitivity (83%, 68–98%), and smallest negative likelihood ratio (.24, .13–.56). The logistic regression model developed for the ≥1 fall classification (i.e., Model 1) included the NBWT, FSST, PLUS-M, and amputation level as predictor variables. Many, but not all, validity indices improved slightly compared to individual clinical balance test scores alone (Table 4). Other candidate predictor variables (e.g., TUG, ABC, MFCL, amputation etiology, age and sex) failed to meet model inclusion criteria (i.e., odds ratio ≥ 2 or less than 0.5, and p < 0.05), during univariate logistic regression.

When fallers were defined as individuals reporting ≥2 falls in the 12 months prior to assessment validity indices (i.e., sensitivity, specificity, and likelihood ratios) associated with each of the clinical balance tests generally improved (Table 4). The NBWT again had the largest AUC (0.89, .78–1.0). The FSST, however, exhibited the largest sensitivity (94%, 83–100%) and smallest negative likelihood ratio (.08, .012–.54), while the TUG had the greatest specificity (83%, 67–98%) and positive likelihood ratio (4.7, 1.9–11.9). The logistic regression model developed for the ≥2 falls classification (i.e., Model 2) included several predictor variables consistent with those included in Model 1 (i.e., NBWT, PLUS-M), and others that were unique to Model 2 (i.e., TUG, MFCL, and amputation etiology). Model 2 did not improve validity indices as markedly as Model 1 did (Table 2). Model 2 was not among the top three results for several of the validity indices (i.e., sensitivity and negative likelihood ratio).

## Discussion

The objective of this study was to establish cutoff scores and associated validity indices for several clinical balance tests administered to established unilateral lower limb prosthesis (LLP) users. Results supported the hypothesis that validity indices, including likelihood ratios, of the NBWT, FSST, and TUG exceeded those for the BBS or ABC scale. Performance above or below identified cutoff scores for the NBWT, FSST, and TUG appear to provide information about small but potentially important shifts in the probability of falling among unilateral LLP users. The distribution of demographic (e.g., age), amputation (e.g., level, etiology, and time since), and activity characteristics (e.g., PLUS-M, MFCL) of participants in the present study were generally comparable to those reported in large national studies of people with lower limb amputation (n=210 to 1568)[24,54–58], although skewed slightly towards a higher percentage of individuals with non-dysvascular and transtibial amputation. The overall similarity, however, with these samples suggests that the current results can be generalized to the broader population of established unilateral LLP users. Although these results can serve as initial benchmarks to reduce uncertainty surrounding the assessment of fall risk in established unilateral LLP users, clinicians and researchers should also consider the reliability and ease of use of these instruments when deciding whether to adopt them in their clinical or academic practices. Additionally, prospective evaluation of the cutoff scores and their validity indices will be required in future studies.

This study provides the first set of cutoff scores and validity indices for performance-based clinical balance tests based on data collected from established unilateral transtibial and transfemoral prosthesis users. The current results build on the work of Dite et al., (2007) who established cutoff scores and validity indices for several clinical balance tests (e.g., FSST, TUG) using data from a sample of short-term (i.e., 6 months post-discharge) transtibial prosthesis users (Dite, 2007). Notable differences were observed between the cutoff scores in the present study and those reported by Dite et al., (2007). Dite et al., (2007), used a faller classification of ≥ 2 falls and reported cutoffs of 19 and 24 seconds for the TUG and FSST, respectively. Cutoff times of 9.25 and 8.71 seconds were identified for the TUG and FSST in the present study. Times were even lower (i.e., 8.17 and 8.49 seconds, respectively) if the "≥ 1 fall" classification was used. The discrepancies in cutoff times noted between the two studies may be attributed to the study samples. The present study included both transtibial and transfemoral prosthesis users, while Dite et al., (2007) included only transtibial prosthesis users. However, the higher prevalence[6] and risk of falls[59] among transfemoral prosthesis users would be expected to increase cutoff times for the FSST and TUG in the present study relative to the prior study, not decrease them. Also, previous studies have not reported statistically significant differences in TUG[16,22] or FSST[16] times between transtibial and transfemoral prosthesis users. Thus, inclusion of transtibial and transfemoral prosthesis users in the present study cannot explain the observed differences in cutoff scores. The observed differences in cutoff times are more likely to be attributed to differences in time since amputation. Dite et al., (2007) studied participants less than a year post amputation (i.e., 6.4 ± 1.5 months after discharge from the rehabilitation unit), while participants in the current study averaged 14.3 years since amputation. This difference implies that distinct cutoff scores may be required to evaluate fall risk at different times post amputation, and that balance ability and fall risk may change markedly after six months of prosthesis use. Longitudinal changes of balance and fall risk may be an important area of future research. The results of these two studies should therefore be considered complimentary rather than conflicting, serving two temporally distinct groups; short-term versus established LLP users.

Existing clinical balance tests appear to provide important information about the probability of falls in established unilateral LLP users. Although none of the likelihood ratios for the clinical balance tests in the present study were sufficient to indicate large and conclusive changes in the probability of a fall event, three of the five clinical balance tests (i.e., NBWT, TUG, FSST) had likelihood ratios that would imply small shifts in the probability of a fall given a positive or negative test (Table 4)[55]. For example, a score equal to or greater than 0.43 on the NBWT would suggest that a LLP user is three times more likely (i.e., LR+ = 3.0) of being a faller than a non-faller. Likelihood ratios for the NBWT, TUG, and FSST were accompanied by 95% confidence intervals that did not overlap with 1.0 (i.e., no change in probability) (Table 4), suggesting that the interpretation of an increase or decrease in the probability of being a faller based on a positive or negative test result can be made with a reasonable level of confidence[1]. While the increased likelihood of a fall associated with a positive test on the NBWT, TUG, and FSST relative to the BBS and ABC may be small, it may still be clinically important given the consequences of falls among LLP users[3,5,7,9,11]. Although the LR were greater in the NBWT, TUG, and FSST, in many cases their confidence intervals overlapped. However, the application of the indices derived in this study in a larger prospective study is needed to more definitively determine the impact of these differences.

Notably, fall risk assessment models that included multiple tests and factors associated with fall risk (i.e., etiology and level of amputation)[8,59] failed to improve likelihood ratios among LLP users compared to scores on the individual balance tests (Table 4). No additional information would therefore appear to be gained regarding the probability of being a faller among LLP users by combining scores from multiple balance tests with the other fall-related demographic or amputation information considered in this study. This suggests that amputation-related factors, including level or etiology of amputation, may provide less information regarding fall risk among established unilateral LLP users than has been historically considered. Additional research is required to verify this result in a prospective study, and to consider other falls-related demographic and amputation-specific information. As a result, clinicians may be best served by assessing fall risk in established unilateral LLP users by administering, scoring, and interpreting performance on a single balance test that possesses sufficient psychometric rigor, is practical for the given setting, and meets the application needs of the clinician (i.e., discriminate, evaluate, or predict). Examples of such applications may include discriminating between fallers and non-fallers in an observational study, evaluating changes in balance ability pre-post therapy, and predicting fall risk to justify the prescription of prosthetic componentry.

Results of this study also indicate that the classification used to categorize participants as fallers or non-fallers can alter the validity indices of the studied clinical balance tests. Validity indices, such as the AUC, generally improved when fallers were classified as individuals reporting ≥ 2 falls. The observed increase in validity indices based on a more conservative faller classification (i.e., ≥ 2 falls vs. ≥ 1 fall) is consistent with previous research[8]. However, it seems important to note that limiting the classification of fallers to those with a history of multiple falls inherently creates a group of individuals with worse balance ability than a group that includes those who have fallen just once in the prior 12 months (Table 3). Adopting a more conservative ≥ 2 falls classification magnifies differences between fallers and non-fallers, increases the magnitude of validity indices, and improves each test’s ability to identify those at risk for additional falls. It does so however, at the cost of being able to identify those at risk for a fall after experiencing only one fall in the prior 12 months, potentially overlooking those with modest balance impairments. The observed differences also suggest that studies reporting validity indices for clinical balance tests based on different faller classifications may not be directly comparable. For this reason, we recommend that investigators report cutoff scores and associated validity indices for both single and multiple fallers. Similarly, if a test’s validity indices are derived using one classification, test administrators should apply the same classification when applying the test clinically to assess fall risk. Cutoff scores for the FSST and TUG increased when fallers were defined as LLP users reporting multiple falls (i.e., ≥ 2 falls), compared to users reporting one or more falls. This indicates that, when using these tests, administrators should use fall classification-specific cutoff scores (Table 4). In contrast, cutoff scores for the NBWT, BBS, and ABC did not change with how fallers were classified. This indicates that for these tests, administrators can use the same cutoff score irrespective of how they define a faller. Having a single cutoff score may simplify and reduce the burden of scoring and interpreting test performance, a reported restriction to the adoption of balance tests among clinicians[60–64]. Whether cutoff scores could be established to differentiate LLP users experiencing no falls, a single, or multiple falls remains to be determined.

## Study Limitations

A number of limitations with this study need to be considered when interpreting the results. First, the sample size, while consistent with similar studies[15,21,65], should be increased in future research. A larger sample would facilitate development of cutoff scores to discriminate non-fallers (i.e., zero falls), from single fallers (i.e., 1 fall), and multiple fallers (i.e., ≥ 2 falls), as well as across score intervals for balance tests with a continuous scale (e.g., the NBWT, TUG, and FSST)[1]. The study sample also consisted of a larger percentage of traumatic LLP users (i.e., 62.5%) than is reported in the literature (i.e., 17% −60.2%)[24,54–58], potentially limiting the generalization of study results. While this is a challenge in most[15,23,25,66,67], but not all[8,21,68–70] research involving LLP users, future efforts to include a larger proportion of dysvascular LLP users is required to increase generalization of study results. The mean age of our sample, 48.7, was slightly younger than that reported in larger, national studies of individuals with lower limb amputation (i.e., mean age from 50 to 55) [54–56,58], yet the range of ages studied, 24 to 70, is consistent with those prior studies. A follow-up study that focuses on older LLP users (e.g., age ≥ 65) may be warranted to examine balance test cutoff scores in LLP users in that specific subpopulation (i.e., Medicare-eligible individuals).

Falls were assessed retrospectively. This may underestimate fall frequency[71], and lend itself to recall bias[72,73]. A prospective study is needed to validate the cutoff scores and validity indices established in this cross-sectional study, and establish a temporal relationship between balance and fall status[74].

Fall-related injuries were not recorded. Determining whether multiple versus non-multiple fallers are more likely to suffer a fall-related injury[3,10], or if existing clinical tests can discriminate between or predict the probability of falls that result in injuries may be an important consideration in future studies.

Only a limited set of sociodemographic, health, and prosthetic-related factors were recorded and included in the multivariate models. Including other known risk factors for falls among lower limb prosthesis users such as strength[75], protective stepping[76], number of medications[9], and sense of vibration[77] may improve model performance in future research.

Finally, psychometric properties, including the validity indices reported here, are population specific. The specific indices presented here therefore do not apply to other patient populations.

## Conclusion

The primary objective of this study was to establish cutoff scores and associated validity indices for several clinical balance tests that may be administered to established unilateral lower limb prosthesis (LLP) users. Given the limited options available to quantitatively assess fall risk among established unilateral LLP users, the proposed cutoff scores and associated likelihood ratios for the NBWT, TUG, and FSST provide clinicians with tools to reduce the uncertainty associated with estimating the probability of a fall among established unilateral LLP users. More studies establishing and assessing the accuracy of cutoff scores for diagnostic tests like these to predict outcomes among LLP users are urgently needed. Additional research to evaluate the relative reliability, utility, and prospective validity (i.e., testing the cutoff scores and indices proposed here) of these tests is needed to facilitate their widespread adoption in clinical care.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

**Funding Support:** Research reported in this publication was supported by the Department of Defense (DoD) under award number W81XWH-17-1-0547 (BH, AS), the National Institutes of Health (NIH) under award number K12HD073945 (AS), and the Orthotics and Prosthetics Education and Research Foundation, Inc. (OPERF) under grant number OPERF-SGA-2016-1 (AS, BH). The content is solely the responsibility of the authors and does not necessarily represent the official views of the DoD, NIH, or OPERF.

## Abbreviations:

| | |
|---|---|
| **ABC** | Activities-specific Balance Confidence scale |
| **AUC** | area under the curve |
| **BBS** | Berg Balance Scale |
| **CI** | confidence intervals |
| **FSST** | Four Square Step Test |
| **LR** | likelihood ratio |
| **MFCL** | Medicare Functional Classification Level |
| **NBWT** | Narrowing Beam Walking Test |
| **PLUS-M** | Prosthetic Limb Users Survey of Mobility |
| **ROC** | receiver operating characteristic |
| **STROBE** | Strengthening The Reporting of Observational studies in Epidemiology |
| **TUG** | Timed Up and Go |

## References

1. Riddle DL, Stratford PW. Interpreting validity indexes for diagnostic tests: an illustration using the Berg balance test. Phys Ther 1999;79(10):939–948. [PubMed: 10499967]
2. 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):670–675. [PubMed: 16410346]
3. Wong C, Chihuri S, Li G. Risk of fall-related injury in people with lower limb amputations: A prospective cohort study. J Rehabil Med 2016;48(1):80–85. [PubMed: 26694526]
4. Kim J, Major MJ, Hafner B, Sawers A. Frequency and Circumstances of Falls Reported by Ambulatory Unilateral Lower Limb Prosthesis Users: A Secondary Analysis. PM&R 2018; in press.
5. Kulkarni J, Wright S, Toole C, Morris J, Hirons R. Falls in patients with lower limb amputations: prevalence and contributing factors. Physiotherapy 1996;82(2):130–136.
6. Gauthier-Gagnon C, Grise MC, Potvin D. Enabling factors related to prosthetic use by people with transtibial and transfemoral amputation. Arch Phys Med Rehabil 1999;80(6):706–713. [PubMed: 10378500]
7. Miller WC, Speechley M, Deathe B. The prevalence and risk factors of falling and fear of falling among lower extremity amputees. Arch Phys Med Rehabil 2001;82(8):1031–1037. [PubMed: 11494181]

8. Wong CK, Chen CC, Blackwell WM, Rahal RT, Benoy SA. Balance ability measured with the Berg balance scale: a determinant of fall history in community-dwelling adults with leg amputation. J Rehabil Med 2015;47(1):80–86. [PubMed: 25223891]
9. Pauley T, Devlin M, Heslin K. Falls sustained during inpatient rehabilitation after lower limb amputation: prevalence and predictors. Am J Phys Med Rehabil 2006;85(6):521–532. [PubMed: 16715022]
10. Chihuri S, Wong CK. Factors associated with the likelihood of fall-related injury among people with lower limb loss. Inj Epidemiol 2018;5(1):42–50. [PubMed: 30417269]
11. Mundell B, Maradit Kremers H, Visscher S, Hoppe K, Kaufman K. Direct medical costs of accidental falls for adults with transfemoral amputations. Prosthet Orthot Int 2017;41(6):564–570. [PubMed: 28641476]
12. van Velzen JM, van Bennekom CA, Polomski W, Slootman JR, van der Woude LH, Houdijk H. Physical capacity and walking ability after lower limb amputation: a systematic review. Clin Rehabil 2006;20(11):999–1016. [PubMed: 17065543]
13. Miller WC, Deathe AB, Speechley M, Koval J. The influence of falling, fear of falling, and balance confidence on prosthetic mobility and social activity among individuals with a lower extremity amputation. Arch Phys Med Rehabil 2001;82(9):1238–1244. [PubMed: 11552197]
14. Pezzin LE, Dillingham TR, MacKenzie EJ. Rehabilitation and the long-term outcomes of persons with trauma-related amputations. Arch Phys Med Rehabil 2000;81(3):292–300. [PubMed: 10724073]
15. Major MJ, Fatone S, Roth EJ. Validity and reliability of the Berg Balance Scale for community-dwelling persons with lower-limb amputation. Arch Phys Med Rehabil 2013;94(11):2194–2202. [PubMed: 23856150]
16. Sawers A, Hafner B. Validation of the Narrowing Beam Walking Test in lower limb prosthesis users. Arch Phys Med Rehabil 2018;99(8):1491–1498. [PubMed: 29653108]
17. Berg K, Wooddauphinee S, Williams JI, Maki B, Gayton D. Measuring balance in the elderly: Validation of an instrument. J Am Geriatr Soc 1991;39(8):A65–A65.
18. Podsiadlo D, Richardson S. The timed “Up & Go”: a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991;39(2):142–148. [PubMed: 1991946]
19. 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):1566–1571. [PubMed: 12422327]
20. Powell LE, Myers AM. The Activities-specific Balance Confidence (ABC) Scale. J Gerontol A Biol Sci Med Sci 1995;50A(1):M28–M34. [PubMed: 7814786]
21. Dite W, Connor HJ, Curtis HC. Clinical identification of multiple fall risk early after unilateral transtibial amputation. Arch Phys Med Rehabil 2007;88(1):109–114.. [PubMed: 17207685]
22. Schoppen T, Boonstra A, Groothoff JW. The Timed “up and go” test: reliability and validity in persons with unilateral lower limb amputation. Arch Phys Med Rehabil 1999;80(7):825–828. [PubMed: 10414769]
23. Miller WC, Deathe AB, Speechley M. Psychometric properties of the activities-specific balance confidence scale among individuals with a lower-limb amputation. Arch Phys Med Rehabil 2003;84(5):656–661. [PubMed: 12736877]
24. Hafner BJ, Morgan SJ, Askew RL, Salem R. Psychometric evaluation of self-report outcome measures for prosthetic applications. J Rehabil Res Dev 2016;53(6):797–812. [PubMed: 28273329]
25. Cardoso JR, Beisheim EH, Horne JR, Sions JM. Test-retest reliability of dynamic balance performance-based measures among adults with a unilateral lower-limb amputation. PM&R 2018; in press.
26. Resnik L, Borgia M. Reliability of outcome measures for people with lower-limb amputations: Distinguishing true change from statistical error. Phys Ther 2011;91(4):555–565. [PubMed: 21310896]
27. Wong CK. Interrater reliability of the Berg Balance Scale when used by clinicians of various experience levels to assess people with lower limb amputations. Phys Ther 2014;94(3):371–378. [PubMed: 24092903]

28. Berg WP, Alessio HM, Mills EM, Tong C. Circumstances and consequences of falls in independent community-dwelling older adults. Age Ageing 1997;26(4):261–268. [PubMed: 9271288]
29. Topper AK, Maki BE, Holliday PJ. Are activity-based assessments of balance and gait in the elderly predictive of risk of falling and/or type of fall? J Am Geriatr Soc 1993;41(5):479–487. [PubMed: 8486878]
30. Elm von E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLOS Med 2007;4(10):e296. [PubMed: 17941714]
31. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377–381. [PubMed: 18929686]
32. Hanspal RS, Fisher K, Nieveen R. Prosthetic socket fit comfort score. Disabil Rehabil 2003;25(22):1278–1280. [PubMed: 14617445]
33. Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. BMJ 1994;309(6948):188. [PubMed: 8044101]
34. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. BMJ 2004;329(7458):168–169. [PubMed: 15258077]
35. Greenhalgh T How to read a paper. Papers that report diagnostic or screening tests. BMJ 1997;315(7107):540–543. [PubMed: 9329312]
36. Hafner BJ, Gaunaurd IA, Morgan SJ, Amtmann D, Salem R, Gailey RS. Construct validity of the Prosthetic Limb Users Survey of Mobility (PLUS-M) in adults with lower limb amputation. Arch Phys Med Rehabil 2016;98(2):277–285. [PubMed: 27590443]
37. Askham J, Glucksman E, Owens P, Swift C. Home and leisure accident research: a review of research on falls among elderly people London, UK: Age Concern Institute of Gerontology, King’s College London, and Department of Trade and Industry; 1990.
38. Lord SR, Ward JA, Williams P, Anstey KJ. Physiological factors associated with falls in older community-dwelling women. J Am Geriatr Soc 1994;42(10):1110–1117. [PubMed: 7930338]
39. Lamb SE, Jørstad-Stein EC, Hauer K, Becker C. Development of a common outcome data set for fall injury prevention trials: the prevention of falls network Europe consensus. J Am Geriatr Soc 2005;53:1618–1622. [PubMed: 16137297]
40. Hauer K, Lamb SE, Jorstad EC, Todd C, Becker C. Systematic review of definitions and methods of measuring falls in randomized controlled fall prevention trials. Age Ageing 2006;35(1):5–10. [PubMed: 16364930]
41. Clemens SM, Gailey RS, Bennett CL, Pasquina PF, Kirk-Sanchez NJ, Gaunaurd IA. The Component Timed-Up-and-Go test: the utility and psychometric properties of using a mobile application to determine prosthetic mobility in people with lower limb amputations. Clin Rehabil 2018;32(3):388–397. [PubMed: 28862042]
42. Portney LG, Watkins MP. Foundations of Clinical Research Upper Saddle River, NJ: Pearson, 2013.
43. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic curve. Radiology 1982;143(1):29–36. [PubMed: 7063747]
44. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983;148(3):839–843. [PubMed: 6878708]
45. Haynes RB. Clinical Epidemiology: How We Do Clinical Practice Research Philadelphia: Lippincott, Williams, and Wilkins; 2006.
46. Akobeng AK. Understanding diagnostic tests 3: Receiver operating characteristic curves. Acta Paediatr 2007;96(5):644–647. [PubMed: 17376185]
47. Schlenstedt C, Brombacher S, Hartwigsen G, Weisser B, Moeller B, Deuschl G. Comparison of the Fullerton Advanced Balance Scale, Mini-BESTest, and Berg Balance Scale to predict falls in Parkinson’s disease. Phys Ther 2016;96(4):494–501. [PubMed: 26381806]
48. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical Epidemiology: A Basic Science for Clinical Medicine Philadelphia: Lippincott, Williams, and Wilkins, 1991.
49. Altman DG, Bland JM. Statistics Notes: Diagnostic tests 1: sensitivity and specificity. BMJ 1994;308(6943):1552–1552. [PubMed: 8019315]

50. Altman DG, Bland JM. Statistics Notes: Diagnostic tests 2: predictive values. BMJ 1994;309(6947):102–102. [PubMed: 8038641]

51. Crombie DL. Diagnostic process. J Coll Gen Pract 1963;6:579–589. [PubMed: 14073679]

52. Sackett DL. A primer on the precision and accuracy of the clinical examination. JAMA 1992;267(19):2638–2644. [PubMed: 1573753]

53. Paul SS, Canning CG, Sherrington C, Lord SR, Close JCT, Fung VSC. Three simple clinical tests to accurately predict falls in people with Parkinson's disease. Mov Disord 2013;28(5):655–662. [PubMed: 23450694]

54. Ziegler-Graham K, MacKenzie EJ, Ephraim PL, Travison TG, Brookmeyer R. Estimating the prevalence of limb loss in the United States: 2005 to 2050. Arch Phys Med Rehabil 2008;89(3): 422–429. [PubMed: 18295618]

55. Pezzin LE, Dillingham TR, MacKenzie EJ, Ephraim P, Rossbach P. Use and satisfaction with prosthetic limb devices and related services. Arch Phys Med Rehabil 2004;85(5):723–729. [PubMed: 15129395]

56. Ehde DM, Czerniecki JM, Smith DG, et al. Chronic phantom sensations, phantom pain, residual limb pain, and other regional pain after lower limb amputation. Arch Phys Med Rehabil 2000;81(8):1039–1044. [PubMed: 10943752]

57. American Board for Certification in Orthotics, Prosthetics, & Pedorthics (ABC). Practice analysis of certified practitioners in the disciplines of orthotics and prosthetics Alexandria, VA 2015;1–60.

58. Wurdeman SR, Stevens PM, Campbell JH. Mobility Analysis of AmpuTees II: Comorbidities and Mobility in Lower Limb Prosthesis Users. Am J Phys Med Rehabil 2018;97(11):782–788. [PubMed: 29781963]

59. Hunter SW, Batchelor F, Hill KD, Hill A-M, Mackintosh S, Payne M. Risk factors for falls in people with a lower limb amputation: A systematic review. PM&R 2017;9(2):170–180. [PubMed: 27485674]

60. Bowman J Challenges to measuring outcomes in occupational therapy: A qualitative focus group study. Br J Occup Ther 2006;69(10):464–472.

61. Gaunaurd I, Spaulding SE, Amtmann D, et al. Use of and confidence in administering outcome measures among clinical prosthetists: Results from a national survey and mixed-methods training program. Prosthet Orthot Int 2015;39(4):314–321. [PubMed: 24827935]

62. Jette DU, Halbert J, Iverson C, Miceli E. Use of standardized outcome measures in physical therapist practice: perceptions and applications. Phys Ther 2009.

63. Duncan EAS, Murray J. The barriers and facilitators to routine outcome measurement by allied health professionals in practice: a systematic review. BMC Health Serv Res 2012;12(1):96–104. [PubMed: 22506982]

64. Valier A, Jennings AL, Parsons JT. Benefits of and barriers to using patient-rated outcome measures in athletic training. J Athl Train 201449(5):674–683. [PubMed: 25098654]

65. Jorgensen V, Opheim A, Halvarsson A, Franzén E, Roaldsen KS. Comparison of the Berg Balance Scale and the Mini-BESTest for assessing balance in ambulatory people with spinal cord injury: Validation study. Phys Ther 2017;97(6):677–687. [PubMed: 28371940]

66. Reid L, Thomson P, Besemann M, Dudek N. Going places: Does the two-minute walk test predict the six-minute walk test in lower extremity amputees? J Rehabil Med 2015;47(3):256–261. [PubMed: 25588644]

67. Lin S-J, Bose NH. Six-minute walk test in persons with transtibial amputation. Arch Phys Med Rehabil 2008;89(12):2354–2359. [PubMed: 18976979]

68. Sakakibara BM, Miller WC, Backman CL. Rasch analyses of the Activities-specific Balance Confidence Scale with individuals 50 years and older with lower-limb amputations. Arch Phys Med Rehabil 2011;92(8):1257–1263. [PubMed: 21704978]

69. Brooks D, Hunter JP, Parsons J, Livsey E, Quirt J, Devlin M. Reliability of the two-minute walk test in individuals with transtibial amputation. Arch Phys Med Rehabil 2002;83(11):1562–1565. [PubMed: 12422326]

70. Mandel A, Paul K, et al. Balance confidence and activity of community-dwelling patients with transtibial amputation. JRRD 2016;53(5):551–560.

71. Hill K, Schwarz J, Flicker L, Carroll S. Falls among healthy, community-dwelling, older women: a prospective study of frequency, circumstances, consequences and prediction accuracy. Aust N Z J Public Health 1999;23(1):41–48. [PubMed: 10083688]
72. Cummings SR, Nevitt MC, Kidd S. Forgetting falls. The limited accuracy of recall of falls in the elderly. J Am Geriatr Soc 1988;36(7):613–616. [PubMed: 3385114]
73. Feldman F, Robinovitch SN. Recalling the mechanics of falls: Young adults cannot accurately describe the sites of impact immediately after a fall occurs. Osteoporos Int 2006;17(6):955–955.
74. Muir SW, Speechley M. Establishing predictive validity of the Fullerton Advanced Balance scale. Arch Phys Med Rehabil 2010;91(7):1147–8. [PubMed: 20599056]
75. Vanicek N, Strike S, McNaughton L, Polman R. Gait patterns in transtibial amputee fallers vs. non-fallers: Biomechanical differences during level walking. Gait Posture 2009;29(3):415–420. [PubMed: 19071021]
76. Crenshaw JR, Kaufman KR, Grabiner MD. Compensatory-step training of healthy, mobile people with unilateral, transfemoral or knee disarticulation amputations: A potential intervention for trip-related falls. Gait Posture 2013;38(3):500–506. [PubMed: 23433547]
77. Quai TM, Brauer SG, Nitz JC. Somatosensation, circulation and stance balance in elderly dysvascular transtibial amputees. Clin Rehabil 2005;19(6):668–676. [PubMed: 16180604]

## Table 1.

Participant demographics stratified by faller status

| Fall Group | Age (yrs) mean (SD) | Sex | Years since Amputationmean (SD) | Amputation Level | Amputation Etiology | PLUS-M mean (SD) | MFCL |
|---|---|---|---|---|---|---|---|
| No falls | 44.9 (14.4) | M (11)<br>F (5) | 16.4 (12.8) | TT (14)<br>TF (2) | Dysvascular (1)<br>Non-dysvascular (15) | 58.5 (9.05) | K1(0) K2(4)<br>K3(10) K4(2) |
| 1 fall | 41.4 (16.3) | M (2)<br>F (5) | 10.8 (4.8) | TT (3)<br>TF (4) | Dysvascular (0)<br>Non-dysvascular (7) | 56.2 (6.35) | K1(0) K2(1)<br>K3(2) K4(4) |
| ≥ 2 falls | 55.4 (11.7) | M (8)<br>F (9) | 13.8 (14.8) | TT (8)<br>TF (9) | Dysvascular (6)<br>Non-dysvascular (11) | 49.2 (6.71) | K1(2) K2(10)<br>K3(5) K4(0) |
| Total | 48.7 (14.6) | M (21)<br>F (19) | 14.3 (12.6) | TT (25)<br>TF (15) | Dysvascular (7)<br>Non-dysvascular (33) | 54.1 (8.68) | K1(2) K2(15)<br>K3(17) K4(6) |

PLUS-M: Prosthetic Limb User's Survey of Mobility; TF: Transfemoral; TT: Transtibial; M: Male; F: Female; Other: trauma, cancer, infection; MFCL: Medicare Functional Classification Level

## Table 2.

Clinical balance test scores stratified by fall history

| Fall Status | NBWT (/1.0) Mean (SD) *Range* | TUG (sec) Median (IQR) *Range* | FSST (sec) Median (IQR) *Range* | BBS (/56) Median (IQR) *Range* | ABC (/100) Median (IQR) *Range* |
|---|---|---|---|---|---|
| No falls (n=16) | .51 (.13)<br>*.27–.76* | 7.86 (2.83)<br>*5.99–13.0* | 7.56 (4.87)<br>*5.26–12.8* | 52.0 (8.50)<br>*40.0–55.0* | 90.3 (18.9)<br>*63.1–100* |
| 1 fall (n=7) | .53 (.16)<br>*.25–.74* | 7.45 (3.04)<br>*5.06–9.18* | 6.57 (3.08)<br>*4.14–11.8* | 52.0 (6.00)<br>*50.0–56.0* | 90.6 (14.1)<br>*53.1–99.2* |
| ≥ 2 falls (n=17) | .25 (.16)[#]<br>*.04–.63* | 10.0 (5.89)[*]<br>*8.30–19.0* | 13.8 (6.76)[*]<br>*6.27–39.3* | 45.0 (12.5)[*]<br>*16.0–55.0* | 74 (26.9)<br>*40.9–96.0* |
| Total (n=40) | .41 (.19)<br>*.04–.76* | 8.75 (3.97)<br>*5.06–19.0* | 8.83 (5.94)<br>*4.14–39.3* | 50.0 (8.50)<br>*16.0–56.0* | 85.0 (23.7)<br>*40.9–100* |

NBWT: Narrowing Beam Walking Test; TUG: Timed Up and Go; FSST: Four Square Step Test; BBS: Berg Balance Scale;

ABC: Activities-specific Balance Confidence Scale; SD: standard deviation; IQR: interquartile range

[#] Test scores significantly different between ≥ 2 falls and 1 fall, as well as ≥ 2 falls and no falls p < .004 (One-way ANOVA)

[*] Test scores significantly different between ≥ 2 falls and 1 fall, as well as ≥ 2 falls and no falls p < .027 (Kruskal-Wallis Test of medians)

## Table 3.

Clinical balance test scores stratified by two faller classifications

**≥ 1 falls (any falls)**

| **Fall Status** | **NBWT (/1.0)**<br>**Mean (SD)**<br>***Range*** | **TUG (sec)**<br>**Median (IQR)**<br>***Range*** | **FSST (sec)**<br>**Median (IQR)**<br>***Range*** | **BBS (/56)**<br>**Median (IQR)**<br>***Range*** | **ABC (/100)**<br>**Median (IQR)**<br>***Range*** |
|---|---|---|---|---|---|
| No falls (n=16) | .51 (.13)<br>*.27−.76* | 7.86 (2.83)<br>*5.99−13.0* | 7.56 (4.87)<br>*5.26−12.8* | 52.0 (8.50)<br>*40.0−55.0* | 90.3 (18.9)<br>*63.1−100* |
| ≥ 1 falls (n=24) | .33 (.20)$^{\#}$<br>*.04−.76* | 9.39 (5.87)<br>*5.99−13.0* | 11.4 (8.54)<br>*4.14 −39.3* | 48.5 (10.8)<br>*16.0−56.0* | 81.8 (24.9)<br>*40.9 −99.1* |

≥ 2 falls (multiple falls)

| Fall Status | NBWT (/1.0)<br>Mean (SD)<br>*Range* | TUG (sec)<br>Median (IQR)<br>*Range* | FSST (sec)<br>Median (IQR)<br>*Range* | BBS (/56)<br>Median (IQR)<br>*Range* | ABC (/100)<br>Median (IQR)<br>*Range* |
|---|---|---|---|---|---|
| 0–1 falls (n=23) | .51 (.14)<br>*.25−.76* | 7.79 (1.66)<br>*5.06–13.0* | 7.46 (3.75)<br>*4.14–12.8* | 52.0 (7.00)<br>*40.0−56.0* | 90.6 (15.6)<br>*53.1−100* |
| ≥ 2 falls (n=17) | .25 (.16)$^{\#}$<br>*.04–.63* | 10.0 (5.89) $^{*}$<br>*8.30–19.0* | 13.8 (6.76) $^{*}$<br>*6.27–39.3* | 45.0 (12.5) $^{*}$<br>*16.0−56.0* | 74 (26.9)<br>*40.9−96.0* |

NBWT: Narrowing Beam Walking Test; TUG: Timed Up and Go; FSST: Four Square Step Test; BBS: Berg Balance Scale;

ABC: Activities-specific Balance Confidence Scale; SD: standard deviation; IQR: interquartile range

[#] Test scores significantly different between No falls and ≥ 1 falls, as well as ≥ 2 falls and 0–1 falls p ≤ .005 (2-sided Independent T-Test)

[*] Test scores significantly different between ≥ 2 falls and 0–1 fall p ≤ .001(Mann Whitney Test of medians)

## Table 4.

Validity indices for performance-based clinical balance tests among unilateral lower limb prosthesis users for two faller classifications

| | | | **≥ 1 falls (any falls)** | | | |
|---|---|---|---|---|---|---|
| **Test** | **AUC (95% CI)** | **Cutoff Score** | **Sensitivity (95% CI)** | **Specificity (95% CI)** | **LR+ (95% CI)** | **LR− (95% CI)** |
| NBWT (/1.00) | .81 (.62–.91) | ≤ 0.43 | 73% (53%–90%) | 76% (54%–96%) | 3.0 (1.5–6.9) | .36 (.19–.77) |
| TUG (sec) | .71 (.54–.88) | ≥ 8.17 | 83% (68%–98%) | 68% (46%–92%) | 2.6 (1.3–5.6) | .24 (.13–.56) |
| FSST (sec) | .70 (.53–.86) | ≥ 8.49 | 74% (58%–92%) | 68% (46%–92%) | 2.4 (1.1–5.2) | .36 (.17–.78) |
| BBS (/56) | .66 (.47–.83) | ≤ 50.5 | 67% (48%–86%) | 62% (39%–86%) | 1.8 (.89–3.6) | .53 (.27–1.1) |
| ABC (/100) | .65 (.47–.82) | ≤ 80.2 | 50% (30%–70%) | 74% (54%–96%) | 1.9 (.78–5.1) | .67 (.41–1.1) |
| Model 1 | .84 (.71–.96) | N/A | 80% (64%–96%) | 73% (51%–96%) | 3.0 (1.3–7.1) | .27 (.12–.63) |

| | | | ≥ 2 falls (multiple falls) | | | |
|---|---|---|---|---|---|---|
| Test | AUC (95% CI) | Cutoff Score | Sensitivity (95% CI) | Specificity (95% CI) | LR+ (95% CI) | LR− (95% CI) |
| NBWT (/1.00) | .89 (.78–1.0) | ≤ 0.43 | 88% (73%–100%) | 79% (65%–92%) | 4.2 (1.7–6.9) | .16 (.042–.60) |
| TUG (sec) | .88 (.78–.98) | ≥ 9.25 | 82% (64%–100%) | 83% (67%–98%) | 4.7 (1.9–11.9) | .21 (.075–.61) |
| FSST (sec) | .88 (.76–.99) | ≥ 8.71 | 94% (83%–100%) | 74% (60%–92%) | 3.6 (1.8–7.3) | .08 (.012–.54) |
| BBS (/56) | .81 (.66–.92) | ≤ 50.5 | 88% (73%–100%) | 70% (51%–88%) | 2.9 (1.5–5.5) | .17 (.045–.64) |
| ABC (/100) | .71 (.54–.87) | ≤ 80.2 | 65% (42%–87%) | 78% (61%–95%) | 3.0 (1.3–7.0) | .45 (.23–.89) |
| Model 2 | .90 (.79–1.0) | N/A | 76% (56%–97%) | 83% (67%–98%) | 4.4 (1.7–11.3) | .28 (.12–.68) |

ABC: Activities-specific Balance Confidence Scale; AUC: Area Under the Curve; BBS: Berg Balance Scale; CI: Confidence Interval; FSST: Four Square Step test; LR-: negative likelihood ratio; LR+: positive likelihood ratio; NBWT: Narrowing Beam Walking Test; TuG: Timed Up and Go.

Model 1: NBWT; FSST; PLUS-M; amputation level

Model 2: NBWT; TUG; PLUS-M; K-level; amputation etiology