

AWARD NUMBER: W81XWH-17-1-0547

TITLE: Can a Novel Beam-Walking Test Improve Fall Risk Assessment in Service Members, Veterans, and Civilians Who Use Lower-Limb Prostheses?

PRINCIPAL INVESTIGATOR: Andrew Sawers, CPO, PhD

CONTRACTING ORGANIZATION: University of Illinois at Chicago, Chicago, IL

REPORT DATE: January 2021

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command  
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;  
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

| REPORT DOCUMENTATION PAGE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                  |                                   |                            | Form Approved<br>OMB No. 0704-0188        |                                                |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------|-----------------------------------|----------------------------|-------------------------------------------|------------------------------------------------|
| Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. <b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                  |                                   |                            |                                           |                                                |
| 1. REPORT DATE<br>January 2021                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                  | 2. REPORT TYPE<br>Final           |                            | 3. DATES COVERED<br>29Sep2017 – 28Sep2020 |                                                |
| 4. TITLE AND SUBTITLE<br><br>Can a Novel Beam-Walking Test Improve Fall Risk Assessment in Service Members, Veterans, and Civilians Who Use Lower-Limb Prostheses?                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                  |                                   |                            | 5a. CONTRACT NUMBER<br>W81XWH-17-1-0547   |                                                |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                  |                                   |                            | 5b. GRANT NUMBER<br>W81XWH-16-OPORP-PORA  |                                                |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                  |                                   |                            | 5c. PROGRAM ELEMENT NUMBER                |                                                |
| 6. AUTHOR(S)<br>Andrew Sawers CPO, PhD; Brian Hafner PhD<br><br>E-Mail: asawers@uic.edu                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                  |                                   |                            | 5d. PROJECT NUMBER                        |                                                |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                  |                                   |                            | 5e. TASK NUMBER                           |                                                |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                  |                                   |                            | 5f. WORK UNIT NUMBER                      |                                                |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br>University of Illinois<br>809 S MARSHFIELD RM 520<br>CHICAGO IL 60612-4305                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                  |                                   |                            | 8. PERFORMING ORGANIZATION REPORT NUMBER  |                                                |
| 9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)<br><br>U.S. Army Medical Research and Materiel Command<br>Fort Detrick, Maryland 21702-5012                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                  |                                   |                            | 10. SPONSOR/MONITOR'S ACRONYM(S)          |                                                |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                  |                                   |                            | 11. SPONSOR/MONITOR'S REPORT NUMBER(S)    |                                                |
| 12. DISTRIBUTION / AVAILABILITY STATEMENT<br><br>Approved for Public Release; Distribution Unlimited                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                  |                                   |                            |                                           |                                                |
| 13. SUPPLEMENTARY NOTES                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                  |                                   |                            |                                           |                                                |
| 14. ABSTRACT<br>The purpose of the proposed research is to improve the assessment of fall risk among unilateral lower limb prosthesis (LLP) users. The scope of the proposed research includes establishing the validity and reproducibility of contemporary clinical balance test, including the Narrowing-Beam Walking Test (NBWT). This is to be accomplished by administering the NBWT and four contemporary performance-based balance tests to 60 lower limb prosthesis users, who are then followed prospectively for 6 months to identify fall events. We have completed recruitment and enrollment of study participants at both study sites (UIC and UW), all prospective data collection, and all data analyses. Dissemination efforts have resulted in nine manuscripts (5 published, 2 in review, 2 in prep), five invited talks or symposia, and seven posters or podium presentations. Our Narrowing Beam Walking Test has also been adopted and implemented by a prosthetic manufacturer (Ohio Willow Wood), and several academic/clinical sites (e.g., Minneapolis VA, Center for the Intrepid, Seattle VA, Delft University, and Emory University). Our research was also featured in the CDMRP FY19 Annual Report. Our principal finding is that a model combining performance-based balance tests and the number of falls recalled over the past year can predict the number of future falls in established, unilateral LLP users. This is the first study to develop a model that clinicians can use to predict future falls in established unilateral LLP users. |                                  |                                   |                            |                                           |                                                |
| 15. SUBJECT TERMS<br>amputee; balance; falls; rehabilitation                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                  |                                   |                            |                                           |                                                |
| 16. SECURITY CLASSIFICATION OF:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                  |                                   | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES                       | 19a. NAME OF RESPONSIBLE PERSON                |
| a. REPORT<br>U<br>Unclassified                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | b. ABSTRACT<br>U<br>Unclassified | c. THIS PAGE<br>U<br>Unclassified |                            |                                           | USAMRMC                                        |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                  |                                   | UU                         | 150                                       | 19b. TELEPHONE NUMBER (include area code)<br>2 |

## Table of Contents

|                                                                           | <b><u>Page</u></b> |
|---------------------------------------------------------------------------|--------------------|
| <b>1. Introduction.....</b>                                               | <b>4</b>           |
| <b>2. Keywords.....</b>                                                   | <b>4</b>           |
| <b>3. Accomplishments.....</b>                                            | <b>4-14</b>        |
| <b>4. Impact.....</b>                                                     | <b>15</b>          |
| <b>5. Changes/Problems.....</b>                                           | <b>16</b>          |
| <b>6. Products, Inventions, Patent Applications, and/or Licenses.....</b> | <b>16-18</b>       |
| <b>7. Participants &amp; Other Collaborating Organizations.....</b>       | <b>18-19</b>       |
| <b>8. Special Reporting Requirements.....</b>                             | <b>19</b>          |
| <b>9. Appendices.....</b>                                                 | <b>19</b>          |

## INTRODUCTION

Existing clinical balance tests are too easy. They do not pose sufficient challenge to expose subtle, but critically important differences in balance that underlie fall risk in Service members, Veterans, and civilians lower limb prosthesis users. Consequently they are unable to discriminate between lower limb prosthesis users with and without a history of falls, or predict who is at risk for a fall. As a result there are no effective or accepted screening tests for diagnosing fall risk in lower limb prosthesis users. To address this gap and improve the assessment of fall risk in lower limb prosthesis users, the purpose of this project is to assess the validity and reliability of a novel Narrowing Beam-Walking Test (NBWT) that provides progressively increasing challenge to balance control. The scope of the proposed research includes establishing the criterion validity (Aim 1), reproducibility (test-retest and inter-rater reliability) (Aim 2), and predictive validity (Aim 3) of a new clinical balance test, the Narrowing-Beam Walking Test (NBWT). This is to be accomplished by administering the NBWT and four contemporary performance-based balance tests to 60 lower limb prosthesis users. These individuals will then be followed prospectively for 6 months to report fall events. This project addresses a notable gap in the DoD/VA lower limb amputation rehabilitation clinical practice guidelines, a valid and reliable test to predict fall risk in lower limb prosthesis users.

## KEYWORDS

Amputee; falls; fall risk; balance; prosthesis; prospective; validity; reliability; rehabilitation

## ACCOMPLISHMENTS

Major goals of the project: The major goals (milestones) of the project as outlined in the SOW were:

Goal 1: Obtain and maintain IRB approvals from UIC, UW, and ORP/HRPO  
(*Target date: Y1Q1, Y1Q4, Y2Q4 – 100% Completed*).

Goal 2: Recruitment, consent, and data collection materials/databases/equipment/staff ready for data collection (*Y1Q1- 100% Completed*).

Goal 3: 60 participants enrolled (*Y2Q2 – 100% completed*).

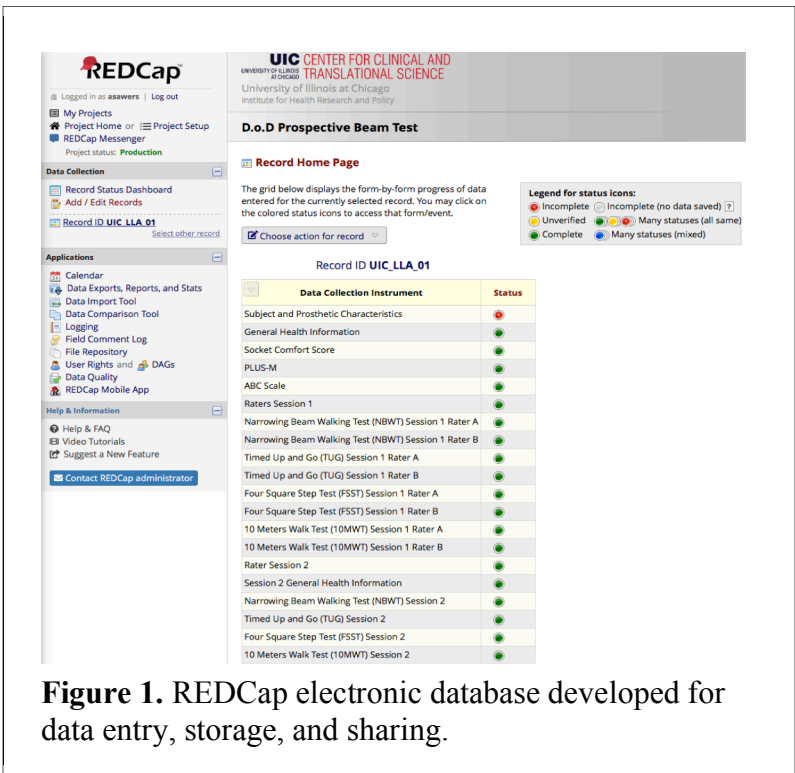
Goal 4: Data collection completed (*Y2Q3 – 100% completed*).

Goal 5: Data entered, processed, and analyzed to address study hypotheses (*Y2Q4 – 100% completed*).

Goal 6: Abstracts presented at a minimum of 2 scientific conferences, manuscripts prepared and submitted for publication, delivery of training material for clinical implementation of beam-walking test, presentation of executive summary to senior VA/DoD clinical personnel.  
(Y3Q4 –100% completed).

Accomplishments under these goals: For each of the goals/milestones outlined in the SOW, we have made significant and timely progress beyond our Y1 Annual Report.

Goal 1: Both study sites (UIC and UW) completed and maintained their annual IRB continuing reviews and their ORP/HRPO annual continuing reviews throughout the study period.



Goal 2: All recruitment, consent, and data collection materials, databases, equipment and staff were prepared by Y1Q1, and maintained or updated throughout the study period. The PI and Co-I of the project prepared, revised, and finalized detailed data collection forms to standardize the data collection procedures across study sites. The PI and other members of the UIC study team members developed a project- specific REDCap database for electronic data entry, storage, and sharing. Co-I and other UW study team members tested the REDCap database for bugs. Feedback provided to the PI and UIC study team members was used to revise the database. The REDCap database served to enter, store, organize, and successfully

share de-identified data between study sites (Figure 1).

The PI and Co-I revised the design of the Narrowing Beam-Walking Test equipment to increase its portability and durability. Specifically this involved changing the material the Narrowing Beam-Walking Test was constructed from, and the manner in which each segment of the Narrowing Beam-Walking Test is connected. The material was changed from a hard pine, which warps over time, to a compressed plastic lumbar. The connections between adjoining segments of the Narrowing Beam-Walking Test segments were changed from an L- bracket to a set of bolt and eyehook fixtures (Figure 2). Representative recruitment materials were distributed to our clinical collaborators at the Chicago (Scheck and Siress Inc.,) and Seattle

(Cornerstone Prosthetics and Orthotics) study sites. Feedback on recruitment materials from local clinicians and ways to maximize recruitment were incorporated into final versions that were re-distributed to our clinical partners.



**Figure 2.** Design and construction of our Narrowing Beam Walking Test (NBWT). (A) View of the entire length of each segment with corresponding distance marks etched in. (B and C) Zoomed in views of each segment, highlighting the bolt & eyehook distal fixtures used to connect adjoining beam segments. (D) View of the NBWT with each segment connected and ready for testing.

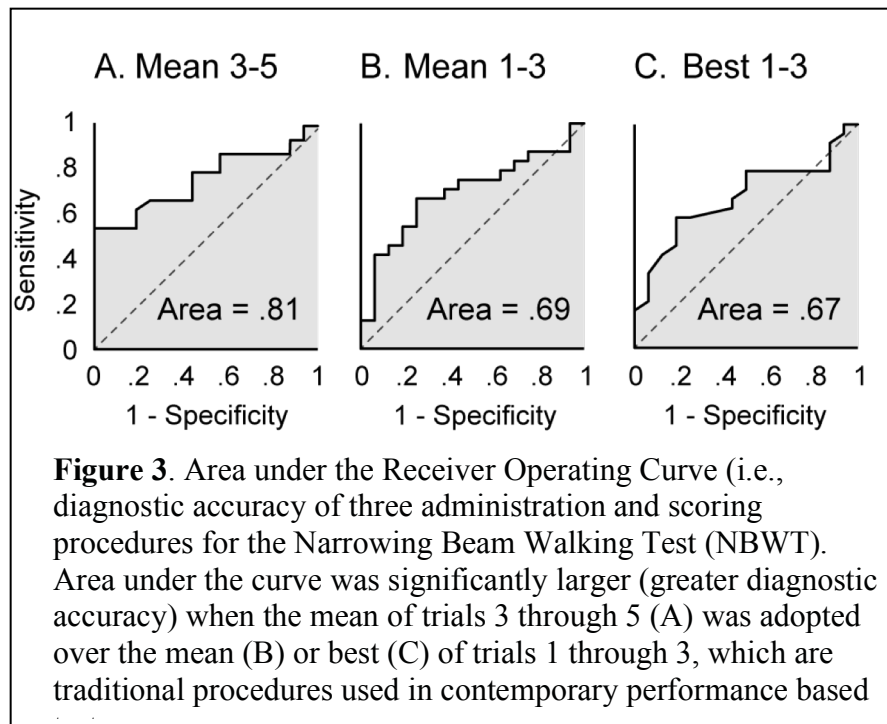
We also created and distributed a detailed fabrication guide for the Narrowing Beam-Walking Test that included a parts list and step-by-step building instructions (Appendix).

Goal 3: We met all quarterly enrollment targets at both study sites. We recruited n=60 participants between the UIC site and the UW site. Because of our success recruiting lower limb amputees for the project we did not have to enhance or expand our recruitment efforts at either study site.

Goal 4: Planned data collection has been completed at both study sites. Namely, clinical balance test data and prospective 6-month fall data have been collected on all 60 participants. We attribute the success of our data collection at both sites to the implementation of a new “mass data collection” procedure. Briefly, this involved recruiting and scheduling multiple participants on a single day, between 6 and 9, and running them through multiple stations each manned by different study personnel. This was done using space in the PI and Co-PI’s labs, local community centers, as well as with our clinical partners. Our success in recruitment and data collection allowed us to extend the prospective follow up period from 6 to 12 months, and ask about fall recall. This will allow us to evaluate fall recall memory decay in people with lower limb amputation, a topic of concern given the extensive use of retrospective fall data in the literature and cognitive deficits in LLP users. To date this topic has been poorly addressed.

Goal 5: All data has been entered, via double entry to minimize data entry errors, processed, and analyzed to address study hypotheses. Each study site (UIC and UW) digitized their data collection sheets and uploaded them to the REDCap database under de-identified, participant-specific numbers. To minimize data entry error, the raw data from the data collection sheets was entered into the REDCap database via double entry. The PI resolved all discrepancies. All demographic, health, amputation, prosthetic, mobility, balance, and falls data have been processed and compiled into a master database, which is accompanied by a data dictionary.

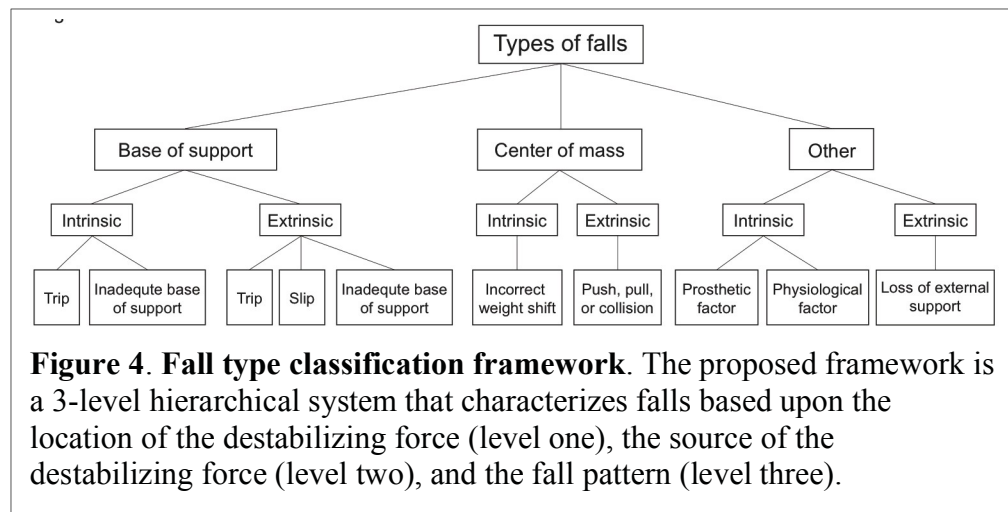
Goal 6: We maintained a productive publication record under this award. To date we have published 5 peer-reviewed articles, with 2 others under review, and 2 others in preparation. All of these manuscripts present results that are relevant to the current project. The first article published under this award, *“Conventional administration and scoring procedures suppress the diagnostic accuracy of a performance-based test designed to assess balance ability in lower limb prosthesis users”*, was published in *Prosthetics and Orthotics International*. The main result from this paper was that if practice



effects on the Narrowing Beam Walking Test were not accommodated through administration and scoring procedures (i.e., taking the mean of trials 3 to 5 rather than the best of the first two trials), the diagnostic accuracy of the Narrowing Beam Walking Test for fall risk was significantly lower (i.e., the area under the Receiver Operating Characteristic curve dropped from .81 to .67) (Figure 3). The take home message from this paper is therefore that the effectiveness of the Narrowing Beam Walking Test at

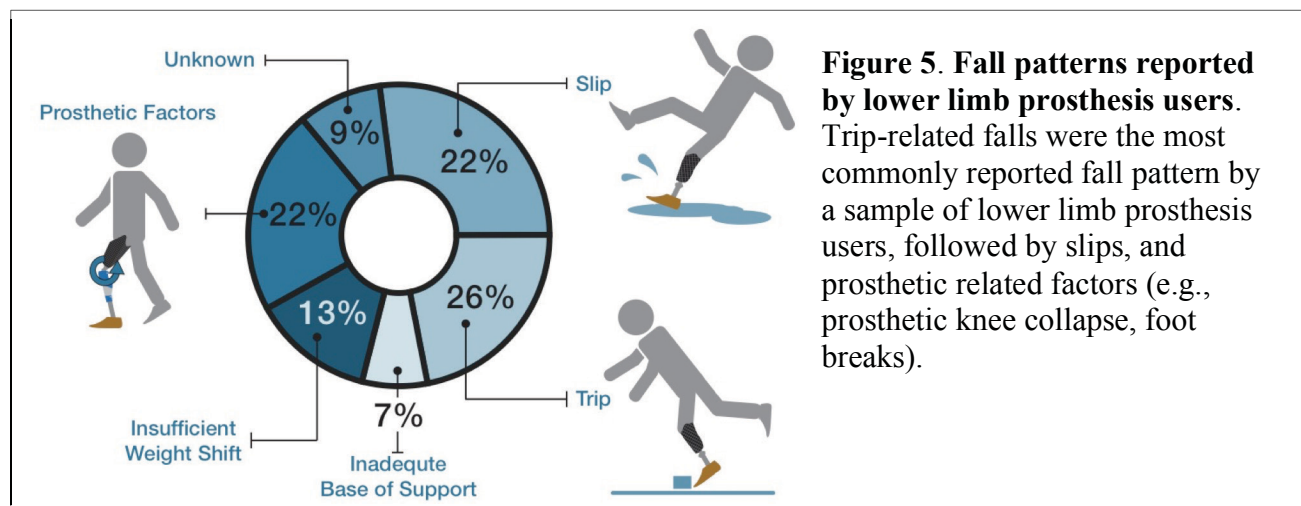
discriminating between fallers and non-fallers was not due to the challenge it presents to balance control, as was originally hypothesized, but rather having accommodated practice effects via administration and scoring procedures. These results suggest that practice effects in other performance-based clinical balance tests may similarly limit their diagnostic accuracy. Our fifth publication under this award provides a preliminary analysis of this problem in other contemporary performance-based tests.

The second manuscript, “*Frequency and circumstances of falls reported by ambulatory unilateral lower limb prosthesis users*”, was published in PM & R. In this paper we developed and tested a fall type classification framework (Figure 4). The framework was then used to classify fall



patterns reported by 66 lower limb prosthesis users. Among the results presented in this manuscript, trip-related falls were found to be the most common fall pattern reported by lower limb

prosthesis users (Figure 5), while walking was the most common activity being performed at the time of a fall. The results in this manuscript informed the questions we asked lower limb prosthesis about their falls during our monthly prospective follow-up phone calls in the current study. The results of this manuscript have also motivated other funded projects in our lab. This includes the development of a structured and clinically meaningful LLP user-specific fall circumstance and consequence survey. This instrument will be administered to a large national sample of lower limb prosthesis users to collect detailed prospective fall circumstance and consequence data, which will guide future research efforts. Other research labs have also implemented the fall type classification framework proposed and used in this manuscript (Anderson et al., 2020).



Anderson C, Miller MJ, Murray AM, Fields TT, So NF, Christiansen CL. (2020) Falls after dysvascular transtibial amputation: A secondary analysis of falling characteristics and reduced physical performance. *PM & R*. XX:1-10.

The third manuscript, “*Using clinical balance tests to assess fall risk in established lower limb prosthesis users: cutoff scores and related validity indices*”, was published in *PM & R*. In this paper we introduced cut-off scores for a suite of performance-based clinical balance tests that facilitate the discrimination between fallers and non-fallers. These cut-off scores were accompanied by a number of validity indices, including likelihood ratios, which clinicians can use to assess the probability that an individual patient will fall in the next 12 months (Table 1). Such validity indices were previously absent from the literature, limiting clinicians’ ability to assign an individual patient a probability of falling. The assignment of a probability is an improvement over previously available metrics that provide a binary outcome, faller or non-faller.

**Table 1.** Validity indices for performance-based clinical balance tests among unilateral lower limb prosthesis users for two faller classifications

| Test         | AUC (95% CI)  | Cutoff Score | <b>≥ 1 falls (any falls)</b>      |                      | LR+ (95% CI)   | LR- (95% CI)   |
|--------------|---------------|--------------|-----------------------------------|----------------------|----------------|----------------|
|              |               |              | Sensitivity (95% CI)              | Specificity (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)  |
| Test         | AUC (95% CI)  | Cutoff Score | <b>≥ 2 falls (multiple falls)</b> |                      | LR+ (95% CI)   | LR- (95% CI)   |
|              |               |              | Sensitivity (95% CI)              | Specificity (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

The fourth manuscript, “*Interrater and test-retest reliability of performance-based clinical tests administered to established users of lower limb prostheses*”, was published in *Physical Therapy*. In this paper we derived reliability indices that clinicians can use to: i) select balance tests that are suitable for making group-level comparisons or individual-level clinical decisions (i.e., test-retest intra-class

correlation coefficients), ii) interpret the precision of individual test scores on performance-based clinical balance tests (i.e., standard error of measurement), and iii) evaluate change over time in balance performance among individual lower limb prosthesis users (i.e., minimal detectable change) Results from this study provide valuable evidence of interrater and test-retest reliability, measurement error, and MDC for performance-based balance tests among established unilateral LLP users. All studied tests demonstrated sufficient reliability to be considered appropriate for group-level comparisons (i.e.,  $ICC \geq 0.70$ ) (Table 2), while the Four Square Step Test, 10 Meter Walk Test, and Narrowing Beam Walking Test were also found to be suitable for individual-level applications among established unilateral LLP users (i.e.,  $ICC \geq 0.90$ ) (Table 2).

**Table 2.** Test-retest reliability of performance-based clinical balance tests in people with lower-limb loss

| Test                | Session 1<br>Score<br>Mean (SD) | Session 2<br>Score<br>Mean (SD) | Score Change<br>Mean (SD) | ICC (2,1)<br>(95% CI) | SEM | MDC <sub>90</sub><br>(95% CI) |
|---------------------|---------------------------------|---------------------------------|---------------------------|-----------------------|-----|-------------------------------|
| NBWT (score, /1.00) | .42 (.21)                       | .44 (.24)                       | .02 (.10)                 | .90 (.83-.94)         | .07 | .16 (.10-.20)                 |
| TUG (time, s)       | 10.4 (2.76)                     | 9.91 (2.75)                     | .50 (1.30)*               | .88 (.78-.93)         | .96 | 2.2 (1.5-2.9)                 |
| FSST (time, s)      | 10.4 (5.27)                     | 10.3 (4.83)                     | .14 (1.32)                | .97 (.94-.98)         | .87 | 2.0 (.79-3.5)                 |
| 10MWT (time, s)     | 8.74 (1.96)                     | 8.58 (2.00)                     | .16 (.54)                 | .96 (.93-.98)         | .39 | .91 (.40-1.4)                 |

\* Test score significantly different between test sessions (2-sided paired t-test)  $p=.004$ . NBWT: Narrowing Beam Walking Test, TUG: Timed Up and Go, FSST: Four Square Step Test, 10MWT: 10 Meter Walk Test, s: seconds, SD: Standard Deviation, CI: Confidence Interval, SEM = Standard Error of Measurement, MDC<sub>90</sub>: Minimal Detectable Change

Standard error of measurement and minimal detectable change estimates of the studied balance tests provide clinicians and researchers with the information necessary to assess differences or changes in balance ability and/or fall risk among individual LLP users. In conjunction with our published validity indices (i.e., manuscripts 3 and 7 under the current award) clinicians and researchers can use the calculated reliability metrics to better select clinical tests for individual LLP users and more effectively interpret the resulting scores to make informed, evidence-based clinical decisions. Importantly, these reliability indices were previously absent from the literature, limiting the ability of clinicians to select and interpret performance-based clinical balance tests among lower limb prosthesis users.

Our fifth manuscript, “*Characterizing practice effects in performance-based tests administered to users of unilateral lower limb prosthesis: A preliminary study*”, was published in PM & R. Across the Four Square Step Test, Timed Up and Go, and 10m Walk Test, we observed practice effects in 45% to 76% of participants. The proportion of participants with practice effects (i.e., occurrence) was significantly greater for the Four Square Step Test than the 10 meter Walk Test ( $p = .008$ ). The median number of trials (i.e., time-course) required for participants to reach a consistent level of performance was not significantly different between tests (Four Square Step Test: 4 trials, Timed Up and Go: 4 trials; 10 meter Walk Test: 3.5 trials;  $p = .481$ ) (Figure 6). Practice effect magnitude (i.e., difference between the mean of trials during the plateau and best performance over the first two trials) was significantly greater than zero for the Four Square Step Test (1.6 s; 16%) and Timed Up and Go (1.4 s; 13%) ( $p<.05$ ) (Table

3). These results suggest that the Four Square Step Test, Timed Up and Go, and 10 meter Walk Test are susceptible to practice effects in LLP users. Practice effects may obscure significant differences in walking and balance ability, and thereby, in the absence of modifications, limit the use of these tests for making individual patient clinical decisions and analyzing group-level data. We are currently pursuing funding to address this concern.

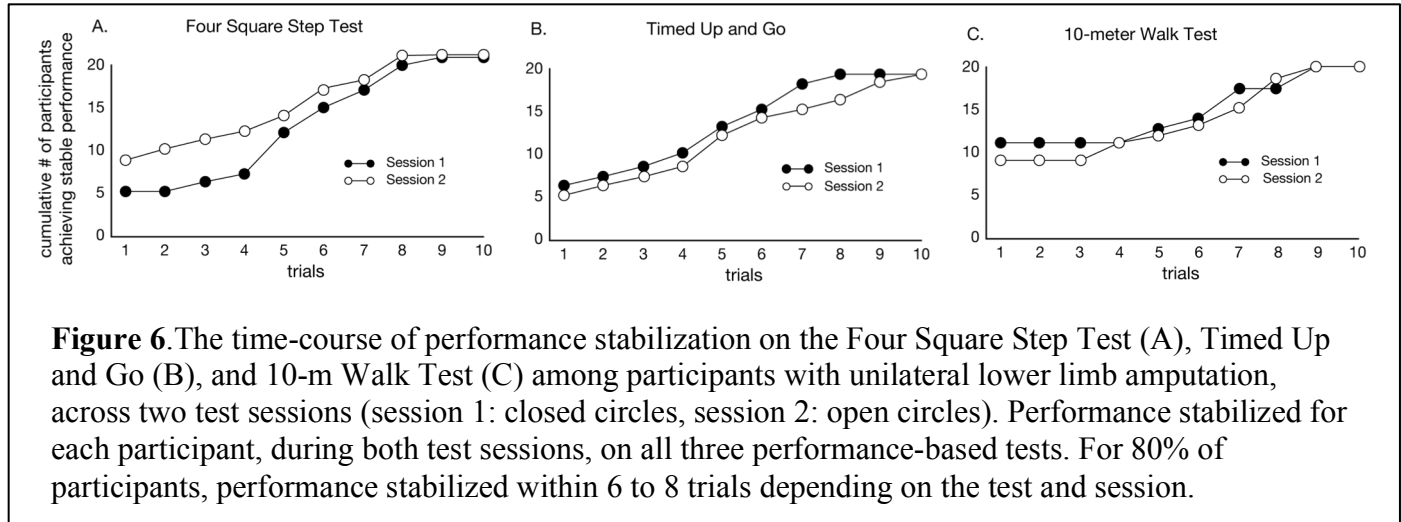


**Table 3.** Conventional and true performance test scores

|                                    | Session 1                |                    |                           | Session 2                |                    |                           |
|------------------------------------|--------------------------|--------------------|---------------------------|--------------------------|--------------------|---------------------------|
|                                    | Best of First Two Trials | Stable Performance | Practice Effect Magnitude | Best of First Two Trials | Stable Performance | Practice Effect Magnitude |
|                                    | Median (Q1, Q3)          |                    | Median (%)                | Median (Q1, Q3)          |                    | Median (%)                |
| Participants with practice effects |                          |                    |                           |                          |                    |                           |
| FSST (time, s)                     | 9.7 (7.1,12)             | 8.1 (6.0,11)       | 1.6 (16%)**               | 9.1 (6,12)               | 8.1 (5.5,10)       | 1.0 (11%)*                |
| TUG (time, s)                      | 11 (8.9,12)              | 9.6 (8.6,12)       | 1.4 (13%)*                | 9.8 (8.7,12)             | 9.6 (8.3,12)       | 0.20 (2%)                 |
| 10mWT (speed, m/s)                 | 1.2 (.98,1.5)            | 1.1 (1.0,1.5)      | 0.10 (8%)                 | 1.1 (1.0,1.3)            | 1.1 (1.0,1.3)      | 0.00 (0%)                 |
| All participants                   |                          |                    |                           |                          |                    |                           |
| FSST (time, s)                     | 9.9 (8.4,12)             | 8.9 (7.5,11)       | 1.01 (10%)**              | 9.3 (7.3,12)             | 8.7 (7.4,11.2)     | 0.55 (6%)                 |
| TUG (time, s)                      | 10.4 (8.6,12)            | 9.6 (8.5,11)       | 0.80 (8%)**               | 9.6 (7.9,12)             | 9.4 (8.1,11)       | 0.20 (2%)                 |
| 10mWT (speed, m/s)                 | 1.3 (1.0,1.5)            | 1.3 (1.1,1.5)      | 0.01 (1%)                 | 1.2 (1.1,1.4)            | 1.3 (1.1,1.5)      | 0.01 (1%)                 |

FSST: Four Square Step Test; m/s: meters per second; Q1: first quartile; Q3: third quartile; s: seconds; TUG: Timed Up and Go; 10mWT: 10 meter Walk Test

\*Practice effect magnitude is significantly greater than zero ( $p \leq .05$ )

\*\*Practice effect magnitude is significantly greater than zero ( $p \leq .002$ )

The sixth publication under this award “*Recalled number of falls in the prior year, combined with perceived mobility, predicts the incidence of future falls in established unilateral lower limb prosthesis users*”, is currently under review in Prosthetics and Orthotics International. To screen for fall risk, clinicians often turn to commonly held beliefs regarding easily accessible attributes thought to imply fall risk in LLP users (e.g., age, amputation level or etiology). While appealing because of the ease with which they can be assessed, evidence to support the association between these assumed risk factors and falls remains limited in established lower limb prosthesis users. The objective of this study was therefore to test whether easily-accessible attributes and/or measurements clinicians might use to assess fall risk (e.g., level of amputation, recalled fall incidence), could predict the future incidence of falls in established unilateral LLP users. An initial model that included six factors (i.e., i) number of falls recalled over the past 12 months, ii) level of amputation, iii) cause of amputation, iv) time since amputation, v) Socket Comfort Score, and vi) PLUS-M T-scores) was reduced using backward stepwise regression. The final model retained two of its original six factors, number of falls recalled, and PLUS-M T-score (i.e., perceived mobility). The combined effect of these two predictors on the number of future falls in established unilateral LLP users can viewed jointly in the final model (EQ.1).

$$\#future\ falls = \exp ((0.126)(\#falls\ recalled) + (-0.052)(PLUS-M\ T-score) + 2.452) \quad [EQ.1]$$

Based on these results, we suggest that the number of recalled falls over the past year, combined with PLUS-M T-scores, can be used to predict the number of falls over the next 6 months in established, unilateral LLP users. The success and simplicity of the final model suggests it may serve as a screening tool for clinicians to assess fall risk. Equally important, we found that many of the risk factors clinicians assume to be associated with falls (e.g., level and cause of amputation, age, time since amputation) were not predictive of future falls in unilateral lower limb prosthesis users. To the best of our knowledge, this is the first study to have examined whether recalled falls can serve as a predictor of future falls in LLP users.

The seventh publication under this award “*Performance-based balance test, combined with the number of falls recalled in the past year, predicts the incidence of future falls in established unilateral transtibial prosthesis users*”, is under review in Archives of Physical Medicine and Rehabilitation. The primary objective of this study was to assess whether common performance-based balance tests could predict the number of future falls in established unilateral LLP users. The secondary objective was to test whether the addition of performance-based test scores to the number of falls recalled over the past year would improve the prediction of future falls in LLP users compared to fall recall alone, which we previously found to predict future falls in established, unilateral LLP users (paper 7 under current award). Based on regression analyses, the key results from this study were: i) no combination of performance-based balance tests included in this study could alone predict the number of future falls in established unilateral lower limb prosthesis users; rather ii) a combination of two balance tests, the Timed Up and Go and Four Square Step Test, plus the number of falls recalled from the past 12 months was required to develop a model that could predict the number of future falls in established unilateral

transtibial prosthesis users (EQ. 2), iii) the addition of performance-based balance test scores to fall recall improved the prediction of future falls in established, unilateral transtibial prosthesis users over fall recall alone, and iv) no combination of factors included in the current study significantly predicted the number of future falls in participants with transfemoral amputation.

$$\# \text{ Future falls} = \exp [(0.202) * (\# \text{ falls recalled past year}) + (0.370) * (TUG) + (-0.154) * (FSST) - 2.90] \text{ [EQ.2]}$$

By administering the Timed Up and Go and Four Square Step Test, and asking about the number of falls recalled over the past year, clinicians can use the predictive model (i.e., EQ. 2) to make predictions about the number of fall(s) that might be expected in the next 6 months for individual unilateral transtibial prosthesis users. For example, an established, unilateral transtibial prosthesis user with a 11.8 s Timed Up and Go time, a 9.7 s Four Square Step Test time, and who recalls having fallen twice in the past year, would be predicted to fall 1.46 times over the next 6 months.

$$\# \text{ Future falls} = \exp [(0.202) * (\# \text{ falls recalled past year}) + (0.370) * (TUG) + (-0.154) * (FSST) - 2.90]$$

$$\# \text{ Future falls} = \exp [(0.202) * (2) + (0.370) * (11.8) + (-0.154) * (9.7) - 2.90]$$

$$\# \text{ Future falls} = \exp [(0.202) * (2) + (0.370) * (11.8) + (-0.154) * (9.7) - 2.90]$$

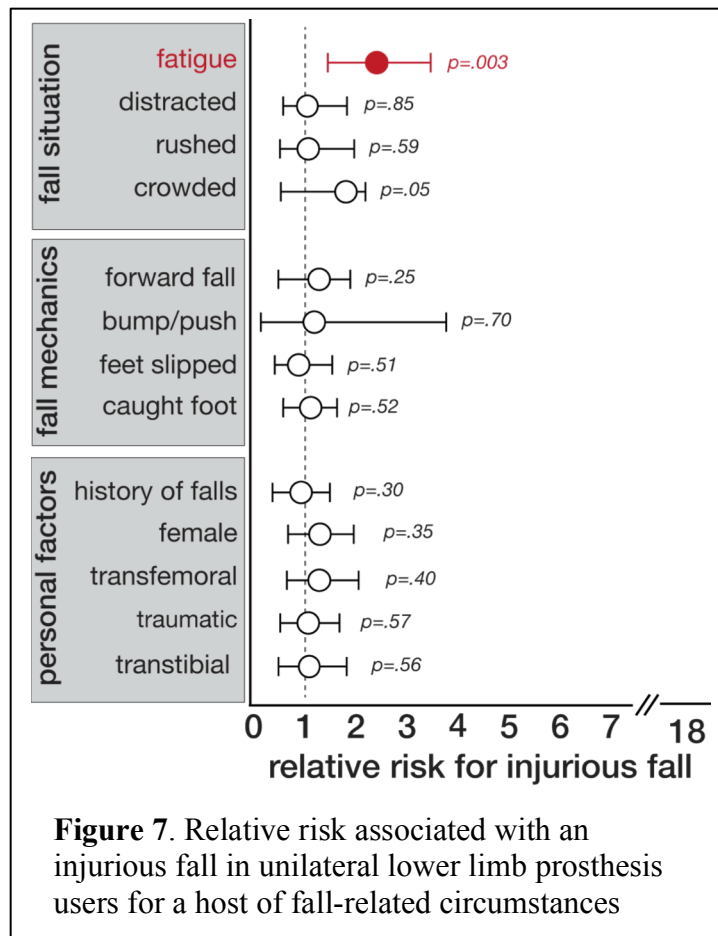
$$\# \text{ Future falls} = \exp [.377]$$

$$\# \text{ Future falls} = 1.46$$

An important result from this study that should not get lost in the success of predicting falls in transtibial prosthesis users, is the inability of any combination of independent variable(s) in the current study to predict the number of future falls in participants with transfemoral amputation. Transfemoral prosthesis users may have different balance deficits than transtibial prosthesis users, and that these deficits may not be assessed by the tests administered in the current study. The activities, causes, or reasons transfemoral prosthesis users fall may differ from those in transtibial prosthesis users. As a result, performance-based tests that are suitable for inferring fall risk in transtibial prosthesis users may not be equally suitable for transfemoral prosthesis users. The assessment of fall risk may therefore need to be stratified by level of amputation. We will be pursuing future funding from the DoD to address this important issue.

We currently have two additional planned manuscripts in preparation. These include: i) “*fall recall accuracy in people with lower limb amputation*”, and ii) “*fall related circumstances associated with consequential falls in lower limb prosthesis users*”. Briefly, in a sample of 48 established, unilateral LLP users, we found that only 79% accurately recalled whether they had fallen or not in the past 6 months, and that only 70% accurately recalled the specific number of falls they had prospectively reported over the past 6 months. Error in fall recall frequency was evenly distributed between over- and under-reporting (i.e., 15% each). Equally troubling, less than 50% of those who fell accurately recalled the activity, environment, situation, or consequence(s) associated with fall(s) they had prospectively reported over the past 6 months. These results suggest that all future research regarding falls in lower limb prosthesis users should make every effort to use prospective rather than retrospectively collected

falls data. Regarding consequential falls, within our prospectively collected falls data from 60 lower limb prosthesis users, we found physical injury to be the most common, costly, and activity-limiting consequence of falling. Remarkably, fatigue due to activity at the time of a fall (i.e., performance fatigability) was the only circumstance significantly associated with injurious falls, increasing the risk of injury by over 2.5 times (Figure 7). Based on these results and other pilot data demonstrating that performance fatigability is greater in LLP users than controls, we will be submitting a proposal to the DoD to determine the extent to and mechanisms by which performance fatigability affects lower limb prosthesis users, and whether a powered ankle-foot prosthesis can reduce these observed deficits in fatigability as well or better than a standardized walking program. We anticipate submitting these two manuscripts for peer-review in the next 6 months. We will forward copies of accepted papers to our assigned Scientific Officer.



Opportunities for training and professional development: Nothing to report.

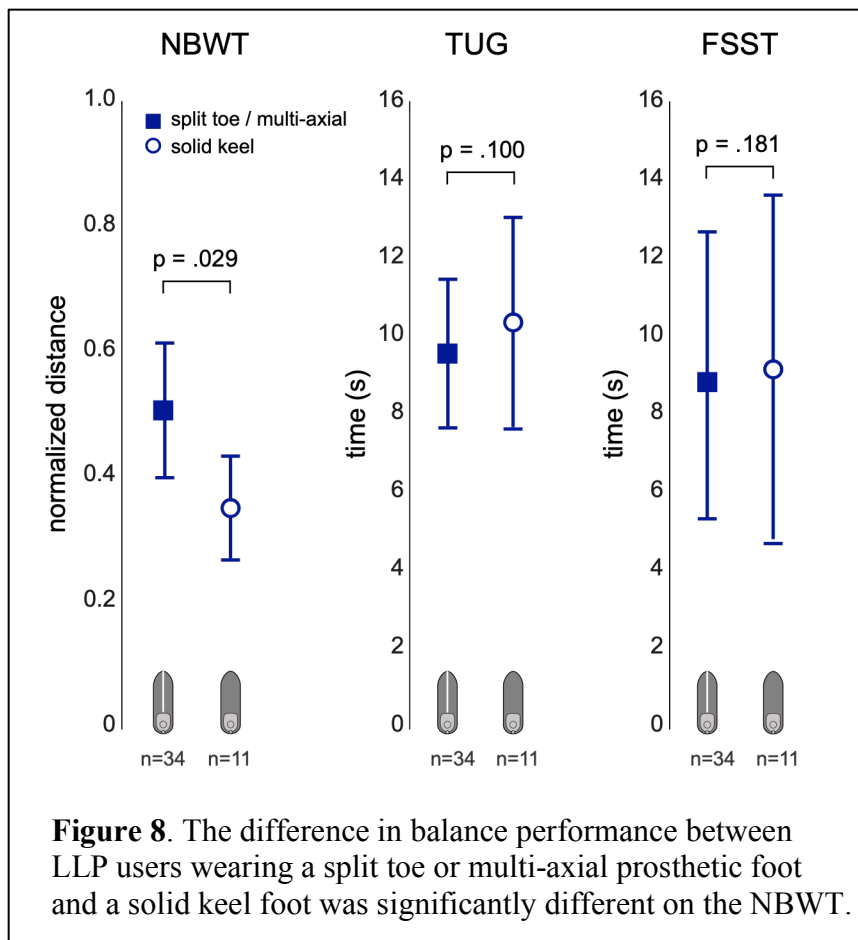
Dissemination of results to communities of interest: Results were disseminated to a national audience of prosthetists at the 2019 American Academy of Prosthetists and Orthotists Annual Meeting and Scientific Conference, and the 2019 Orthotics and Prosthetics Innovative Technologies Conference, to local prosthetists and lower limb prosthesis users at the 2019 Scheck and Siress Education Fair, national researchers at the 2019 International Society of Posture and Gait Research (ISPGR), and the 2019 American Congress of Rehabilitation Medicine (ACRM), as well as military and civilian rehabilitation experts at the 2019 MHSRS conference. An abstract focusing on our fall risk prediction results (i.e., papers 6 and 7 under this award) is in preparation and will be submitted to the 18<sup>th</sup> World Congress

of the International Society for Prosthetics and Orthotics (ISPO) in 2021. We are also working to complete an executive summary of the research conducted under this award on fall risk prediction in lower limb prosthesis users to be presented to senior VA/DoD clinical personnel including the VA Amputation System of Care, and VA/DOD Extremity Trauma and Amputation Center of Excellence. We hope that this executive summary can contribute to filling the gap regarding fall risk assessment in the DoD/VA lower limb amputation rehabilitation clinical practice guidelines.

Plan to accomplish goals during over next reporting period: Nothing to report.

## IMPACT

The performance-based balance test that we have designed, developed, and tested, the Narrowing Beam Walking Test, is now being used in a number of clinical, industry, and academic centers. This includes but is not limited to: Ohio Willow Wood, the Minneapolis VA Health Care System (Adaptive Design and Engineering Program), Center for the Intrepid, VA Puget Sound Health Care System (Center for Limb Loss and Mobility), Delft University, and Emory University for a variety of purposes. The regression equation model in paper 7 under the current award represent the first, and only, published model that clinicians can use to predict the number of future falls in established unilateral transtibial prosthesis users. The fall type classification framework published in the first manuscript under this award has been implemented by other research labs in their falls data analyses, and was the motivation



for additional funding our lab has received to develop a structured and clinically meaningful LLL users-specific fall circumstance and consequence survey. Data from the current study served as pilot data for two additional funded OPORP grants. The first was awarded to Dr. David Morgenroth to study the effects of multi-axial prosthetic feet on medial lateral stability while walking in unilateral transtibial prosthesis users. Pilot data derived from the current award demonstrating that the Narrowing Beam Walking Test could discriminate between transtibial prosthesis users wearing multi-axial versus non-multi-axial feet, suggesting better balance performance in those with multi-axial feet, was included in the proposal (Figure 8). The PI of the

current award (Dr. Andrew Sawers) is a Co-I on that award. Additionally, discussions during the current award between the PI (Dr. Andrew Sawers) and Co-I (Dr. Brian Hafner) resulted in the submission and funding of a new OPORP award to address the need for better self-report fall outcome measures to study prosthetic interventions. In that award we will be developing and validating self-report measures of fall-related self-efficacy and interference in life activities. These measures will then been used as key

outcomes in the largest clinical trial of microprocessor knees to date. Dr. Sawers is also a Co-I on that award.

## **CHANGES/PROBLEMS**

None to report.

## **PRODUCTS**

### Journal publications

1. Sawers A, Hafner BJ. (2019) Conventional administration and scoring procedures suppress the diagnostic accuracy of a performance-based test designed to assess balance ability in lower limb prosthesis users. *Prosthet Orthot Int.* 43(4):402-408.
2. Kim, J, Hafner, BJ, Major M, Sawers A. (2019) Frequency and circumstances of falls reported by ambulatory unilateral lower limb prosthesis users: A secondary analysis. *PM & R* 11(4):344-353.
3. Sawers A, Hafner BJ. (2020) Using clinical balance tests to assess fall risk in established lower limb prosthesis users: cutoff scores and related validity indices. *PM & R* 12(1):16-25.
4. Sawers A, Kim J, Balkman GS, Hafner BJ. (2020). Inter-rater and test-retest reliability of performance-based clinical balance tests administered to established unilateral lower limb prosthesis users. *Physical Therapy* 100(7):1206-1216.
5. Sawers A, and Hafner BJ. (2020). Characterizing practice effects in performance-based tests administered to unilateral lower limb prosthesis users: A preliminary study. *PM&R (Online ahead of print)*.
6. Tobaigy M, Hafner BJ, Sawers A. (in review). Recalled number of falls in the prior year, combined with perceived mobility, predicts the incidence of future falls in established unilateral lower limb prosthesis users. *Prosthetics and Orthotics International*.
7. Sawers A, and Hafner BJ. (In review). Performance-based balance tests, combined with the number of falls recalled in the past year, predicts the incidence of future falls in established unilateral transtibial prosthesis users. *Archives of Physical Medicine and Rehabilitation*.
8. Sawers A, and Hafner BJ. (In preparation). Fall recall accuracy in people with lower limb amputation.
9. Tobaigy M, Hafner BJ, Sawers A. (In preparation). Fall related circumstances associated with consequential falls in lower limb prosthesis users.

## Conference presentations

1. Sawers A, Hafner, BJ. Ossur Instructional Course. Increasing awareness of knowledge around falls in lower limb prosthesis users. Reykjavik, Iceland, 2019 (*invited speaker*).
2. Sawers A, Kim J, Balkman GS, Hafner BJ. Orthotic and Prosthetic Innovative Technologies Conference. Evidence-based selection, use, and interpretation of performance-based clinical balance tests to improve fall risk assessment in lower limb prosthesis users. Ann Arbor MI, 2019 (*invited speaker*).
3. Sawers A, Kim J, Balkman GS, Hafner BJ. Scheck and Siress Prosthetics and Orthotics Education Fair. Reducing barriers to clinical selection, use, and interpretation of performance-based clinical balance tests. Chicago IL, 2019 (*invited speaker*).
4. Sawers A, Kim J, Major MJ, Hafner BJ. 45<sup>th</sup> American Academy of Orthotists and Prosthetists Annual Meeting and Scientific Symposium. Falls by lower limb prosthesis users: Refocusing through an epidemiological lens. Orlando, FL, 2019 (*invited symposium*).
5. Sawers, A. Northwestern University's Neuromechanics of Rehabilitation for Lower Limb Loss Symposium. Falls in lower limb prosthesis users: Refocusing through an epidemiological lens. Chicago, IL, 2018 (*invited speaker*).
6. Sawers, A., Kim, J., Hafner, B. Exploring practice effects in performance-based clinical tests administered to lower limb prosthesis users: Potential clinical significance. 8<sup>th</sup> Annual movement and Rehabilitation Science Training Day, Chicago, IL, 2019 (*Poster*).
7. Sawers A, Kim J, Major MJ, Hafner BJ. 45<sup>th</sup> American Academy of Orthotists and Prosthetists Annual Meeting and Scientific Symposium. Falls by lower limb prosthesis users: Refocusing through an epidemiological lens. Orlando, FL, 2019 (*invited symposium*).
8. Sawers A, Kim J, Balkman G, Hafner BJ. Deriving Reproducibility Indices To Facilitate Evidence-based Selection And Interpretation Of Performance-based Balance Tests For Lower Limb Prosthesis Users. Military Health System Research Symposium. Kissimmee, FL, 2019 (*Podium*).
9. Sawers A, Kim J, Hafner BJ. Administration and scoring procedures for performance-based clinical balance tests do not accommodate practice effects among lower limb prosthesis users. International Society of Posture and Gait Research World Congress. Edinburgh, Scotland, 2019 (*Poster*).

10. Kim J, Major MJ, Hafner BJ, Sawers A. The frequency and circumstances of falls reported by unilateral lower limb prosthesis users. International Society of Posture and Gait Research World Congress Edinburgh, Scotland, 2019 (*Poster*).
11. Sawers A, Kim J, Hafner BJ. Practice-effects in performance-based clinical balance tests limit their application among lower limb prosthesis users. American Congress of Rehabilitation Medicine 96<sup>th</sup> Annual Conference. Chicago, IL 2019 (*Podium*).
12. Kim J, Major MJ, Hafner BJ, Sawers A. Characterizing fall circumstances as reported by lower limb prosthesis users: A secondary analysis. American Congress of Rehabilitation Medicine 96<sup>th</sup> Annual Conference Chicago, IL 2019 (*Poster*).

#### Other Products

1. Walker N, Hafner BH, Sawers A. The American Academy of Orthotists and Prosthetists “How-To” Video Series. The Narrowing Beam Walking Test (NBWT). [Web Link](#)
2. Kim J, Hafner BH, Sawers A. (2019). The American Academy of Orthotists and Prosthetists Outcomes Research Committee Online Webinar. Classifying falls in patients that use lower limb prosthesis. [Web Link](#)

### **PARTICIPANTS AND OTHER COLLABORATING ORGANIZATIONS**

#### Individuals who worked on the project

|                              |                                                                                                                                                                |
|------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Name:                        | Andrew Sawers, PhD                                                                                                                                             |
| Project Role:                | Principal Investigator (UIC)                                                                                                                                   |
| Researcher Identifier:       |                                                                                                                                                                |
| Nearest person month worked: | 1                                                                                                                                                              |
| Contribution to Project:     | Dr. Sawers has been responsible for overseeing all aspects of the project (IRB, managing recruitment, enrollment, and data collection, manuscript preparation) |
| Funding Support:             | N/A                                                                                                                                                            |

|                              |                                                                                                                                                                                  |
|------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Name:                        | Brian Hafner, PhD                                                                                                                                                                |
| Project Role:                | Co-Investigator (UW)                                                                                                                                                             |
| Researcher Identifier:       |                                                                                                                                                                                  |
| Nearest person month worked: | 1                                                                                                                                                                                |
| Contribution to Project:     | Dr. Hafner has been responsible for overseeing aspects of the project at the UW study site (IRB, managing recruitment, enrollment, and data collection, manuscript preparation). |

|                  |     |
|------------------|-----|
| Funding Support: | N/A |
|------------------|-----|

|                              |                                                                                                                                                                                                                                                                                                                 |
|------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Name:                        | Janis Kim, MPT                                                                                                                                                                                                                                                                                                  |
| Project Role:                | Graduate Student (UIC)                                                                                                                                                                                                                                                                                          |
| Researcher Identifier:       |                                                                                                                                                                                                                                                                                                                 |
| Nearest person month worked: | 1                                                                                                                                                                                                                                                                                                               |
| Contribution to Project:     | Ms. Kim has performed work in the area of screening of participants, enrolling participants, consent, data collection, data entry, data analysis, and dissemination at the University of Illinois at Chicago study site. She has also managed the prospective fall follow up phone calls at the UIC study site. |
| Funding Support:             | N/A                                                                                                                                                                                                                                                                                                             |

|                              |                                                                                                                 |
|------------------------------|-----------------------------------------------------------------------------------------------------------------|
| Name:                        | Geoff Balkman                                                                                                   |
| Project Role:                | Graduate Student (UW)                                                                                           |
| Researcher Identifier:       |                                                                                                                 |
| Nearest person month worked: | <1                                                                                                              |
| Contribution to Project:     | Mr. Balkman has been responsible for data collection and data entry at the University of Washington study site. |
| Funding Support:             | N/A                                                                                                             |

Change in the active other support for the PD/PI(s) or senior/key personnel

Nothing to report

Other organizations involved as partners

Nothing to report

## **SPECIAL REPORTING REQUIREMENTS**

N/A

## **APPENDICES**

Appendix 1: Quad Chart

Appendix 2: Manuscripts

Appendix 3: NBWT Fabrication Guide

## Appendix 1: Quad Chart

# Can a novel beam-walking test improve fall risk assessment in Service Members, Veterans, and civilians who use lower limb prostheses?

OP160027

W81XWH-16-OPORP-PORA (Funding Level 1)

PI: Andrew Sawers, PhD, CPO

Org: The University of Illinois at Chicago

Award Amount: \$300,000



## Study Aim(s)

- Assess and compare the criterion validity of beam walking in lower limb prosthesis (LLP) users to existing balance tests.
- Assess and compare the reliability of beam walking in LLP users to existing clinical balance tests.
- Assess and compare the predictive validity of beam walking in LLP users to existing clinical balance tests.

## Approach

We will conduct cross-sectional (Aim 1) and longitudinal (Aims 2 & 3) studies to establish the validity and reliability of a novel beam walking-test to assess fall risk in LLP users. We will compare the ability of the beam-walking test with existing clinical balance tests to retrospectively discriminate between LLP users with and without a history of falls (Aim 1), and prospectively classify fallers and non-fallers over a 6 month follow-up period (Aim 3). The reliability of the beam-walking test will be tested between raters and days (Aim 2).

$$\# \text{ Future falls} = \exp [(0.202) * (\# \text{ falls recalled past year}) + (0.370) * (TUG) + (-0.154) * (FSST) - 2.90] \text{ [EQ. 2]}$$

By administering the Timed Up and Go and Four Square Step Test, and asking about the number of falls recalled over the past year, clinicians can use the predictive model (i.e., EQ. 2) to make predictions about the number of fall(s) that might be expected in the next 6 months for individual unilateral transtibial prosthesis users. For example, an established, unilateral transtibial prosthesis user with a 11.8 s Timed Up and Go time, a 9.7 s Four Square Step Test time, and who recalls having fallen twice in the past year, would be predicted to fall 1.46 times over the next 6 months.

$$\# \text{ Future falls} = \exp [(0.202) * (\# \text{ falls recalled past year}) + (0.370) * (TUG) + (-0.154) * (FSST) - 2.90]$$

$$\# \text{ Future falls} = \exp [(0.202) * (2) + (0.370) * (11.8) + (-0.154) * (9.7) - 2.90]$$

$$\# \text{ Future falls} = \exp [(0.202) * (2) + (0.370) * (11.8) + (-0.154) * (9.7) - 2.90]$$

$$\# \text{ Future falls} = \exp [3.77]$$

$$\# \text{ Future falls} = 1.46$$

**Equation.** Using negative binomial regression, and prospectively collected falls data from 60 lower limb prosthesis users, we developed a predictive model based on performance-based balance tests and the number of falls recalled over the past year that can predict the number of future falls transtibial prosthesis users are likely to experience in the next 6 months. This is the first prospectively derived model to predict falls in established transtibial prosthesis users. Importantly, the same model, or any model, did not predict future falls in transfemoral prosthesis users, suggesting a different approach to fall risk assessment in transfemoral prosthesis users may be required.

## Timeline and Cost

| Activities                            | CY           | 17          | 18           | 19          |
|---------------------------------------|--------------|-------------|--------------|-------------|
| Human subjects approval & train sites |              | ■           |              |             |
| Participant recruitment               |              |             | ■            |             |
| Conduct data collection procedure     |              | ■           | ■            |             |
| Analyze data and disseminate results  |              |             | ■            |             |
| <b>Estimated Budget (\$K)</b>         | <b>\$300</b> | <b>\$75</b> | <b>\$150</b> | <b>\$75</b> |

Updated: (01/07/2010)

## Goals/Milestones

**CY17 Goal** – Study preparation and participant recruitment

- ✓ Obtain human subjects approval
- ✓ Equip and train study staff at UIC & UW

**CY18 Goals** – Ongoing recruitment, data collection, and analysis

- ✓ Recruit additional 40 participants
- ✓ Conduct data collection and analysis procedures
- ✓ Disseminate initial results at national conference

**CY19 Goal** – Analysis and dissemination

- ✓ Recruit and collect data from final 10 participants
- ✓ Analyze final data set
- ✓ Disseminate final study results

## Comments/Challenges/Issues/Concerns

- None to report

## Budget Expenditure to Date

**\$300,000 (including F&A)**

## Appendix 2: Manuscripts

Original Research—CME

# Frequency and Circumstances of Falls Reported by Ambulatory Unilateral Lower Limb Prosthesis Users: A Secondary Analysis

Janis Kim, MPT, Matthew J. Major, PhD, Brian Hafner, PhD, Andrew Sawers, PhD, CPO

## Abstract

**Background:** More than 50% of lower limb prosthesis (LLP) users report falling at least once a year, placing them at high risk for adverse health outcomes such as decreased mobility and diminished quality of life. Efforts to decrease falls in LLP users have traditionally focused on developing clinical tests to assess fall risk, designing prosthetic components to improve patient safety, and identifying risk factors to recognize potential fallers. Little attention has been directed toward recording, reporting, and characterizing the circumstances of falls in LLP users. Identifying the most common types of falls could help guide and prioritize clinical and research needs.

**Objective:** To characterize the frequency and circumstances of falls reported by unilateral LLP users.

**Design:** Secondary analysis of data from 2 cross-sectional studies.

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

**Participants:** Ambulatory unilateral transtibial and transfemoral LLP users (N = 66).

**Intervention:** None.

**Outcome:** A fall-type classification framework was developed based on biomechanical theory and published falls terminology. Self-reported falls and accompanying narrative descriptions of LLP users' falls in the previous 12 months were analyzed with the framework. Frequencies, estimated proportions, and estimated counts were compared across fall circumstances using 95% confidence intervals.

**Results:** Thirty-eight participants (57.6%) reported 90 falls during the previous year. All reported falls were successfully categorized using the proposed framework. Most falls occurred from disruptions to the base of support, intrinsic destabilizing factors, and a diverse set of fall patterns. Walking on level terrain was the most common activity at the time of a fall.

**Conclusion:** This secondary analysis showed that falls remain frequent in ambulatory LLP users and that clinicians and researchers might wish to prioritize falls owing to disruptions of the base of support that occur while walking. Additional research with a larger sample is required to confirm and expand these results.

**Level of Evidence:** III

## Introduction

More than 50% of lower limb prosthesis (LLP) users report falling at least once a year [1-5], placing them at high risk for adverse health outcomes such as decreased mobility and diminished quality of life [3,6,7]. Attempts to lower the prevalence of falls in LLP users have traditionally focused on developing and validating clinical tests to assess fall risk [8-12], designing and testing prosthetic components to improve patient safety [13-16], characterizing the biomechanics of key balance

strategies to identify deficits in those at risk for falls [17-23], and identifying risk factors to help recognize potential fallers [2,5,7,24,25]. However, little attention has been directed toward recording, reporting, and characterizing the circumstances of falls in LLP users [1,4,7]. For example, are falls caused primarily by external factors imposed by the environment or are falls initiated by internal physiologic factors? Are LLP users more susceptible to trips, slips, or prosthetic failures? What activities are LLP users most commonly engaged in at the time of a fall? Answers to these questions could help direct

treatment to the most prevalent and consequential types of falls [26-29], prioritize research needs in areas related to fall assessment (eg, screening methods and diagnostic tests) [30], and generate evidence to develop and revise reimbursement policies (eg, prosthetic components that decrease falls). Therefore, the primary objective of this project was to provide an initial characterization of the circumstances of falls in ambulatory unilateral LLP users.

The characterization of fall-related circumstances in LLP users is limited by the lack of a universal fall-type classification framework and taxonomy suited to LLP users. The use of incomplete and study-specific terminology [1,4] has limited comparisons between studies and prevented a more comprehensive characterization of fall circumstances in LLP users. Several classification frameworks have been used to characterize falls in older adults [26,31,32], but they are not entirely applicable to the experiences of LLP users [33]. A fall-type classification framework for LLP users would provide a structured approach to soliciting, recording, reporting, and studying falls experienced by LLP users. Therefore, a secondary objective of this project was to propose and evaluate an LLP user-specific fall-type classification framework and taxonomy.

## Methods

### Study Design

A secondary analysis of self-report falls data collected in 2 previous studies [11,12] was conducted to address the study objectives. Data were combined owing to similar study instruments, data collection methods, and participants. The 2 studies sought to evaluate the validity of clinical balance tests in ambulatory LLP users. No a priori power analysis was performed for this secondary analysis. Original study protocols were reviewed and approved by institutional review boards at the University of Illinois at Chicago and Northwestern University. All individuals provided written informed consent before participation.

## Participants

Participants who met the following inclusion and exclusion criteria were selected from the datasets in the 2 studies [11,12]: age at least 18 years; unilateral, transtibial, or transfemoral amputation; at least 1 year since amputation; and able to walk with or without an assistive device. Exclusion criteria included congenital or upper extremity limb loss and comorbidities limiting mobility (eg, knee replacement, severe pulmonary disease, ulcers).

## Measurements

### Participant Demographics, Prosthetic-related Information, and Retrospective Falls Data

Data on age, gender, amputation level, cause of amputation, and time since amputation were collected by interview with a study investigator. Study investigators also measured height and weight. To assess fall history, each participant was asked, "In the past 12 months have you had any falls including a slip or trip in which you inadvertently lost your balance and landed on the ground or lower level?" [34]. Participants reporting a fall were asked to provide a narrative description of the event (ie, circumstances of their fall[s]) [29,31].

### Development of Fall-type Classification Framework and Taxonomy

A fall-type classification framework and taxonomy was developed based on biomechanical theory, published terminology, and fall descriptions reported by or observed among LLP users [1,4,33] and older adults [26,27,29,31,35-38]. Fall circumstances reported by older adults were referenced when developing the framework to supplement the limited reporting of fall circumstances in LLP users.

The fall-type classification framework (Figure 1) is a 3-level hierarchical system describing the location of the destabilizing force, the source of the destabilization, and the ensuing fall pattern. The first level of the

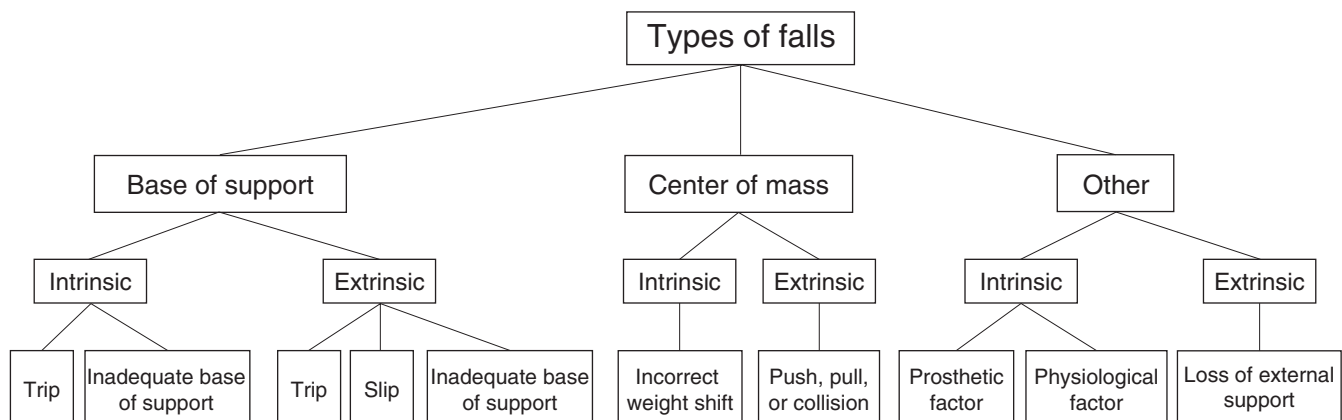


**Figure 1.** Fall-type classification framework. The proposed framework is a 3-level hierarchical system that characterizes falls based on the location of the destabilizing force (level 1), the source of the destabilizing force (level 2), and the fall pattern (level 3).

framework characterizes the location of the destabilization with respect to the body. First-level descriptors are based on a simple biomechanical model in which falling results from uncorrected disruptions between the base of support (BoS) and the center of mass (CoM) [27,29,39-42]. Disruption can occur from a destabilizing force acting to displace the BoS (ie, feet) from beneath the CoM (ie, a BoS fall) or displace the CoM (ie, body) beyond its existing BoS (ie, a CoM fall) [4,27,29,33,36]. Falls with no apparent biomechanical disruption between the BoS and the CoM are classified as “other” types of falls [27]. Characterizing the most common location of destabilizing forces leading to a fall could help determine the most suitable targets for mitigating fall risk through balance training, clinical balance assessment, and/or prosthetic design [1,28].

The second level of the framework characterizes the source of destabilization. A destabilizing force that acts on the CoM or BoS can originate from intrinsic or extrinsic sources. Intrinsic sources are personal factors, often physiologic in nature (eg, muscle weakness). Extrinsic sources are external factors imposed by the environment that result in unexpected disruptions (eg, icy walkway, being bumped by someone) [1,29,31,35,43]. Categorizing falls as intrinsic or extrinsic can guide clinical decision making in several ways. First, it can highlight potential preventative strategies. Falls attributed to extrinsic sources are often addressed through patient education and/or environmental modifications [44-46]. Falls attributed to intrinsic sources might require interventions that target specific neuromusculoskeletal deficits. Second, individuals who experience falls from intrinsic sources are more likely to fall again because such sources often indicate systemic issues [47]. Thus, categorizing falls as intrinsic or extrinsic could help identify individuals who are likely to be recurrent fallers. Third, contemporary clinical balance tests largely probe intrinsic over extrinsic factors. If most falls experienced by LLP users are attributed to extrinsic sources of destabilization, then clinically feasible tests that replicate such extrinsic sources of destabilization might need to be developed.

The third level of the framework describes fall patterns [1,20,28]. For example, a destabilizing force acting on the BoS from an extrinsic source can produce a fall from a trip, a slip, or an inadequate BoS [26]. This level of the framework also includes prosthetic-related factors (eg, prosthetic knee buckling) [1], physiologic factors (eg, collapse from muscle weakness), and loss of external support (eg, support structure or assistive device moving unexpectedly) [26,31]. Characterizing fall patterns could be important for selecting treatment and preventative strategies with a level of specificity that is needed for balance training, yet beyond what is offered by the first 2 levels of the framework [28,48-51]. If certain fall patterns are more prevalent, then resources should be directed toward their investigation. Inclusion of fall patterns within the framework also introduces terminology

consistent with that used by clinicians and patients. A complete set of definitions and examples for each component of the fall-type classification framework is provided in Appendix S1.

### *Classifying Falls*

Two reviewers independently classified each reported fall using the fall-type framework. Participants' narrative descriptions of fall events were analyzed to determine whether the BoS or CoM was disrupted (level 1). If participants described neither, then the fall was assigned to “other.” Each type of level 1 fall was further characterized based on whether participants described extrinsic (ie, environmental) or intrinsic (ie, personal) factors as the source of the destabilization (level 2). If intrinsic and extrinsic factors were reported, then the source was categorized as intrinsic [31]. Falls were subsequently classified with greater detail based on reported movements during the fall (level 3). For example, if a participant reported catching his or her foot in a crack in the sidewalk and fell forward, then the fall was classified as a BoS extrinsic trip.

In addition to fall type, activity at the time of the fall was categorized. The same 2 reviewers extracted details of the specific form of activity at the time of the fall from the narrative description provided by each participant. Previously reported activities at the time of a fall (eg, walking, transfer, reaching) [7,26,33,36] were used as an initial guideline for categorization. Falls were classified as “unknown” when participants could not recall the events concerning the fall.

### *Statistical Analysis*

For each fall type and activity at the time of a fall, we tabulated the total number of reported falls (ie, frequency), the proportion of participants falling at least once (ie, estimated proportion), and the average number of falls per participant (ie, estimated count). Frequencies were expressed as a number and a percentage of all reported falls. Estimated proportions were calculated as the ratio of participants who reported at least 1 fall for each type or activity with respect to all participants, multiplied by 100. Estimated counts were calculated as the ratio of all falls reported for each type or activity with respect to the total number of participants. To determine whether there were any differences in the estimated proportions or counts across the various types of falls and activities at the time of a fall, 95% confidence intervals (CIs) were computed from a *t*-distribution and compared across fall types and activities. Overlap between 95% CIs was taken as a conservative estimate of no significant difference. All statistical analyses were performed using SPSS 24 (SPSS, Inc, Chicago, IL).

We also sought to ascertain the ability of the proposed fall-type classification framework to accurately capture the range of fall types reported by LLP users. To that

end, after classification, the framework was reviewed for (1) categories that did not align with any reported fall events, which would suggest they might be unnecessary and could be removed, and (2) falls that did not fit into the framework, which would suggest that additional categories might be required.

## Results

### Participants

Records of 70 LLP users who participated in the original studies [11,12] were reviewed. Three participants were excluded because of amputation level (2 bilateral; 1 ankle disarticulation), and the fourth participant was excluded because time since amputation was less than 1 year. Data from 66 participants were included (Table 1).

### Fall Prevalence

Sixty-six participants reported a total of 90 falls (ie, 1.36 falls per participant). Thirty-eight participants (57.6%) reported at least 1 fall in the past 12 months. Fourteen participants (22.1%) reported 1 fall, 9 (13.6%) reported 2 falls, and 15 (22.7%) reported at least 3 falls in the past 12 months. Based on the narrative description provided by the participants of their fall(s) and activities at the time of their falls, all reported falls were considered to have occurred while wearing the prosthesis.

### Suitability of Fall-type Classification Framework

Eight falls (9%) were categorized as “unknown” (Figure 2), 7 owing to incomplete data collection forms and 1 owing to a participant’s inability to recall the specific circumstances of a fall. Three framework categories did not align with at least 1 reported fall event: “a loss of external support,” “push, pull, or collision,” and “inadequate base of support.”

### BoS vs CoM Falls

The BoS was the most commonly reported location of a destabilizing force leading to a fall, accounting for 54%

(49 of 90) of all reported falls (Figure 2). The proportion of participants (36.4%, 95% CI 26.4-48.4) and the average number of falls per participant (0.75, 95% CI 0.63-0.85) owing to disruption of the BoS were significantly larger than those arising from disruption of the CoM (proportion 9.1%, 95% CI 1.9-16.3; average number per participant 0.18, 95% CI 0.09-0.28) or those with no apparent disruption to the BoS or CoM (ie, “other”; proportion 16.7%, 95% CI 7.4-26.0; average number per participant 0.32, 95% CI 0.20-0.43; Table 2).

### Intrinsic vs Extrinsic Falls

Intrinsic (ie, personal) factors were the most commonly reported source of destabilizing forces leading to a fall (Figure 2). Fifty-two of the 90 reported falls (58%) were attributed to intrinsic sources (eg, missed step, poor foot clearance), whereas 30 of the reported falls (33%) were attributed to extrinsic sources (eg, icy surface, cracked sidewalk, uneven terrain; Figure 2). However, neither the proportion of participants reporting a fall nor the average number of falls per participant was significantly different between intrinsic and extrinsic sources (Table 2).

### Fall Patterns

Slips were the most commonly reported fall pattern, accounting for 23 of the 90 reported falls (26%). This was followed by trips and prosthetic factors, each accounting for 22% of all reported falls (Figure 2). The proportion of participants reporting slips (18.2%, 95% CI 8.6-27.8) and trips (16.7%, 95% CI 7.4-26.0) and the average number of slips (0.35, 95% CI 0.23-0.47) and trips (0.30, 95% CI 0.19-0.42) per participant were significantly larger than the proportion of participants reporting falls from physiologic factors (1.5%, 95% CI 0-4.6) and the number of falls per participant from physiologic factors (0.02, 95% CI 0-0.05) or an inadequate BoS (0.09, 95% CI 0-0.16). No other significant differences in the proportion of participants falling or number of falls per participant were identified across fall patterns (Table 2).

**Table 1**  
Sociodemographic and amputation-related characteristics

|       | Age (y) | Height (cm) | Weight (kg) | Time Since Amputation (y) | Gender             | Amputation Level                  | Amputation Etiology                                                     |
|-------|---------|-------------|-------------|---------------------------|--------------------|-----------------------------------|-------------------------------------------------------------------------|
| Mean  | 50.6    | 173         | 81.6        | 16.3                      | 37 men<br>29 women | 36 transtibial<br>30 transfemoral | 39 trauma<br>13 dysvascular<br>5 cancer<br>7 infections<br>2 congenital |
| SD    | 14.1    | 8.65        | 16.9        | 13.0                      |                    |                                   |                                                                         |
| Range | 21-76   | 155-189     | 104-266     | 1.6-57                    |                    |                                   |                                                                         |

SD = standard deviation.

|                           |                             |                |                       |                        |                                        |                                |                   |
|---------------------------|-----------------------------|----------------|-----------------------|------------------------|----------------------------------------|--------------------------------|-------------------|
| Level 1<br>(location)     | Base of Support<br>49 (54%) |                |                       |                        | Center<br>of Mass<br>12 (13%)          | Other<br>21 (23%)              | Unknown<br>8 (9%) |
| Level 2<br>(source)       | Extrinsic<br>30 (33%)       |                | Intrinsic 52 (58%)    |                        |                                        |                                |                   |
|                           |                             |                | Intrinsic<br>19 (21%) | Intrinsic<br>12 (13%)  | Intrinsic<br>21 (23%)                  |                                |                   |
| Level 3<br>(fall pattern) | Slip<br>23 (26%)            | Trip 20 (22%)  |                       | Small<br>BoS<br>6 (7%) | Inadequate<br>Weight Shift<br>12 (13%) | Prosthetic Factors<br>20 (22%) |                   |
|                           |                             | Trip<br>7 (8%) | Trip<br>13 (14%)      |                        |                                        |                                |                   |
| Physiological 1 (1%)      |                             |                |                       |                        |                                        |                                |                   |

**Figure 2.** Number and frequency of fall types in lower limb prosthesis users as categorized by the classification framework. Base of support was the most common location of a destabilizing force resulting in a fall (level 1). Intrinsic factors were the most common source of a destabilizing force resulting in a fall (level 2). Slip, trips, and prosthetic factors accounted for nearly three-fourths of reported fall patterns (level 3).

### Activities Performed at the Time of a Fall

Participants most commonly reported walking on a level terrain when they experienced a fall. Falls that occurred while participants were walking on a level terrain accounted for 45.6% (41 of 90) of all falls (Figure 3). The proportion of participants who reported falling while walking on a level terrain (31.8%, 95% CI 20.2-43.4) and the mean number of falls while walking per participant (0.62, 95% CI 0.50-0.74) were significantly larger than falls that occurred while participants' were engaged in any other reported activity (eg, walking on an uneven terrain, stairs, transfers; Table 3). Thirteen (14%) falls occurred during an "unknown" activity (Figure 3).

### Discussion

This secondary analysis was conducted to provide an initial characterization of the circumstances of falls in

ambulatory unilateral LLP users and propose a fall-type classification framework. The prevalence of falls in this sample of LLP users exceeded 50%. Most of these falls occurred from disruptions to the BoS, intrinsic destabilizing factors, and a diverse set of fall patterns (eg, slips, trips, inadequate weight shift). Although additional work is necessary to confirm and expand these results in a larger sample, the present findings represent an important first step in developing specific targets for clinical assessment and treatment and future research directions.

### Proposed Classification Framework Successfully Categorized Fall Types Reported by LLP Users

The proposed framework offers 3 advantages. First, the 2 reviewers independently arrived at and agreed on the fall categorizations, suggesting that the framework and accompanying taxonomy are reliable. Second, the taxonomy appears comprehensive. All reported falls

**Table 2**

Proportion of participants who fell at least once and average number of falls per participant from various fall types\*

|                                        | Proportion of Participants Who Fell |           | Number of Falls per Participant |           |
|----------------------------------------|-------------------------------------|-----------|---------------------------------|-----------|
|                                        | Estimated Proportion, % (SE)        | 95% CI    | Estimated Count, n (SE)         | 95% CI    |
| <b>Location of destabilizing force</b> |                                     |           |                                 |           |
| Base of support                        | 36.4 (5.9)                          | 26.4-48.4 | 0.75 (0.05)                     | 0.63-0.85 |
| Center of mass                         | 9.1 (3.5)                           | 1.9-16.3  | 0.18 (0.05)                     | 0.09-0.28 |
| Other                                  | 16.7 (4.6)                          | 7.4-26.0  | 0.32 (0.06)                     | 0.20-0.43 |
| <b>Source of destabilizing force</b>   |                                     |           |                                 |           |
| Extrinsic                              | 27.3 (5.5)                          | 16.2-38.4 | 0.46 (0.06)                     | 0.33-0.58 |
| Intrinsic                              | 39.4 (6.0)                          | 27.2-51.6 | 0.78 (0.05)                     | 0.69-0.89 |
| <b>Fall pattern</b>                    |                                     |           |                                 |           |
| Slip                                   | 18.2 (4.7)                          | 8.6-27.8  | 0.35 (0.06)                     | 0.23-0.47 |
| Trip                                   | 16.7 (4.6)                          | 7.4-26.0  | 0.30 (0.06)                     | 0.19-0.42 |
| Inadequate weight shift                | 9.1 (3.5)                           | 1.9-16.3  | 0.18 (0.05)                     | 0.09-0.28 |
| Inadequate base of support             | 7.6 (3.3)                           | 0-14.2    | 0.09 (0.04)                     | 0-0.16    |
| Prosthetic factors                     | 15.2 (4.4)                          | 6.2-24.1  | 0.30 (0.06)                     | 0.19-0.42 |
| Physiologic factors                    | 1.5 (1.5)                           | 0-4.6     | 0.02 (0.02)                     | 0-0.05    |
| Unknown                                | 10.6 (3.8)                          | 2.9-18.3  | 0.12 (0.04)                     | 0.04-0.20 |

SE = standard error; CI = confidence interval.

\* Estimated proportions and counts might not sum across fall types to 100% or 1.0, respectively, because participants can have more than 1 fall type if they reported multiple falls.

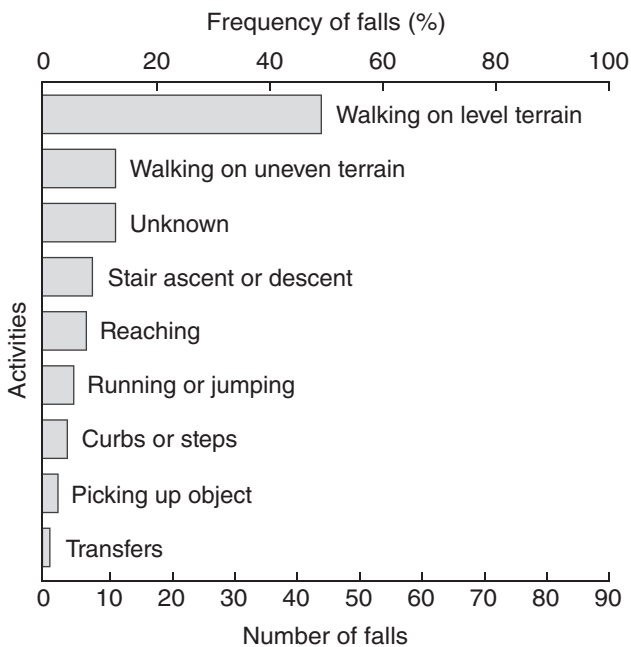


**Figure 3.** Number and frequency of falls reported by lower limb prosthesis users based on the activity at the time of a fall. Walking was the most commonly reported activity at the time of a fall. More falls occurred while walking on level vs uneven terrain.

were successfully categorized. Only 8 (9%) of the reported falls were categorized as “unknown,” well below 19% in older adults [52]. Third, 3 fall patterns, “push, pull, or collision,” “loss of external support,” and “inadequate base of support,” were not used. These fall patterns have been previously reported by LLP users [4] and older adults [26]; as such, their removal from the framework is not recommended until they can be assessed in a larger sample.

### Falls Remain Frequent in LLP Users

Historically, more than 50% of LLP users report falling at least once a year [1,2,4,5,25]. The results of this secondary analysis showed that the prevalence of falls in

community-living ambulatory LLP users remains high, at 57.6%. The prevalence of recurrent falls, 36.3%, also remains at or above historically reported values for LLP users (ie, 23.4%-39.0%) [1,3,5]. The relevance of the increased frequency of falls in LLP users should be judged not only on its magnitude alone but also on the severity and occurrence of adverse outcomes. Between 8.0% and 57.0% of LLP users report a fall-related injury [1,4,7,25,53]; 14.6% seek medical attention and 7.3% require surgery [53] after a fall. Coupled with the importance LLP users place on balance [54], these consequences indicate that falls negatively affect the lives of a substantial portion of LLP users. Research exploring new approaches guided by the types of falls experienced by LLP users might be needed to decrease falls in LLP users.

### LLP Users Are More Susceptible to BoS Than CoM Falls

BoS falls are more common than CoM falls in LLP users. These findings are consistent with those in older adults without lower limb amputation, in whom BoS falls consistently outnumber CoM falls [27,31,36-38]. Environmental barriers and/or limitations in modern prosthetic components could contribute to the prevalence of BoS falls in LLP users. Environmental hazards that disrupt the BoS (eg, cracks in the sidewalk) might be more common than those that disrupt the CoM (eg, collisions in crowds). The lack of active ankle dorsiflexion in most contemporary prosthetic feet could limit the functional shortening of the prosthetic leg required during the swing phase to achieve sufficient toe clearance, increasing the likelihood of a BoS disruption (ie, trip) [13]. Minimum toe clearance of the prosthetic leg during the swing phase is 50% lower in transtibial LLP users with a history of trip-related (ie, BoS) falls than without [33]. LLP users also could be susceptible to BoS falls because of limited or ineffective responses to BoS disruptions. For example, when the prosthetic leg is obstructed during a trip, transfemoral LLP users only select the “lowering” strategy, ignoring the “elevating” strategy, even when the

**Table 3**

Proportion of participants who fell at least once and average number of falls per participant for activities reported at the time of a fall\*

|                           | Proportion of Participants Who Fell |           | Number of Falls per Participant |           |
|---------------------------|-------------------------------------|-----------|---------------------------------|-----------|
|                           | Estimated Proportion, % (SE)        | 95% CI    | Estimated Count, n (SE)         | 95% CI    |
| Walking on level terrain  | 31.8 (5.7)                          | 20.2-43.4 | 0.62 (0.06)                     | 0.50-0.74 |
| Unknown                   | 12.1 (4.0)                          | 4.0-20.3  | 0.20 (0.05)                     | 0.10-0.30 |
| Stairs                    | 10.6 (3.8)                          | 2.9-18.3  | 0.11 (0.04)                     | 0.03-0.18 |
| Walking on uneven terrain | 9.1 (3.5)                           | 1.9-16.3  | 0.20 (0.05)                     | 0.10-0.30 |
| Running or jumping        | 6.1 (2.9)                           | 0.1-12.0  | 0.06 (0.03)                     | 0-0.12    |
| Reaching                  | 6.1 (2.9)                           | 0-12.0    | 0.09 (0.04)                     | 0.02-0.16 |
| Curbs or steps            | 3.0 (2.1)                           | 0-7.3     | 0.05 (0.03)                     | 0-0.10    |
| Picking up object         | 1.5 (1.5)                           | 0-4.6     | 0.03 (0.02)                     | 0-0.07    |
| Transfers                 | 1.5 (1.5)                           | 0-4.6     | 0.02 (0.02)                     | 0-0.05    |

SE = standard error; CI = confidence interval.

\* Estimated proportions and counts might not sum across activities to 100% or 1.0, respectively, because participants can be in more than 1 group if they reported multiple falls.

“elevating” strategy is more suitable [20]. In addition, the “lowering” strategy is often unsuccessful in transfemoral prosthesis users because of excessive stance phase flexion (ie, buckling) of the prosthetic knee [20,55]. Therefore, new prosthetic designs with active powered control might be necessary for LLP users to successfully use the full range of available balance response strategies [17]. Clinical balance tests that probe responses to BoS disruptions [27] might be needed to improve fall risk assessment in LLP users [30].

### ***Prevalence of Intrinsic Falls in LLP Users Could Explain the High Rate of Recurrent Falls***

Consistent with prior research [1], the results of this secondary analysis confirmed that ambulatory LLP users experience more falls initiated by intrinsic personal factors than imposed by extrinsic environmental factors. A larger proportion of intrinsic falls are typically associated with advanced age (ie, >75 years old) [31,43,45,56]. However, the relatively young mean age of the present sample (50.4 years) suggests that intrinsic falls are unlikely to be attributable to age. Instead, challenges and demands unique to prosthetic ambulation, such as controlling a prosthetic knee (ie, “prosthetic factors”) and a decrease in strength [3,28,57], are more likely to be responsible for intrinsic falls. The prevalence of intrinsic falls, which are associated with systemic issues and multiple falls [47], could explain the high rate of recurrent falls (36.3%) that were observed in this analysis. In addition, by placing greater emphasis and quickly identifying and treating intrinsic factors related to fall risk, clinicians could decrease recurrent falls in LLP users. Although modifiable (eg, pain, general health, mobility, strength) [3,53,57] and non-modifiable (eg, level of amputation, cause of amputation, increasing age, altered somato-sensation, and gender) [3,53,57] intrinsic personal factors have been associated with increased fall risk, a history of falls, fall-related injuries in LLP users, other important and modifiable intrinsic factors including prosthetic design, and reaction time [21,58] remain unexamined. How intrinsic factors change across stages of recovery [59] and their synthesis into a clinical test for predicting susceptibility to intrinsic falls could improve fall risk assessment. A full characterization of intrinsic personal factors and their association with falls in LLP users is warranted [60].

### ***LLP Users Reported Different Fall Patterns***

Nearly three-fourths of the falls reported by LLP users were categorized as slips (26%), trips (22%), or prosthetic factors (22%). The prevalence of slips and trips corroborates the findings of Rosenblatt et al [33] who reported that, in a small sample of transtibial LLP users, one-fourth of falls were trips and one-fourth were slips. Kulkarni et al [1] reported that 12% of falls in their sample of LLP users were attributed to prosthetic-related factors,

approximately half that in the present analysis. These prosthetic-related falls in the prior study included “prosthetic failures” (eg, foot breaking) and “prostheses not working as expected.” This suggests that LLP users might be attempting activities that their prostheses do not allow them to do and that new components might increase patient safety. These results also expand the diversity of fall patterns previously identified. Although previously reported in older adults [26,27], incorrect weight shifts and a small BoS, which accounted for 20% of all reported falls, had not previously been described in LLP users. Owing to this variety, neither assessment nor treatment of falls in LLP users can currently be prioritized based on the prevalence of a specific fall pattern.

### ***Walking Was the Most Common Activity at the Time of a Fall***

Participants in this study were generally engaged in some form of locomotion, mainly on a level terrain, at the time of a fall. This is consistent with a previous study of a small sample transtibial LLP users in which most falls and stumbles occurred while walking, 83% of which were on level ground [33]. The prevalence of falls while walking is likely attributable to the prominence of walking in our lives and the inherently unstable biomechanics of walking [61]. Falls in LLP users might be more common on a level terrain owing to opportunity (ie, there is more level than uneven terrain), avoiding an uneven terrain, or gait modifications made to preserve “stability” on an uneven terrain. Although challenging [62], when walking on an uneven terrain LLP users adopt a cautious, conservative gait pattern (ie, decreased speed, wider BoS, increased double support time, lowered CoM, and increased upper extremity motion) [22,63-66]. Although potentially imposing additional costs or demands (ie, metabolic), these changes appear to mitigate additional risk associated with walking over an uneven terrain [22,66].

### ***Limitations***

LLP users were asked to recall fall events during the past 12 months (ie, self-report). This could have introduced recall bias and led to an underestimation of the prevalence of falls [67,68]. There is currently no consensus regarding recall period when asking LLP users about falls. In addition, commonly adopted definitions of a fall [1,3,5] do not contain prosthetic-specific language (eg, “a fall with or without your prosthesis”). Similar to guidelines proposed for older adults [69], future research could benefit from establishing the time window during which LLP users can accurately recall fall events and their related details and a fall definition that is meaningful to LLP users.

The present analysis is limited by our understanding of the situation in which falls occurred and the role, if any, of prosthetic componentry on fall type. For example, the mental (eg, distraction, urgency) or physical (eg, fatigue) state of

LLP users at the time of a fall remains unknown. Additional research characterizing these situations could clarify why falls while walking on a level terrain were prevalent. Future research would benefit from a clearer separation of activity and physical environment and from additional details characterizing the physical environment at the time of a fall (eg, lighting, time of day, indoor or outdoor). Likewise, the role of prosthetic componentry in contributing to or decreasing certain fall types remains to be determined. A comprehensive analysis of the physical, social, economic, and psychological consequences of falls in LLP users is needed to better determine the impact of falls in LLP users. Characterizing fall consequences also could provide a more robust approach to prioritizing clinical and research directions with respect to specific fall patterns.

## Conclusion

The primary objective of this secondary analysis was to provide an initial characterization of the circumstances concerning falls in ambulatory unilateral LLP users. The results of this secondary analysis suggest that falls remain frequent in ambulatory LLP users and that clinicians and researchers might wish to prioritize falls from disruptions of the BoS that occur while walking. The secondary objective was to propose and evaluate an LLP user-specific fall-type classification framework and taxonomy. The proposed fall-type classification framework successfully classified all self-reported falls, suggesting that it is comprehensive. Future falls research in LLP users might wish to use a similar taxonomy and framework. With the frequency of falls remaining higher than 50% and the rate of amputation expected to increase [70], falls in LLP users are likely to be a substantial burden on the U.S. health care system. An understanding of the circumstances and consequences associated with falls could lead to previously unidentified targets for assessing and treating fall risk.

## Supplementary Data

Supplementary data associated with this article can be found in the online version at <https://doi.org/10.1016/j.pmrj.2018.08.385>.

## References

- Kulkarni J, Wright S, Toole C, Morris J, Hiron R. Falls in patients with lower limb amputations: Prevalence and contributing factors. *Physiotherapy*. 1996;82:130-136.
- Gauthier-Gagnon C, Grisé MC, Potvin D. Enabling factors related to prosthetic use by people with transtibial and transfemoral amputation. *Arch Phys Med Rehabil*. 1999;80:706-713.
- 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;82:1031-1037.
- Ülger Ö, Topuz S, Bayramlar K, Erbahçeci F, Sener G. Risk factors, frequency, and causes of falling in geriatric persons who has had a limb removed by amputation. *Top Geriatr Rehabil*. 2010;26:156-163.
- 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;47:80-86.
- McWhinnie DL, Gordon AC, Collin J. Rehabilitation outcome 5 years after 100 lower-limb amputations. *Br J Surg*. 1994;81:1596-1599.
- Pauley T, Devlin M, Heslin K. Falls sustained during inpatient rehabilitation after lower limb amputation: Prevalence and predictors. *Am J Phys Med Rehabil*. 2006;85:521-532.
- Schoppen T, Boonstra A, Groothoff JW, de Vries J, Goeken L, Eisma WH. The Timed "Up and Go" test: Reliability and validity in persons with unilateral lower limb amputation. *Arch Phys Med Rehabil*. 1999;80:825-828.
- 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;84:656-661.
- Dite W, Connor HJ, Curtis HC. Clinical identification of multiple fall risk early after unilateral transtibial amputation. *Arch Phys Med Rehabil*. 2007;88:109-114.
- 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;94:2194-2202.
- Sawers A, Hafner BJ. A study to assess whether fixed-width beam walking provides sufficient challenge to assess balance ability across lower limb prosthesis users. *Clin Rehabil*. 2018;32:483-492.
- Rosenblatt NJ, Bauer A, CPO DR, CPO MDG. Active dorsiflexing prostheses may reduce trip-related fall risk in people with transtibial amputation. *J Rehabil Res Dev*. 2014;51:1229-1242.
- Hafner BJ, Willingham LL, Buell NC, Allyn KJ, Smith DG. Evaluation of function, performance, and preference as transfemoral amputees transition from mechanical to microprocessor control of the prosthetic knee. *Arch Phys Med Rehabil*. 2007;88:207-217.
- Hafner BJ, Smith DG. Differences in function and safety between Medicare Functional Classification Level-2 and -3 transfemoral amputees and influence of prosthetic knee joint control. *J Rehabil Res Dev*. 2009;46:417-434.
- Kahle JT, Highsmith MJ, Hubbard SL. Comparison of nonmicroprocessor knee mechanism versus C-Leg on Prosthesis Evaluation Questionnaire, stumbles, falls, walking tests, stair descent, and knee preference. *J Rehabil Res Dev*. 2008;45:1-14.
- Shirota C, Simon AM, Kuiken TA. Transfemoral amputee recovery strategies following trips to their sound and prosthesis sides throughout swing phase. *J Neuroeng Rehabil*. 2015;12:79.
- D. Segal A, K. Klute G. Lower-limb amputee recovery response to an imposed error in mediolateral foot placement. *J Biomech*. 2014;47:2911-2918.
- Bolger D, Ting LH, Sawers A. Individuals with transtibial limb loss use interlimb force asymmetries to maintain multi-directional reactive balance control. *Clin Biomech*. 2014;29:1039-1047.
- Crenshaw JR, Kaufman KR, Grabiner MD. Trip recoveries of people with unilateral, transfemoral or knee disarticulation amputations: Initial findings. *Gait Posture*. 2013;38:534-536.
- Curtze C, Hof AL, Otten B, Postema K. Balance recovery after an evoked forward fall in unilateral transtibial amputees. *Gait Posture*. 2010;32:336-341.
- Curtze C, Hof AL, Postema K, Otten B. The relative contributions of the prosthetic and sound limb to balance control in unilateral transtibial amputees. *Gait Posture*. 2012;36:276-281.
- Yang J, Jin D, Ji L, et al. The reaction strategy of lower extremity muscles when slips occur to individuals with trans-femoral amputation. *J Electromyogr Kinesiol*. 2007;17:228-240.
- 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;29:415-420.
- 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;82:1238-1244.

26. Robinovitch SN, Feldman F, Yang Y, et al. Video capture of the circumstances of falls in elderly people residing in long-term care: An observational study. *Lancet*. 2013;381:47-54.
27. 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;41:479-487.
28. Vlahov D, Myers AH, Al-Ibrahim MS. Epidemiology of falls among patients in a rehabilitation hospital. *Arch Phys Med Rehabil*. 1990;71:8-12.
29. Wild D, Nayak US, Isaacs B. Description, classification and prevention of falls in old people at home. *Rheumatol Rehabil*. 1981;20:153-159.
30. Hart-Hughes S, Latlief GA, Phillips S, Groer S, Highsmith MJ. A review of clinical outcome assessment instruments for gait, balance, and fall risk in persons with lower extremity amputation. *Top Geriatr Rehabil*. 2014;30:70-76.
31. Lach HW, Reed AT, Arfken CL, et al. Falls in the elderly: Reliability of a classification system. *J Am Geriatr Soc*. 1991;39:197-202.
32. Bleijlevens MHC, Diederiks JPM, Hendriks MRC, van Haastregt JCM, Crebolder HFJM, van Eijk JTM. Relationship between location and activity in injurious falls: An exploratory study. *BMC Geriatr*. 2010;10:40.
33. Rosenblatt NJ, Bauer A, Grabiner MD. Relating minimum toe clearance to prospective, self-reported, trip-related stumbles in the community. *Prosthet Orthot Int*. 2016;41:387-392.
34. Askham J, Glucksman E, Owens P, Swift C, Tinker A, Yu G. Home and leisure accident research: a review of research on falls among elderly people. *Age Concern Institute of Gerontology, King's College London, and Department of Trade and Industry*. London, UK; 1990.
35. Haga H, Shibata H, Shichita K, Matsuzaki T, Hatano S. Falls in the institutionalized elderly in Japan. *Arch Gerontol Geriatr*. 1986;5:1-9.
36. Berg WP, Alessio HM, Mills EM, Tong C. Circumstances and consequences of falls in independent community-dwelling older adults. *Age Ageing*. 1997;26:261-268.
37. 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;23:41-48.
38. Crenshaw JR, Bernhardt KA, Achenbach SJ, et al. The circumstances, orientations, and impact locations of falls in community-dwelling older women. *Arch Gerontol Geriatr*. 2017;73:240-247.
39. MacKinnon CD, Winter DA. Control of whole-body balance in the frontal plane during human walking. *J Biomech*. 1993;26:633-644.
40. Eng JJ, Winter DA. Kinetic analysis of the lower limbs during walking: what information can be gained from a three-dimensional model? *J Biomech*. 1995;28:753-758.
41. Hof AL, Gazendam MGJ, Sinke WE. The condition for dynamic stability. *J Biomech*. 2005;38:1-8.
42. Carpenter R, Reddi B. *The Control of Posture*. 5th ed. Neurophysiology. Boca Raton, FL: CRC Press; 2012:215-224.
43. Niino N, Tsuzuku S, Ando F, Shimokata H. Frequencies and circumstances of falls in the National Institute for Longevity Sciences, Longitudinal Study of Aging (NILS-LSA). *J Epidemiol*. 2000;10:S90-S94.
44. Wolf-Klein GP, Silverstone FA, Basavarajo N, Foley CJ, Pascara A, Ma PH. Prevention of falls in the elderly population. *Arch Phys Med Rehabil*. 1988;69:689-691.
45. Morfitt JM. Falls in old people at home: Intrinsic versus environmental factors in causation. *Public Health*. 1983;97:115-120.
46. Clark AN. Factors in fracture of the female femur. A clinical study of the environmental, physical, medical and preventive aspects of this injury. *Gerontol Clin (Basel)*. 1968;10:257-270.
47. Campbell AJ, Reinken J, Allan BC, Martinez GS. Falls in old age: a study of frequency and related clinical factors. *Age Ageing*. 1981;10:264-270.
48. Grabiner MD, Bareither ML, Gatts S, Marone J, Troy KL. Task-specific training reduces trip-related fall risk in women. *Med Sci Sports Exerc*. 2012;44:2410-2414.
49. Oddsson LI, Karlsson R, Konrad J, Ince S, Williams SR, Zemkova E. A rehabilitation tool for functional balance using altered gravity and virtual reality. *J Neuroeng Rehabil*. 2007;4:25-27.
50. Giboin L-S, Gruber M, Kramer A. Task-specificity of balance training. *Hum Mov Sci*. 2015;44:22-31.
51. Kaufman KR, Wyatt MP, Sessoms PH, Grabiner MD. Task-specific fall prevention training is effective for warfighters with transtibial amputations. *Clin Orthop Relat Res*. 2014;472:3076-3084.
52. Blake AJ, Morgan K, Bendall MJ, et al. Falls by elderly people at home: prevalence and associated factors. *Age Ageing*. 1988;17:365-372.
53. 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;48:80-85.
54. Legro MW, Reiber G, del Aguila M, et al. Issues of importance reported by persons with lower limb amputations and prostheses. *J Rehabil Res Dev*. 1999;36:155-163.
55. Bellmann M, Schmalz T, Blumentritt S. Comparative biomechanical analysis of current microprocessor-controlled prosthetic knee joints. *Arch Phys Med Rehabil*. 2010;91:644-652.
56. Nickens H. Intrinsic factors in falling among the elderly. *Arch Intern Med*. 1985;145:1089-1093.
57. Quai TM, Brauer SG, Nitz JC. Somatosensation, circulation and stance balance in elderly dysvascular transtibial amputees. *Clin Rehabil*. 2005;19:668-676.
58. Richardson JK, Eckner JT, Allet L, Kim H, Ashton-Miller JA. Complex and simple clinical reaction times are associated with gait, balance, and major fall injury in older subjects with diabetic peripheral neuropathy. *Am J Phys Med Rehabil*. 2017;96:8-16.
59. Steinberg N, Gottlieb A, Siev-Ner I, Plotnik M. Fall incidence and associated risk factors among people with a lower limb amputation during various stages of recovery—A systematic review. *Disabil Rehabil*. 2018;16:1-10.
60. 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;9:170-180.e1.
61. Winter DA. *Biomechanics and Motor Control of Human Movement*. 4th ed. Hoboken, NJ: John Wiley & Sons; 2009.
62. Viton JM, Mouchino L, Mille ML, et al. Equilibrium and movement control strategies in trans-tibial amputees. *Prosthet Orthot Int*. 2000;24:108-116.
63. Paysant J, Beyaert C, Dati A-M, Martinet N, Andr J-M. Influence of terrain on metabolic and temporal gait characteristics of unilateral transtibial amputees. *J Rehabil Res Dev*. 2006;43:153-110.
64. Kendall C, Lemaire ED, Dudek NL, Kofman J. Indicators of dynamic stability in transtibial prosthesis users. *Gait Posture*. 2010;31:375-379.
65. Gates DH, Dingwell JB, Scott SJ, Sinitski EH, Wilken JM. Gait characteristics of individuals with transtibial amputations walking on a destabilizing rock surface. *Gait Posture*. 2012;36:33-39.
66. Gates DH, Scott SJ, Wilken JM, Dingwell JB. Frontal plane dynamic margins of stability in individuals with and without transtibial amputation walking on a loose rock surface. *Gait Posture*. 2013;38:570-575.
67. Mackenzie L, Byles J. Validation of self-reported fall events in intervention studies. *Clin Rehabil*. 2006;20:331-339.
68. Jenkins P, Earle-Richardson G, Slingerland DT, May J. Time dependent memory decay. *Am J Ind Med*. 2002;41:98-101.
69. 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.
70. Ziegler-Graham K, MacKenzie EJ, Ephraim PL, Trivison TG, Brookmeyer R. Estimating the prevalence of limb loss in the United States: 2005 to 2050. *Arch Phys Med Rehabil*. 2008;89:422-429.

---

*This journal-based CME activity is designated for 1.0 AMA PRA Category 1 Credit and can be completed online at <http://www.me.aapmr.org>. This activity is FREE to AAPM&R members and available to nonmembers for a nominal fee. For assistance with claiming CME for this activity, please contact (847) 737-6000.*

---

---

## Disclosure

**J.K.** Department of Kinesiology and Nutrition, University of Illinois at Chicago, Chicago, IL  
Disclosure: nothing to disclose

**M.J.M.** Feinberg School of Medicine, Northwestern University and Jesse Brown VA Medical Center, Chicago, IL  
Disclosure: nothing to disclose

**B.H.** Department of Rehabilitation Medicine, University of Washington, Seattle, WA  
Disclosure: nothing to disclose

**A.S.** Department of Kinesiology and Nutrition, University of Illinois at Chicago, 1919 W Taylor Street, Chicago, IL, IL 60612. Address correspondence to: A.S.; e-mail: [asawers@uic.edu](mailto:asawers@uic.edu)  
Disclosure: nothing to disclose

Disclosure related to this publication: Award W81XWH-17-1-0547, Department of

Defense (J.K., B.H., A.S.); award K12HD073945, National Institutes of Health (A.S.); grant OPERF-SGA-2016-1, Orthotics and Prosthetics Education and Research Foundation, Inc (A.S., B.H.); award IK2 RX001322, U.S. Department of Veterans Affairs (M.J.M.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Department of Defense, National Institutes of Health, Orthotics and Prosthetics Education and Research Foundation, or US Department of Veterans Affairs.

All financial disclosures and CME information related to this article can be found on the me<sup>®</sup> site (<http://me.aapmr.org/>) prior to accessing the activity.

Submitted for publication June 14, 2018; accepted August 25, 2018

---

## CME Question

According to this report, the most common combination of factors leading to falls in unilateral lower limb prosthetic users is:

- a. Center of mass destabilization due to intrinsic factors
- b. Base of support destabilization due to intrinsic factors
- c. Base of support destabilization due to extrinsic factors
- d. Center of mass destabilization due to extrinsic factors

Answer online at <http://me.aapmr.org>

# Conventional administration and scoring procedures suppress the diagnostic accuracy of a performance-based test designed to assess balance ability in lower limb prosthesis users

*Prosthetics and Orthotics International*

2019, Vol. 43(4) 402–408

© The International Society for

Prosthetics and Orthotics 2019

Article reuse guidelines:

[sagepub.com/journals-permissions](http://sagepub.com/journals-permissions)

DOI: 10.1177/0309364619848274

[journals.sagepub.com/home/poi](http://journals.sagepub.com/home/poi)Andrew Sawers<sup>1</sup>  and Brian J Hafner<sup>2</sup> 

## Abstract

**Background:** Practice effects have been observed among performance-based clinical tests administered to prosthesis-users. Their impact on test applications remains unknown.

**Objective:** To determine whether scoring a clinical balance test using conventional procedures that do not accommodate practice effects reduces its diagnostic accuracy relative to scoring it using recommended procedures that do accommodate practice effects.

**Study Design:** Cross-sectional study.

**Methods:** Narrowing Beam Walking Test data from 40 prosthesis users was scored using recommended methods (i.e. average of trials 3–5) and conventional methods applied to other tests (i.e. mean or best of trials 1–3). Area under the receiver operating characteristic curve for each method was compared to 0.50, to determine if it was better than chance at identifying prosthesis-users with a history of falls, and to 0.80, to determine if it surpassed a threshold recommended for diagnostic accuracy.

**Results:** Receiver operating characteristic curve area decreased when the Narrowing Beam Walking Test was scored using conventional rather than recommended procedures. Furthermore, when scored using conventional procedures, the NBWT no longer discriminated between prosthesis-users with and without a history of falls with a probability greater than chance, or exceeded recommended diagnostic thresholds.

**Conclusion:** Scoring the Narrowing Beam Walking Test using conventional procedures that do not accommodate practice effects decreased its diagnostic accuracy among prosthesis-users relative to recommended procedures. Conventional scoring procedures may limit the effectiveness of performance-based tests used to screen for fall risk in prosthesis-users because they do not mitigate practice effects. The influence of practice effects on other tests, and test applications (e.g. clinical evaluation and prediction), is warranted.

## Clinical relevance

Scoring a clinical balance test using conventional procedures that do not mitigate practice effects reduced its diagnostic accuracy. Changing administration and scoring procedures to accommodate practice effects should be considered to improve the diagnostic accuracy of other performance-based balance tests.

## Keywords

Amputee, accidental falls, psychometrics, rehabilitation

Date received: 25 January 2019; accepted: 12 April 2019

<sup>1</sup>Department of Kinesiology, University of Illinois at Chicago, Chicago, IL, USA

<sup>2</sup>Department of Rehabilitation Medicine, University of Washington, Seattle, WA, USA

## Corresponding author:

Andrew Sawers, Department of Kinesiology, University of Illinois at Chicago, 1919 West Taylor Street, Rm. 646, Chicago, IL 60612, USA.  
Email: [asawers@uic.edu](mailto:asawers@uic.edu)

## Background

Over 50% of lower limb prosthesis (LLP) users report falling at least once a year,<sup>1–3</sup> placing them at a high risk for adverse health outcomes that include injury,<sup>2,4</sup> reduced mobility,<sup>5</sup> and diminished quality of life.<sup>6,7</sup> Falls are therefore a considerable problem that negatively affect the lives of a substantial portion of LLP users. A major barrier to reducing falls among LLP users has been effective screening of those at risk.<sup>8</sup> Existing performance-based clinical balance tests do not generally pose sufficient challenge to identify differences in balance control that underlie fall risk,<sup>9</sup> reducing their ability to discriminate between LLP users with and without a history of falls.<sup>8</sup>

A standardized performance-based clinical balance test termed the Narrowing Beam Walking Test (NBWT) was developed to address this limitation among LLP users.<sup>10,11</sup> The NBWT was designed to safely challenge and detect differences in balance across a wide range of abilities by restricting the base-of-support and/or reducing the width of the support surface.<sup>8,11</sup> The NBWT requires tested individuals walk along four low beams (2.5 cm height, 1.8 m long), each narrower than the last (20 cm, 10 cm, 4.0 cm, and 2.0 cm). Validation testing among LLP users demonstrated that the NBWT was the only one of several clinical balance tests to discriminate between LLP users with and without a history of falls with a probability greater than chance, and above a recommended threshold for diagnostic accuracy.<sup>8</sup> The accuracy with which the NBWT discriminated between LLP users with and without a history of falls was attributed to the increased challenge beam walking imposes on balance control.<sup>8,12</sup>

Beyond increasing the challenge to balance control, development of the NBWT included the study practice effects associated with beam-walking performance by LLP users.<sup>9,11</sup> Practice effects are systematic changes in performance that occur between consecutive trials or repeated assessments with the same test in the absence of any intervention.<sup>13,14</sup> Practice effects need to be accommodated (e.g. by allowing tested individuals to perform several practice trials) to avoid taking measurements during the time when individuals are improving on the test. Although practice effects have been observed in performance-based tests among LLP users and other patient populations,<sup>15–20</sup> few clinical test developers explicitly state whether their recommended administration and scoring procedures address possible practice effects,<sup>21</sup> and none have examined the role practice effects may have on the discriminative, evaluative, or predictive applications of performance-based clinical tests.

During development of the NBWT, investigators used data from 10 consecutive trials to evaluate practice effects in LLP users, and create an administration and scoring procedure (i.e. the mean of trials 3 through 5) that accurately estimated LLP users' consistent (i.e. typical) performance.<sup>11</sup> The developers, however, did not compare the diagnostic accuracy of NBWT scores obtained with their

recommended administration and scoring procedure to scores obtained using conventional methods for scoring clinical balance tests<sup>11</sup> (e.g. mean or best of trials 1–3).<sup>22–25</sup> It remains unknown therefore whether the superior diagnostic accuracy of the NBWT among LLP users<sup>8</sup> is due to the challenge the NBWT imposes on balance control, as originally hypothesized, or use of an administration and scoring procedure that mitigates practice effects.

The primary objective of this study was to determine whether scoring the NBWT using methods generally recommended for other clinical balance tests (i.e. mean or best of trials 1–3) would affect its ability to discriminate LLP fallers and non-fallers. It was hypothesized that scoring the NBWT using conventional scoring methods would adversely affect its diagnostic accuracy (i.e. reduce the area under the receiver operating characteristic (ROC) curve). To test this hypothesis, the diagnostic accuracy of NBWT scores derived using three different administration and scoring procedures—mean of trials 3–5, mean of trials 1–3, and best of trials 1–3 were compared. Significant differences in diagnostic accuracy would indicate that practice effects limit the effectiveness of the NBWT, and that additional research examining whether practice effects influence the diagnostic accuracy of similar performance-based balance tests may be required. A secondary objective was to determine whether factors previously associated with practice effects in performance-based clinical tests (e.g. baseline performance, age, level of motor deficit)<sup>15,26,27</sup> could explain practice effects observed among LLP users performing the NBWT.

## Methods

### Study design

A secondary analysis of cross-sectional data from a sample of 40 LLP users collected in a previous study<sup>7</sup> was conducted to address the study objectives. Original study protocols were reviewed and approved by institutional review boards at the University of Illinois at Chicago. All individuals provided written informed consent prior to participation.

### Participants

Inclusion criteria for the prior study<sup>11</sup> were as follows: 18 years of age or older; unilateral transtibial or transfemoral amputation due to trauma, dysvascular complications, tumor, or infection; one year or more of experience using a prosthesis, regular use of a prosthesis to ambulate; and ability to walk 10 feet over level terrain with the prosthesis and without an assistive device. Exclusion criteria included amputation of a second limb, complications on the contralateral leg (e.g. knee replacement, ulcers), severe pulmonary disease, an advanced neurologic disorder, or congenital limb absence or difference. People with congenital limb absence or difference were excluded because their motor

**Table 1.** Sociodemographic and prosthetic-related characteristics.

|           | Age (years) | Height (cm) | Weight (kg) | Years Since Limb Loss | Sex    | Amputation Level | Amputation Etiology                            | MFCL                         |
|-----------|-------------|-------------|-------------|-----------------------|--------|------------------|------------------------------------------------|------------------------------|
| Mean (SD) | 48.7 (14.6) | 173 (9.10)  | 78.5 (14.1) | 14.3 (12.6)           | M (21) | TT (25)          | Trauma (25)                                    | K1 (2)                       |
| Range     | 24–70       | 155–189     | 47–107      | 2–57                  | F (19) | TF (15)          | Dysvascular (7)<br>Cancer (5)<br>Infection (3) | K2 (15)<br>K3 (17)<br>K4 (6) |

MFCL: Medicare Functional Classification Level (K-level); M: male; F: female; TT: transtibial; TF: transfemoral.

control, and potential balance, may be fundamentally different than someone who loses a limb later in life.

## Measurements

**Participant demographic and characterization measures.** Sociodemographic information (i.e. height, weight, age, and sex) was collected 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. amputation level, amputation etiology, and time since amputation) was obtained via interview with a study investigator. To identify participants with a history of recent falls, each was asked, “In the past year have you experienced any falls including a slip or trip where you inadvertently lost your balance and landed on the ground or lower level?”<sup>28–30</sup> Participants reporting one or more falls in the past 12 months were classified as “fallers.”<sup>8,31–33</sup>

**Assessment of practice effects in narrowing beam walking.** Participants’ performance across 10 narrowing beam-walking trials was examined in a prior study in order to develop administration and scoring procedures for the NBWT.<sup>11</sup> Performance on each trial was quantified by the distance walked prior to the participant stepping off the beam or uncrossing their arms, whichever came first. Practice effects were identified as statistically significant changes in the slope of each participant’s cumulative trial-by-trial performance.<sup>11,34</sup> Participants’ narrowing beam-walking performance was considered stable and consistent once the cumulative trial-by-trial record reached a terminal slope (i.e. the participant exhibited no further changes in performance across trials).<sup>11</sup> Participants were classified into one of two groups; LLP users who did and did not experience significant changes in performance over the 10 beam walking trials (i.e. participants who did or did not exhibit practice effects). The distance walked along the NBWT on the first trial was selected as a measure of baseline performance.

An administration and scoring procedure was then developed to mitigate the observed narrowing beam-walking practice effects,<sup>11</sup> and ensure that the NBWT assessed balance ability based on an accurate estimate of tested individuals’ stable and consistent performance. Specifically, the

mean of trials 3–5 was found to more accurately estimate stable consistent narrowing beam-walking performance than methods used by most contemporary clinical balance tests (i.e. best or mean of trials 1–3).<sup>11</sup> The former method was therefore recommended for administering and scoring the NBWT (11).

## Statistical analysis

Distributions of NBWT scores derived using each scoring method were evaluated for normality using the Shapiro-Wilk test. A one-way repeated measures analysis of variance (ANOVA) was used to compare NBWT scores across the three scoring methods (i.e. mean of trials 3–5, mean of trials 1–3, best of trials 1–3). Diagnostic accuracy of each method was evaluated using the area under the respective receiver operating characteristic (ROC) curve.<sup>35</sup> Areas were first compared to a null value of 0.50 to assess whether the procedure was better than chance at identifying participants with a history of falls. Areas were then compared to a threshold of 0.80, a recommended threshold for clinically acceptable diagnostic accuracy.<sup>36</sup> The significance level was adjusted with a Bonferroni correction ( $\alpha < .0167$ ) to account for multiple comparisons. To test whether factors previously linked to practice effects (i.e. age, baseline performance, motor deficits)<sup>15,27</sup> differed between LLP users who did and did not exhibit practice effects on the NBWT, continuous variables (e.g. age, baseline performance) were compared between groups using 2-sided t-tests, while nominal variables (e.g. amputation level, fall history) were compared using Fisher’s exact tests. Statistical analyses were performed using SPSS v.24 (Chicago, IL).

## Results

Narrowing beam-walking data from 40 LLP users (Table 1) who participated in the original study<sup>8</sup> were analyzed. Practice effects, lasting between one and seven trials, were observed among 17 participants. NBWT scores (mean  $\pm$  95% confidence interval (CI)) for the mean of trials 3–5 ( $0.40 \pm 0.13$ ), the best of trials 1–3 ( $0.44 \pm 0.15$ ), and the mean of trials 1–3 ( $0.36 \pm 0.14$ ) were all normally distributed (mean 3–5:  $W=0.964$ ,  $p=0.24$ ; best 1–3:  $W=0.973$ ,  $p=0.45$ ; mean 1–3:  $W=0.959$ ,  $p=0.21$ ). Mauchly’s test of sphericity indicated that the assumption of sphericity was violated,  $\chi^2(2)=6.508$ ,

**Table 2.** Participant characteristics potentially contributing to practice effects.

|                              | Age (years)   | NBWT baseline first trial | MFCL* |      | Amputation level* |    | Amputation etiology* |     | Fall history* |         |
|------------------------------|---------------|---------------------------|-------|------|-------------------|----|----------------------|-----|---------------|---------|
|                              | Mean (SD)     | Mean (SD)                 | K1&2  | K3&4 | TT                | TF | DV                   | NDV | Fall          | No Fall |
| Practice effects (n = 17)    | 49.35 (12.52) | 0.405 (0.232)             | 7     | 10   | 13                | 4  | 1                    | 16  | 8             | 9       |
| No practice effects (n = 23) | 48.26 (16.17) | 0.290 (0.209)             | 10    | 13   | 12                | 11 | 7                    | 16  | 16            | 7       |
| p-value                      | 0.818         | 0.109                     | 1.00  |      | 0.187             |    | 0.107                |     | 0.199         |         |

MFCL: Medicare Functional Classification Level (K-level); TT: transtibial; TF: transfemoral; DV: dysvascular; NDV: non-dysvascular.

\*Statistical significance was assessed with a Fisher's Exact Test.

$p=0.039$ . Using the Greenhouse-Geisser correction, NBWT scores were found to be significantly different between scoring methods,  $F(1.73, 67.39)=16.87$ ,  $p<0.001$ . After a Bonferroni correction, the mean of trials 3–5 was significantly lower than the best of trials 1–3,  $p=0.041$ , and significantly greater than the mean of trials 1–3,  $p=0.017$ . The best of trials 1–3 was also significantly greater than the mean of trials 1–3,  $p<0.001$ .

The area under the ROC curve (area, 95% CI) was larger for the mean of trials 3–5 (0.81, 0.62–0.91) than for the best of trials 1–3 (0.67, 0.50–0.84) or the mean of trials 1–3 (0.69, 0.52–0.85). The area under the ROC curve was significantly greater than 0.50 for the mean of trials 3–5 ( $p=0.006$ ) and the mean of trials 1–3 ( $p=0.041$ ), but not for the best of trials 1–3 ( $p=0.66$ ). Only the mean of trials 3–5 had an area under the ROC curve that exceeded 0.80 (Supplemental Figure 1). No significant differences ( $p<0.05$ ) in factors previously reported to influence practice effects were identified between LLP users who did and did not exhibit practice effects on the NBWT (Table 2). Similarly, LLP users who exhibited practice effects had a greater NBWT score (i.e. mean of trials 3–5: 0.458) than those who did not (i.e. 0.361); however, this difference was not statistically significant ( $p=0.118$ ).

## Discussion

This study was conducted to test the hypothesis that scoring the NBWT using conventional methods (i.e. mean or best of trials 1–3) would adversely affect its diagnostic accuracy. Results show that the diagnostic accuracy of the NBWT was reduced significantly when scored using methods that do not accommodate practice effects. When the NBWT was scored using conventional methods, it was no longer able to discriminate between fallers and non-fallers with a probability greater than chance (i.e. the best of trials 1–3), or exceed an established threshold for diagnostic accuracy (i.e. mean and best of trials 1–3). These results indicate that accommodating practice effects, rather than the challenge posed by NBWT, may be responsible for its superior diagnostic accuracy. Furthermore, these results suggest that the influence of practice effects on the diagnostic ability of other performance-based clinical tests should be investigated. This may have important implications for the use of perfor-

mance-based clinical tests in clinical trials, as well as day-to-day clinical practice.

*Practice effects influence the diagnostic accuracy of the NBWT.* Several studies have documented practice effects in performance-based clinical tests.<sup>15–20</sup> We are not aware, however, of any research that has assessed the influence of practice effects on the discriminative, evaluative, or predictive ability of performance-based clinical tests. The results of this study therefore contribute several key results to the body of knowledge. First, the diagnostic accuracy of the NBWT dropped to suboptimal levels reported for other clinical balance tests applied among LLP users<sup>8</sup> when it was scored in a manner used by other contemporary balance tests (i.e. best of trials 1 and 2).<sup>22–25</sup> Second, the NBWT's ability to identify people with a history of falls dropped below a recommended threshold of accuracy<sup>36</sup> when it was scored in this manner. Third, the NBWT was unable to identify LLP users with a history of falls better than chance when it was scored using the best of trials 1–3. In contrast, when the NBWT is scored using its recommended methods that mitigate practice effects (i.e. mean of trials 3–5), its diagnostic accuracy was both significantly better than chance and surpassed the recommended threshold for diagnostic accuracy.<sup>36</sup>

The current results, while focused on one performance-based balance test, represent an important, albeit initial, step in understanding how practice effects can limit the effectiveness of clinical tests. Presence of practice effects in performance-based tests indicates that an individual's performance is not consistent, but is instead changing over repeated administrations (i.e. due to familiarity with the test or practice). The consequence of practice effects, whether in clinical practice or research, is that an individual's initial baseline measurements may not represent their true ability. Correspondingly, changes in measured performance (e.g. after an intervention or over time) may not be accurate. Thus, tests that exhibit practice effects may give erroneous information about whether or how the individual has changed. Examples of tests that may exhibit practice effects in LLP users have begun to appear in Halsne et al.<sup>16</sup>

Whereas previous studies have advocated increasing the challenge to balance control as a way to improve the diagnostic accuracy of clinical tests,<sup>9,37</sup> these results suggest that mitigating practice effects may also facilitate discrimination

between LLP users with and without a history of falls. Had increasing the challenge to balance control alone been sufficient, the NBWT scored using the best or mean of trials 1–3 would have had an ROC curve area greater than 0.80. Rather, it appears that it is necessary to mitigate practice effects, in this case via administration and scoring procedures, in order to achieve an acceptable level of diagnostic accuracy. The ability of the NBWT to discriminate between LLP users with and without a history of falls with greater accuracy than existing clinical balance tests<sup>8</sup> may therefore be attributed not only to increasing the challenge to balance control,<sup>9,10</sup> but also to mitigating practice effects through administration of additional trials.<sup>11</sup> Interestingly, practice effects associated with narrowing beam walking appear to affect either an overestimation (i.e. the best of trials 1–3), or underestimation (i.e. the mean of trials 1–3) of individuals' consistent NBWT performance among LLP users. Controlling or reducing the variance attributable to practice effects appears to be a critically important consideration when using performance-based clinical balance tests like the NBWT to evaluate fall risk among LLP users. Whether similar results would be observed with other performance-based clinical balance tests remains to be determined.

*The cause of practice effects may dictate how they ought to be addressed and potentially interpreted.* Practice effects in performance-based clinical tests have been associated with individuals who are younger,<sup>15,27</sup> exhibit fewer motor deficits,<sup>15,27</sup> demonstrate better baseline performance,<sup>26</sup> and receive greater verbal encouragement.<sup>15</sup> In this study, practice effects on the NBWT could not be explained by differences in age, baseline performance, mobility, or level of motor deficit (i.e. amputation level, etiology) (Table 2). Thus, it would not be feasible to address practice effects by adjusting NBWT scores based on these factors. Rather, developing appropriate administration and scoring procedures would appear to be the most suitable approach to resolving practice effects.

While practice effects in a few performance-based clinical tests have been related to a number of factors described above,<sup>15,26,27</sup> the reasons for practice effects on a given test remain largely unknown. Considering potential explanations for practice effects in performance-based clinical tests may have important implications for how they are identified, interpreted, and best addressed. For example, if practice effects are related to issues of motivation, effort, or movement strategy selection, they may be best addressed by modifying test instructions. If specific motor tasks were found to be susceptible to practice effects, then they could be removed from existing clinical tests and avoided when developing new tests. Each of these solutions, however, assumes an understanding of practice effects and their causes that has yet to be established.

While the focus of this study was on resolving practice effects, practice effects themselves may provide important and useful information about the integrity and adaptability

of the systems mediating an individual's performance. For example, LLP users who exhibited practice effects in the current study walked further on their first NBWT trial, reported fewer falls, and had a greater NBWT score (i.e. mean of trials 3–5) (Table 2). While these results did not reach statistical significance, they do suggest the possibility that LLP users who can adapt their gait, measured here through practice effects, may be safer ambulators. Additional research is required to examine this hypothesis in a larger sample. Depending on the setting therefore, practice effects may be conceptualized either as a source of unwanted variance needing to be removed, or as another factor to be considered in clinical decisions.<sup>38</sup>

*Study limitations.* The current study only assessed the influence of practice effects on one performance-based clinical test (i.e. the NBWT), one application (i.e. discrimination), one construct (i.e. balance), and one clinical population (i.e. LLP users). Whether similar results would be observed with other balance tests (e.g. timed performance tests, clinical rating scales), applications (i.e. prediction and evaluation), constructs (i.e. mobility and endurance), and/or clinical populations (e.g. incomplete spinal cord injury, stroke, older adults) remains to be determined. Practice effects observed in other performance-based clinical tests,<sup>15,17,18</sup> including those applied to LLP users,<sup>16,19,20</sup> suggest that the current results extend beyond this study. The sample size used in the present analysis ( $n=40$ ), while consistent with similar studies,<sup>23,39</sup> should be increased in future work to confirm the current results and add greater certainty to their interpretation. Finally, only a limited set of variables was studied in an effort to explain why practice effects may be occurring. Studied variables were limited to participant demographics and characterization measures collected in the original study.<sup>8</sup> The collection of physiological measures of mood, effort, attention, cognition, and movement strategies in future studies may yield further insights into the basis of practice effects among performance-based clinical tests.

## Conclusion

The results of the present study suggest that the information obtained from performance-based clinical tests may be limited by how they are administered and scored. Rescoring the NBWT using methods often used with other clinical tests significantly decreased its diagnostic accuracy. The ability of the NBWT to discriminate between LLP users with and without a history of falls with greater accuracy than existing clinical balance tests can therefore be attributed not only to having imposed greater challenge to balance control, but also to mitigating practice effects. In light of their prominent use in clinical trials and clinical practice, results of this study suggest that consideration should be given to whether the discriminative, predictive, or evaluative applications of other performance-based clinical tests are influenced by practice effects.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: Research reported in this publication was supported by the National Institutes of Health (NIH) under award number K12HD073945, by the Orthotics and Prosthetics Education and Research Foundation, Inc. (OPERF) under grant number OPERF-SGA-2016-1, and by the Department of Defense (DoD) under award number W81XWH-17-1-0547. This content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, OPERF, or DoD.

## Supplemental material

Supplemental material for this article is available online.

## ORCID iDs

Andrew Sawers  <https://orcid.org/0000-0002-3493-304X>

Brian J Hafner  <https://orcid.org/0000-0001-6175-1869>

## References

- Kim J, Hafner BJ, Major MJ, et al. Frequency and circumstances of falls reported by ambulatory unilateral lower limb prosthesis users: a secondary analysis. *PM R* 2019; 11: 344–353.
- Kulkarni J, Wright S, Toole C, et al. Falls in patients with lower limb amputations: prevalence and contributing factors. *Physiotherapy* 1996; 82: 130–136.
- Miller WC, Speechley M and 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.
- Mundell B, MaraditKremers H, Visscher S, et al. Direct medical costs of accidental falls for adults with transfemoral amputations. *Prosthet Orthot Int* 2017; 41(6): 564–570.
- van Velzen JM, van Bennekom CA, Polonski W, et al. Physical capacity and walking ability after lower limb amputation: a systematic review. *Clin Rehabil* 2006; 20(11): 999–1016.
- Pezzin LE, Dillingham TR and MacKenzie EJ. Rehabilitation and the long-term outcomes of persons with trauma-related amputations. *Arch Phys Med Rehabil* 2000; 81(3): 292–300.
- Barnett CT, Vanicek N and Polman RCJ. Temporal adaptations in generic and population-specific quality of life and falls efficacy in men with recent lower-limb amputations. *J Rehabil Res Dev* 2013; 50(3): 437–48.
- Sawers A and Hafner B. Validation of the narrowing beam walking test in lower limb prosthesis users. *Arch Phys Med Rehabil* 2018; 99(8): 1491–1498.e1.
- Sawers A and Ting LH. Beam walking can detect differences in walking balance proficiency across a range of sensorimotor abilities. *Gait Posture* 2015; 41(2): 619–623.
- Sawers A and Hafner BJ. A study to assess whether fixed-width beam walking provides sufficient challenge to assess balance ability across lower limb prosthesis users. *Clin Rehabil* 2018; 32(4): 483–492.
- Sawers A and Hafner BJ. Narrowing beam walking is a clinically feasible approach for assessing balance ability in lower-limb prosthesis users. *J Rehabil Med* 2018; 50(5): 457–464.
- Sawers A, Allen JL and Ting LH. Long-term training modifies the modular structure and organization of walking balance control. *J Neurophysiol* 2015; 114(6): 3359–3373.
- Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med* 2000; 30(1): 1–15.
- Bartels C, Wegrzyn M, Wiedl A, et al. Practice effects in healthy adults: a longitudinal study on frequent repetitive cognitive testing. *BMC Neurosci* 2010; 11: 118–112.
- Guyatt GH, Pugsley SO, Sullivan MJ, et al. Effect of encouragement on walking test performance. *Thorax* 1984; 39(11): 818–822.
- Halsne EG, McDonald CL, Morgan SJ, et al. Assessment of low- and high-level task performance in people with transtibial amputation using crossover and energy-storing prosthetic feet: a pilot study. *Prosthet Orthot Int* 2018; 42: 583–591.
- Kristensen MT, Ekdahl C, Kehlet H, et al. How many trials are needed to achieve performance stability of the timed up & Go test in patients with hip fracture? *Arch Phys Med Rehabil* 2010; 91: 885–889.
- Bloch ML, Jönsson LR and Kristensen MT. Introducing a third timed up & go test trial improves performances of hospitalized and community-dwelling older individuals. *J Geriatr Phys Ther* 2007; 40: 121–126.
- Lin SJ and Bose NH. Six-minute walk test in persons with transtibial amputation. *Arch Phys Med Rehabil* 2008; 89(12): 2354–2359.
- Kristensen MT, Nielsen AØ, Topp UM, et al. Number of test trials needed for performance and interrater reliability of the one leg stand test in patients with a major non-traumatic lower limb amputation. *Gait Posture* 2014; 39: 424–429.
- Ashendorf L and Horwitz JE. Abbreviating the finger tapping test. *Arch Clin Neuropsychol* 2015; 30(2): 99–104.
- Podsiadlo D and Richardson S. The timed “up & go”: a test of basic functional mobility for frail elderly persons. *JAGS* 1991; 39(2): 142–148.
- Dite W, Connor HJ and Curtis HC. Clinical identification of multiple fall risk early after unilateral transtibial amputation. *Arch Phys Med Rehabil* 2007; 88(1): 109–114.
- Deathe AB and Miller WE. The L test of functional mobility: measurement properties of a modified version of the timed “up & go” test designed for people with lower-limb amputations. *Phys Ther* 2005; 85(7): 625–635.
- Wolfson L, Whipple R, Derby C, et al. Balance and strength training in older adults: intervention gains and Tai Chi maintenance. *J Am Geriatr Soc* 1996; 44(5): 498–506.
- Adsett J, Mullins R, Hwang R, et al. Repeat six-minute walk tests in patients with chronic heart failure: are they clinically necessary. *Eur J Cardiovasc Prev Rehabil* 2011; 18(4): 601–606.

27. Liu J, Drutz C, Kumar R, et al. Use of the six-minute walk test post-stroke: is there a practice effect? *Arch Phys Med Rehabil* 2008; 89: 1686–1692.
28. Askham J, Glucksman E, Owens P, et al. *Home and leisure accident research: a review of research on falls among elderly people*. London: Age Concern Institute of Gerontology, King's College London, and Department of Trade and Industry, 1990.
29. Hauer K, Lamb SE, Jorstad EC, et al. Systematic review of definitions and methods of measuring falls in randomized controlled fall prevention trials. *Age and Ageing* 2006; 35(1): 5–10.
30. Lamb SE, Jorstad-Stein EC, Hauer K, et al. 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(9): 1618–1622.
31. Wong C, Chihuri S and 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.
32. Topper AK, Maki BE and Holliday PJ. Are activity-based assessments of balance and gait in the elderly predictive of risk of falling and/or type of fall? *JAGS* 1993; 141(5): 479–487.
33. Berg WP, Alessio HM, Mills EM, et al. Circumstances and consequences of falls in independent community-dwelling older adults. *Age Ageing* 1997; 26(4): 261–268.
34. Gallistel CR, Fairhurst S and Balsam P. The learning curve: implications of a quantitative analysis. *Proc Natl Acad Sci U S A* 2004; 101(36): 13124–13131.
35. Hanley JA and McNeil BJ. A method of comparing the areas under Receiver Operating Characteristic curves derived from the same cases. *Radiology* 1983; 148(3): 839–843.
36. Haynes RB. *Clinical epidemiology: how we do clinical practice research*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins, 2006.
37. Bandinelli S, Pozzi M, Lauretani F, et al. Adding challenge to performance-based tests of walking. *Am J Phys Med Rehabil* 2006; 85(12): 986–991.
38. McCaffrey RJ and Westervelt HJ. Issues associated with repeated neuropsychological assessments. *Neuropsychol Rev* 1995; 5(3): 203–221.
39. Jorgensen V, Opheim A, Halvarsson A, et al. 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.



## Original Research—CME

# 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 , Brian J. Hafner, PhD

---

## Abstract

**Background:** Clinicians are routinely required to make decisions about fall risk among lower limb prosthesis (LLP) users. 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 nonfallers 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 (ie, 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 ( $\leq .43/1.0$ ), TUG ( $\geq 8.17$  seconds), FSST ( $\geq 8.49$  seconds), BBS ( $\leq 50.5/56$ ), and ABC ( $\leq 80.2/100$ ). Validity indices (ie, area under the 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.

**Level of Evidence:** III

---

## 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.<sup>1</sup> 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 continuous 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.<sup>2</sup> Validity indices (eg, 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.<sup>3,4</sup> Over 50% of LLP users report falling at least once a year, with up to 39% reporting multiple falls a year.<sup>4-8</sup> Falls among LLP users frequently lead to adverse health outcomes including injury,<sup>3,5,7,9,10</sup> financial costs,<sup>11</sup> reduced mobility,<sup>12,13</sup> and diminished quality of life.<sup>14</sup> A major barrier to reducing falls among LLP users has been effective screening of those at risk.<sup>15,16</sup> 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),<sup>17</sup> timed up and go (TUG),<sup>18</sup> four square step test (FSST),<sup>19</sup> and activities-specific balance confidence (ABC) scale.<sup>20</sup> A number of studies have assessed the validity and reliability of these instruments among LLP users,<sup>15,16,21-27</sup> but cutoff scores to infer fall risk have been established only for transtibial prosthesis users with limited prosthetic experience (ie, 6 months of prosthetic use).<sup>21</sup> Clinicians, therefore, lack the necessary information to assess fall risk among individual patients within the larger LLP user population (ie, established transtibial and transfemoral prosthesis users).

Two different faller classifications, either  $\geq 1$  fall<sup>3,16,28,29</sup> or  $\geq 2$  falls<sup>3,15,21</sup> 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 BBS, increased when a classification of  $\geq 2$  falls was used in a prior study.<sup>8</sup> Adopting a  $\geq 2$  falls classification inherently creates a “faller” group that is likely to possess worse balance ability than if it also included individuals with 1 fall (ie,  $\geq 1$  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.<sup>16,21</sup> Based on this prior research, therefore, it was hypothesized that the validity indices (eg, 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 nonfallers,<sup>3,15,16,21</sup> a secondary objective of this study was to determine whether cutoff scores and associated validity indices differed across two common fall classifications (ie,  $\geq 1$  fall vs  $\geq 2$  falls over the past 12 months). It was hypothesized that classifying fallers as  $\geq 2$  falls would increase validity indices.<sup>8</sup>

## 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<sup>30</sup> was followed during the collection and reporting of study data. All data were stored and managed using a REDCap database hosted at the University of Illinois at Chicago.<sup>31</sup> Study protocols were reviewed and approved by a University of Illinois at Chicago institutional review board. Study participants provided written informed consent prior to participation.

### Participants

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 (ie, established users); ability to ambulate at least 10 ft without an upper extremity assistive device (eg, cane); and use of a comfortable prosthesis (assessed with the Socket Comfort Scale).<sup>32</sup> Participants with complications to their contralateral leg (eg, 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.<sup>1,33-35</sup>

## Measurements

### Participant Demographic and Characterization Measures

Age, height, weight, and sex were collected from study participants via self-report. Medicare functional classification level (MFCL) (ie, K-level) was determined by a certified prosthetist via interview and physical evaluation, and amputation-related information (ie, level, etiology, and time since) was obtained via interview with a study investigator. Perceived mobility was assessed with the Prosthetic Limb Users Survey of Mobility (PLUS-M).<sup>36</sup> 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?"<sup>37-40</sup> 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 were analyzed using two faller classifications:  $\geq 1$  fall<sup>3,16,28,29</sup> and  $\geq 2$  falls<sup>3,15,21</sup> over the past 12 months.

### Clinical Balance Tests

Five clinical balance instruments, the ABC Scale,<sup>20</sup> the TUG,<sup>18</sup> the FSST,<sup>19</sup> the BBS,<sup>17</sup> and the NBWT<sup>16</sup> were administered and scored according to the developers' instructions (Appendix 1). Each balance instrument has demonstrated acceptable levels of validity<sup>15,16,21-23</sup> and/or reliability (ie, intraclass correlation coefficient range .70-.99)<sup>15,22,23,25-27,41</sup> among LLP users.

### Statistical Analysis

Clinical balance test scores were compared across participant subgroups defined by their reported fall history (ie, 0 falls, 1 fall, or  $\geq 2$  falls), and within commonly adopted fall classifications (ie, 0 falls vs  $\geq 1$  fall, or 0-1 falls vs  $\geq 2$  falls) using parametric (ie, analysis of variance) or nonparametric equivalents (Kruskal-Wallis Test).

Receiver operating characteristic (ROC) curves were used to identify cutoff scores<sup>33,42</sup> and related validity indices (ie, 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 nonfaller.<sup>43</sup> The larger the AUC, the greater the test's general discriminative ability.<sup>43</sup> The AUC was selected as a summary measure of diagnostic accuracy.<sup>44</sup> The AUC can assume any value between 0 and 1, with a value of 0.5 representing chance,<sup>2</sup> and values greater than or equal to 0.8 recommended as the limit of clinical acceptability.<sup>45</sup> Areas for each test were therefore compared to a threshold of 0.8 to determine the clinical

acceptability of a test. Although 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.<sup>46</sup> 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.<sup>46,47</sup> 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.<sup>2</sup> Values for sensitivity (ie, how often a diagnostic test is positive when a condition is present), and specificity (ie, 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.<sup>48</sup> Sensitivity, for example, does not describe how often patients with positive tests have the condition of interest (ie, a fall). Sensitivity indices, therefore, do not indicate the probability of the event or condition occurring,<sup>49</sup> information that is likely to be of greatest value to clinicians.<sup>1</sup> 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.<sup>50</sup> Likelihood ratios overcome these aforementioned limitations by quantifying how much the obtained test score increases or decreases the probability of an event occurring,<sup>1,35</sup> independent of its prevalence in the sample (ie, 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.<sup>34,35,51,52</sup> Unlike the AUC, likelihood ratios provide clinicians with actionable information for individual patients (ie, 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, whereas the likelihood ratio for a negative test (LR-) was computed as 1-specificity/sensitivity.<sup>49</sup> Likelihood ratios greater than 5 or less than 0.2 indicate moderate changes in the probability of an event (eg, fall), while likelihood ratios between 5 and 2, or 0.5 and 0.2 result in small, but potentially important shifts in probability.<sup>1</sup> 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 (ie,  $\geq 1$  fall = Model 1,  $\geq 2$  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

(eg, 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  $\geq 2$  or  $< 0.5$ , and a  $P$  value of  $< .05$  were retained for further consideration.<sup>53</sup> To avoid multicollinearity, any of the independent variables retained from the univariate logistic regression that had a strong correlation (ie, Spearman rho  $\geq .75$ ) and a lower odds ratio than the other independent variable with which it was correlated were excluded from the multivariate logistic regression model.<sup>53</sup> A multivariate logistic regression was then performed with all remaining independent variables included. Validity indices described previously were computed from the regression output. Cut-off scores and validity indices were computed for each of the faller classifications (ie,  $\geq 1$  fall and  $\geq 2$  falls). 95% confidence intervals (CI) were computed to determine the precision of each validity index.<sup>52</sup> All statistical analyses were conducted using SPSS v.25 (SPSS, Inc., Chicago, IL).

## Results

The sample included 40 established unilateral LLP users (height:  $173 \pm 9.10$  cm; weight:  $78.5 \pm 14.1$  kg; socket comfort score [SCS]:  $7.8 \pm 1.4$ ) (Table 1). All participants were able to perform and complete all tests. Scores for each of the studied tests, stratified by participant fall history and fall classification are presented in Tables 2 and 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 1 and 0 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 0 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  $\geq 1$  fall in the

12 months prior to assessment, the NBWT had the greatest area under the ROC curve (AUC) (0.81, 0.62-0.91), 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 (0.24, 0.13-0.56). The logistic regression model developed for the  $\geq 1$  fall classification (ie, 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 (eg, TUG, ABC, MFCL, amputation etiology, age and sex) failed to meet model inclusion criteria (ie, odds ratio  $\geq 2$  or less than 0.5, and  $P < .05$ ), during univariate logistic regression.

When fallers were defined as individuals reporting  $\geq 2$  falls in the 12 months prior to assessment validity indices (ie, 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, 0.78-1.0). The FSST, however, exhibited the largest sensitivity (94%, 83%-100%) and smallest negative likelihood ratio (0.08, 0.012-0.54), whereas 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  $\geq 2$  falls classification (ie, Model 2) included several predictor variables consistent with those included in Model 1 (ie, NBWT, PLUS-M), and others that were unique to Model 2 (ie, 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 (ie, 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 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

**Table 1**  
Participant demographics stratified by faller status

| Fall Group     | Age (y)     |  | Sex           | Years since Amputation |                 | Amputation Etiology                 | PLUS-M      |                               |  |  |
|----------------|-------------|--|---------------|------------------------|-----------------|-------------------------------------|-------------|-------------------------------|--|--|
|                | Mean (SD)   |  |               | Mean (SD)              | Level           |                                     | Mean (SD)   | MFCL                          |  |  |
| No falls       | 44.9 (14.4) |  | M (11) F (5)  | 16.4 (12.8)            | TT (14) TF (2)  | Dysvascular (1) Nondysvascular (15) | 58.5 (9.05) | K1 (0) K2 (4) K3 (10) K4 (2)  |  |  |
| 1 fall         | 41.4 (16.3) |  | M (2) F (5)   | 10.8 (4.8)             | TT (3) TF (4)   | Dysvascular (0) Nondysvascular (7)  | 56.2 (6.35) | K1 (0) K2 (1) K3 (2) K4 (4)   |  |  |
| $\geq 2$ falls | 55.4 (11.7) |  | M (8) F (9)   | 13.8 (14.8)            | TT (8) TF (9)   | Dysvascular (6) Nondysvascular (11) | 49.2 (6.71) | K1 (2) K2 (10) K3 (5) K4 (0)  |  |  |
| Total          | 48.7 (14.6) |  | M (21) F (19) | 14.3 (12.6)            | TT (25) TF (15) | Dysvascular (7) Nondysvascular (33) | 54.1 (8.68) | K1 (2) K2 (15) 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)<br>Mean (SD) Range | TUG (s) Median<br>(IQR) Range      | FSST (s) Median<br>(IQR) Range     | BBS (/56) Median<br>(IQR) Range    | ABC (/100) Median<br>(IQR) Range |
|-------------------|--------------------------------|------------------------------------|------------------------------------|------------------------------------|----------------------------------|
| No falls (n = 16) | 0.51 (0.13) .27-.76            | 7.86 (2.83) 5.99-13.0              | 7.56 (4.87) 5.26-12.8              | 52.0 (8.50) 40.0-55.0              | 90.3 (18.9) 63.1-100             |
| 1 fall (n = 7)    | 0.53 (0.16) .25-.74            | 7.45 (3.04) 5.06-9.18              | 6.57 (3.08) 4.14-11.8              | 52.0 (6.00) 50.0-56.0              | 90.6 (14.1) 53.1-99.2            |
| ≥2 falls (n = 17) | 0.25 (0.16)* .04-.63           | 10.0 (5.89) <sup>†</sup> 8.30-19.0 | 13.8 (6.76) <sup>†</sup> 6.27-39.3 | 45.0 (12.5) <sup>†</sup> 16.0-55.0 | 74 (26.9) 40.9-96.0              |
| Total (n = 40)    | 0.41 (0.19) .04-.76            | 8.75 (3.97) 5.06-19.0              | 8.83 (5.94) 4.14-39.3              | 50.0 (8.50) 16.0-56.0              | 85.0 (23.7) 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).

<sup>†</sup>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 fall (any falls)       |                                |                                    |                                    |                                    |                                  |
|---------------------------|--------------------------------|------------------------------------|------------------------------------|------------------------------------|----------------------------------|
| Fall Status               | NBWT (/1.0) Mean<br>(SD) Range | TUG (s) Median<br>(IQR) Range      | FSST (s) Median<br>(IQR) Range     | BBS (/56) Median<br>(IQR) Range    | ABC (/100) Median<br>(IQR) Range |
| No falls (n = 16)         | 0.51 (.13) .27-.76             | 7.86 (2.83) 5.99-13.0              | 7.56 (4.87) 5.26-12.8              | 52.0 (8.50) 40.0-55.0              | 90.3 (18.9) 63.1-100             |
| ≥1 falls (n = 24)         | 0.33 (.20)* .04-.76            | 9.39 (5.87) 5.99-13.0              | 11.4 (8.54) 4.14-39.3              | 48.5 (10.8) 16.0-56.0              | 81.8 (24.9) 40.9-99.1            |
| ≥2 falls (multiple falls) |                                |                                    |                                    |                                    |                                  |
| Fall Status               | NBWT (/1.0) Mean<br>(SD) Range | TUG (s) Median<br>(IQR) Range      | FSST (s) Median<br>(IQR) Range     | BBS (/56) Median<br>(IQR) Range    | ABC (/100) Median<br>(IQR) Range |
| 0-1 falls (n = 23)        | 0.51 (.14) .25-.76             | 7.79 (1.66) 5.06-13.0              | 7.46 (3.75) 4.14-12.8              | 52.0 (7.00) 40.0-56.0              | 90.6 (15.6) 53.1-100             |
| ≥2 falls (n = 17)         | 0.25 (.16)* .04-.63            | 10.0 (5.89) <sup>†</sup> 8.30-19.0 | 13.8 (6.76) <sup>†</sup> 6.27-39.3 | 45.0 (12.5) <sup>†</sup> 16.0-56.0 | 74 (26.9) 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 fall, as well as ≥2 falls and 0-1 falls  $P \leq .005$  (2-sided Independent  $t$  Test).

<sup>†</sup>Test scores significantly different between ≥2 falls and 0-1 fall  $P \leq .001$  (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 fall (any falls)       |                |              |                      |                      |                |                 |
|---------------------------|----------------|--------------|----------------------|----------------------|----------------|-----------------|
| Test                      | AUC (95% CI)   | Cutoff Score | Sensitivity (95% CI) | Specificity (95% CI) | LR+ (95% CI)   | LR- (95% CI)    |
| NBWT (/1.00)              | 0.81 (.62-.91) | ≤0.43        | 73% (53%-90%)        | 76% (54%-96%)        | 3.0 (1.5-6.9)  | 0.36 (.19-.77)  |
| TUG (s)                   | 0.71 (.54-.88) | ≥8.17        | 83% (68%-98%)        | 68% (46%-92%)        | 2.6 (1.3-5.6)  | 0.24 (.13-.56)  |
| FSST (s)                  | 0.70 (.53-.86) | ≥8.49        | 74% (58%-92%)        | 68% (46%-92%)        | 2.4 (1.1-5.2)  | 0.36 (.17-.78)  |
| BBS (/56)                 | 0.66 (.47-.83) | ≤50.5        | 67% (48%-86%)        | 62% (39%-86%)        | 1.8 (.89-3.6)  | 0.53 (.27-1.1)  |
| ABC (/100)                | 0.65 (.47-.82) | ≤80.2        | 50% (30%-70%)        | 74% (54%-96%)        | 1.9 (.78-5.1)  | 0.67 (.41-1.1)  |
| Model 1                   | 0.84 (.71-.96) | N/A          | 80% (64%-96%)        | 73% (51%-96%)        | 3.0 (1.3-7.1)  | 0.27 (.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)              | 0.89 (.78-1.0) | ≤0.43        | 88% (73%-100%)       | 79% (65%-92%)        | 4.2 (1.7-6.9)  | 0.16 (.042-.60) |
| TUG (s)                   | 0.88 (.78-.98) | ≥9.25        | 82% (64%-100%)       | 83% (67%-98%)        | 4.7 (1.9-11.9) | 0.21 (.075-.61) |
| FSST (s)                  | 0.88 (.76-.99) | ≥8.71        | 94% (83%-100%)       | 74% (60%-92%)        | 3.6 (1.8-7.3)  | 0.08 (.012-.54) |
| BBS (/56)                 | 0.81 (.66-.92) | ≤50.5        | 88% (73%-100%)       | 70% (51%-88%)        | 2.9 (1.5-5.5)  | 0.17 (.045-.64) |
| ABC (/100)                | 0.71 (.54-.87) | ≤80.2        | 65% (42%-87%)        | 78% (61%-95%)        | 3.0 (1.3-7.0)  | 0.45 (.23-.89)  |
| Model 2                   | 0.90 (.79-1.0) | N/A          | 76% (56%-97%)        | 83% (67%-98%)        | 4.4 (1.7-11.3) | 0.28 (.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.

probability of falling among unilateral LLP users. The distribution of demographic (eg, age), amputation (eg, level, etiology, and time since), and activity characteristics (eg, 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 = 210-1568),<sup>24,54-58</sup> although skewed slightly toward 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,<sup>21</sup> who established cutoff scores and validity indices for several clinical balance tests (eg, FSST, TUG) using data from a sample of short-term (ie, 6 months postdischarge) transtibial prosthesis users. Notable differences were observed between the cutoff scores in the present study and those reported by Dite et al<sup>21</sup>. Dite et al<sup>21</sup> used a faller classification of  $\geq 2$  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 (ie, 8.17 and 8.49 seconds, respectively) if the " $\geq 1$  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, whereas Dite et al<sup>21</sup> included only transtibial prosthesis users. However, the higher prevalence<sup>6</sup> and risk of falls<sup>59</sup> 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<sup>16,22</sup> or FSST<sup>16</sup> 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<sup>21</sup> studied participants less than a year after amputation (ie,  $6.4 \pm 1.5$  months after discharge from the rehabilitation unit), whereas 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 after amputation and that balance ability and fall risk may change markedly after 6 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 complementary 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 (ie, 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).<sup>55</sup> 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 (ie,  $LR+ = 3.0$ ) to be a faller than a nonfaller. Likelihood ratios for the NBWT, TUG, and FSST were accompanied by 95% CIs that did not overlap with 1.0 (ie, 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.<sup>1</sup> Although 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.<sup>3,5,7,9,11</sup> Although the LR were greater in the NBWT, TUG, and FSST, in many cases their CIs 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 (ie, etiology and level of amputation)<sup>8,59</sup> 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 (ie, discriminate, evaluate, or predict). Examples of such applications may include discriminating between fallers and nonfallers 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 nonfallers 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  $\geq 2$  falls. The observed increase in validity indices based on a more conservative faller classification (ie,  $\geq 2$  falls vs  $\geq 1$  fall) is consistent with previous research.<sup>8</sup> 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  $\geq 2$  falls classification magnifies differences between fallers and nonfallers, 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 (ie,  $\geq 2$  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.<sup>60-64</sup> 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, although consistent with similar studies,<sup>15,21,65</sup> should be increased in future research. A larger sample would facilitate development of cutoff scores to discriminate nonfallers (ie, zero falls), from single fallers (ie, 1 fall), and multiple fallers (ie,  $\geq 2$  falls), as well as across score intervals for balance tests with a continuous scale (eg, the NBWT, TUG, and FSST).<sup>1</sup> The study sample also consisted of a larger percentage of traumatic LLP users (ie, 62.5%) than is reported in the literature (ie, 17%-60.2%),<sup>24,54-58</sup> potentially limiting the generalization of study results. Although this is a challenge in most,<sup>15,23,25,66,67</sup> but not all<sup>8,21,68-70</sup> 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 years, was slightly younger than that reported in larger, national studies of individuals with lower limb amputation (ie, mean age from 50 to 55),<sup>54-56,58</sup> 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 (eg, age  $\geq 65$ ) may be warranted to examine balance test cutoff scores in LLP users in that specific subpopulation (ie, Medicare-eligible individuals).

Falls were assessed retrospectively. This may underestimate fall frequency<sup>71</sup> and lend itself to recall bias.<sup>72,73</sup> 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.<sup>74</sup>

Fall-related injuries were not recorded. Determining whether multiple versus nonmultiple fallers are more likely to suffer a fall-related injury,<sup>3,10</sup> 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,<sup>75</sup> protective stepping,<sup>76</sup> number of medications,<sup>9</sup> and sense of vibration<sup>77</sup> 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 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 (ie, testing the cutoff scores and indices proposed here) of these tests is needed to facilitate their widespread adoption in clinical care.

### Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

### 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.

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.
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.
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*. 2019;11(4):344-353.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
17. Berg K, Wooddauphinee S, Williams JL, 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.
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.
20. Powell LE, Myers AM. The activities-specific balance confidence (ABC) scale. *J Gerontol A Biol Sci Med Sci*. 1995;50A(1):M28-M34.
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.
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.
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.
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.
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*. 2019;11(3):243-251.
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.
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.
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.
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.
30. Elm v 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.
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.
32. Hanspal RS, Fisher K, Nieveen R. Prosthetic socket fit comfort score. *Disabil Rehabil*. 2003;25(22):1278-1280.
33. Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. *BMJ*. 1994;309(6948):188.
34. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ*. 2004;329(7458):168-169.
35. Greenhalgh T. How to read a paper. Papers that report diagnostic or screening tests. *BMJ*. 1997;315(7107):540-543.
36. Hafner BJ, Gaunaud 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.
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.
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.
40. Hauer K, Lamb SE, Jørstad 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.
41. Clemens SM, Gailey RS, Bennett CL, Pasquina PF, Kirk-Sanchez NJ, Gaunaud 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.
42. Portney LG, Watkins MP. *Foundations of Clinical Research*. Upper Saddle River, NJ: Pearson; 2009.
43. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic curve. *Radiology*. 1982;143(1):29-36.
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.
45. Haynes RB. *Clinical Epidemiology: how we Do Clinical Practice Research*. Philadelphia, PA: Lippincott, Williams, and Wilkins; 2006.
46. Akobeng AK. Understanding diagnostic tests 3: receiver operating characteristic curves. *Acta Paediatr*. 2007;96(5):644-647.

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.
48. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical Epidemiology: A Basic Science for Clinical Medicine*. Philadelphia, PA: Lippincott, Williams, and Wilkins; 1991.
49. Altman DG, Bland JM. Statistics notes: diagnostic tests 1: sensitivity and specificity. *BMJ.* 1994;308(6943):1552-1552.
50. Altman DG, Bland JM. Statistics notes: diagnostic tests 2: predictive values. *BMJ.* 1994;309(6947):102-102.
51. Crombie DL. Diagnostic process. *J Coll Gen Pract.* 1963;6:579-589.
52. Sackett DL. A primer on the precision and accuracy of the clinical examination. *JAMA.* 1992;267(19):2638-2644.
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.
54. Ziegler-Graham K, MacKenzie EJ, Ephraim PL, Trivison 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.
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.
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.
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.
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.
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. Gaunaud 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.
62. Jette DU, Halbert J, Iverson C, Miceli E. Use of standardized outcome measures in physical therapist practice: perceptions and applications. *Phys Ther.* 2009;89:125-135.
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.
64. Valier A, Jennings AL, Parsons JT. Benefits of and barriers to using patient-rated outcome measures in athletic training. *J Athl Train.* 2014;49(5):674-683.
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.
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.
67. Lin S-J, Bose NH. Six-minute walk test in persons with transtibial amputation. *Arch Phys Med Rehabil.* 2008;89(12):2354-2359.
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.
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.
70. Mandel A, Paul K, et al. Balance confidence and activity of community-dwelling patients with transtibial amputation. *J Rehabil Res Dev.* 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.
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.
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):1147-1148.
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.
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.
77. Quai TM, Brauer SG, Nitz JC. Somatosensation, circulation and stance balance in elderly dysvascular transtibial amputees. *Clin Rehabil.* 2005;19(6):668-676.

---

**This journal-based CME activity is designated for 1.0 AMA PRA Category 1 Credit and can be completed online at <http://www.me.aapmr.org>. This activity is FREE to AAPMR members and available to nonmembers for a nominal fee. For assistance with claiming CME for this activity, please contact (847) 737-6000.**

---

## Disclosure

A.S. Department of Kinesiology and Nutrition, University of Illinois at Chicago, Chicago, IL. Address correspondence to: A. S. Department of Kinesiology and Nutrition, University of Illinois at Chicago, 1919 W Taylor Street, Chicago, IL 60612, USA. Tel: 312-996-1427, Fax: 312-413-0319, e-mail: asawers@uic.edu  
Disclosure: none

B.J.H. Department of Rehabilitation Medicine, University of Washington, Seattle, WA  
Disclosure: none

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.

All financial disclosures and CME information related to this article can be found on the me<sup>®</sup> site (<http://me.aapmr.org/>) prior to accessing the activity.

Submitted for publication January 10, 2019; accepted March 18, 2019.

**CME Question**

According to this article, a score equal to or greater than 0.43 on the narrowing-beam walking test (NBWT) would suggest a lower limb prosthesis (LLP) user is how many times more likely to be a faller than a non-faller?

- a. 0
- b. 1
- c. 2
- d. 3

**Answer online at** <http://me.aapmr.org>

A. Sawers, CPO, PhD, Department of Kinesiology, University of Illinois at Chicago, 1919 W. Taylor Street, Chicago, IL 60612 (USA). Address all correspondence to Dr. Sawers at: [asawers@uic.edu](mailto:asawers@uic.edu).

J. Kim, PT, MPT, Department of Kinesiology, University of Illinois at Chicago.

G.S. Balkman, CPO, MSPO, Department of Rehabilitation Medicine, University of Washington, Seattle, Washington.

B.J. Hafner, PhD, Department of Rehabilitation Medicine, University of Washington.

[Sawers A, Kim J, Balkman G, Hafner BJ. Interrater and test-retest reliability of performance-based clinical tests administered to established users of lower limb prostheses. *Phys Ther*. 2020;100:1206–1216.]

© 2020 American Physical Therapy Association

Published Ahead of Print:

April 13, 2020

Accepted: February 3, 2020

Submitted: May 7, 2019

## Interrater and Test-Retest Reliability of Performance-Based Clinical Tests Administered to Established Users of Lower Limb Prostheses

Andrew Sawers, Janis Kim, Geoff Balkman, Brian J. Hafner

**Objective.** A major barrier to reducing falls among users of lower limb prostheses (LLP) has been an absence of statistical indices required for clinicians to select and interpret scores from performance-based clinical tests. The study aimed to derive estimates of reliability, measurement error, and minimal detectable change values in performance-based clinical tests administered to unilateral LLP users.

**Methods.** A total of 60 unilateral LLP users were administered the Narrowing Beam Walking Test, Timed “Up and Go” (TUG), Four Square Step Test (FSST), and 10-Meter Walk Test on 2 occasions, 3 to 9 days apart. Intraclass correlation coefficients (ICCs) were calculated to assess interrater and test-retest reliability, while standard error of measurement (SEM) and minimal detectable change (MDC<sub>90</sub>) were derived to establish estimates of measurement error in individual scores or changes in score for each test.

**Results.** Interrater reliability ICCs (1,1) were high for all tests (ie,  $\geq 0.98$ ). Test-retest ICCs (2,1) varied by test, ranging from .88 for the TUG to .97 for the FSST. SEM and MDC<sub>90</sub> varied between .39 and .96 and between .91 seconds and 2.2 seconds for the time-based tests (FSST, TUG, 10-Meter Walk Test). SEM and MDC<sub>90</sub> for the Narrowing Beam Walking Test were .07 and .16, respectively.

**Conclusion.** With the exception of the TUG, studied tests had test-retest ICCs (2,1) that exceeded the minimum required threshold to be considered suitable for group- and individual-level applications (ie, ICC  $\geq 0.70$  and  $\geq 0.90$ , respectively). Future research on individuals with dysvascular and transfemoral amputations or in specific age categories is required.

**Impact.** Along with published validity indices, these reliability, error, and change indices can help clinicians select balance tests suitable for LLP users. They can also help clinicians interpret test scores to make informed, evidence-based clinical decisions.



Post a comment for this article at:  
<https://academic.oup.com/ptj>

The use of valid and reliable instruments to document clinical outcomes is increasingly expected by employers and third-party payers to justify and evaluate rehabilitation services and interventions provided to lower limb prosthesis (LLP) users.<sup>1-3</sup> Despite the demand for robust documentation, several barriers inhibit regular use of standardized outcome measures in clinical practice.<sup>2-6</sup> Selection of suitable outcome measures for different clinical applications<sup>5,7-9</sup> and interpretation of test scores to guide clinical decision-making<sup>6,7</sup> are frequently cited as challenges to routine use of outcome measures in clinical practice. These barriers may limit, among other things, regular assessment of balance ability and fall risk among LLP users, potentially contributing to the persistence of falls<sup>10-14</sup> and their related consequences<sup>15-17</sup> among LLP users.

Indices of reproducibility (eg, estimates of reliability, error, and detectable change) in the population of interest (ie, LLP users)<sup>18-20</sup> are needed to formulate recommendations for performance-based clinical tests suited to the care of LLP users.<sup>18,19,21</sup> Interrater and test-retest reliability, often quantified using intraclass correlation coefficients (ICC),<sup>22</sup> are estimates of the stability of measurement between raters and over a period of time when no changes in health are believed to have occurred.<sup>19,23,24</sup> ICCs therefore provide test administrators with important information about the consistency of measurement(s) produced by a test within a given population. The magnitude of reliability estimates can also be used to guide clinicians in selecting performance-based tests for specific applications. To be recommended for group-level comparisons, a clinical test should generally demonstrate a test-retest ICC of .70 or greater.<sup>25-28</sup> For applications that involve individual-level decisions, a test should have higher reliability, recommended as .90 or greater.<sup>19,25,28</sup> Reliability estimates do not, however, help clinicians interpret test scores or provide the information needed to make evidence-based decisions for individual patients. The interpretation of individual test scores requires knowledge of the standard error of measurement (SEM) and minimal detectable change (MDC) of a test.

The SEM reflects the error associated with a single measurement.<sup>19</sup> The SEM can therefore be used to estimate the precision of an individual's score on a test,<sup>19,29,30</sup> as it can be used to derive confidence intervals (CIs) around a score. The MDC is an estimate of the amount of change needed to exceed the error associated with 2 measurements on the same test within the same individual.<sup>31</sup> An individual patient is considered to have changed only when the difference between the previous score and the current score exceeds the MDC of the test.<sup>32</sup> The SEM and MDC can therefore help clinicians interpret test scores among individual patients by monitoring whether there is a real difference in their health or

performance with respect to a reference value (ie, using SEM) and/or whether their health or performance has changed between testing sessions (ie, using MDC).<sup>19</sup>

Few performance-based clinical tests with evidence of validity among LLP users have been assessed for evidence of reproducibility.<sup>33-36</sup> In the absence of reliability indices like test-retest ICCs or SEM and MDC values, clinicians cannot select tests suitable for individual-level decision-making or interpret test scores of individual LLP users in clinical practice. Efforts to study the reproducibility of performance-based clinical tests administered to LLP users have been limited by several factors. First, sample sizes are generally small<sup>33-35</sup> and often below recommendations for establishing reliability indices.<sup>37</sup> Second, evidence of reproducibility for some tests is dated and may not reflect best practices for assessing test-retest reliability.<sup>35</sup> Lastly, study methodology (eg, selection of raters and corresponding analysis) may limit the generalizability of study results to the broader population of LLP users and test administrators.<sup>34,36</sup> Additionally, a number of common performance-based clinical tests (eg, FSST, 10MWT, TUG) have not yet been assessed for interrater reliability in LLP users. Many performance-based tests also evaluate unique aspects of balance and walking (ie, straight ahead walking [10-Meter Walk Test, 10MWT], turns [Timed "Up and Go", TUG], balancing along beam [Narrowing Beam Walking Test, NBWT], and multi-planar stepping [Four Square Step Test, FSST]). As such, it is important to evaluate each test and determine if it and its components are suitable for clinical practice. Finally, given that test-retest reliability is relative to the sample of participants from which it is derived,<sup>38</sup> and the evaluation of a test's reliability is a continual process (ie, tests are not reliable or unreliable; they exhibit more or less evidence of reliability), it is beneficial to accumulate additional evidence of reliability for a test by comparing, contrasting, and confirming values across different samples of LLP users. Therefore, insufficient psychometric evidence exists to generate recommendations for the selection and interpretation of common performance-based clinical tests administered to LLP users.

The objective of this study was therefore to assess the reproducibility of performance-based clinical tests administered to established unilateral LLP users, deriving estimates of reliability indices to help clinicians and researchers select and interpret the studied tests. Specifically, the ICCs for interrater and test-retest reliability were assessed, from which the standard SEM and MDC of each of the studied tests were derived. We hypothesized that the studied tests (ie, NBWT, FSST, TUG, and 10MWT) would demonstrate evidence of reliability matching or exceeding that required for group-level comparisons (ie,  $ICC \geq 0.7$ ) and individual decision-making (ie,  $ICC \geq 0.9$ ) in assessments of balance and fall risk.

## Methods

### Study Design

A 2-site (Chicago and Seattle) repeated-measures study was conducted between February 2018 and November 2018 to assess the reproducibility of the NBWT,<sup>39</sup> TUG,<sup>40</sup> FSST,<sup>41</sup> and 10MWT<sup>42</sup> when administered to LLP users. Best practices in conducting observational studies (ie, Strengthening the Reporting of Observational Studies in Epidemiology<sup>43</sup>) as well as the collection and reporting of reliability data (ie, Consensus-Based Standards for the Selection of Health Measurement Instruments)<sup>19,20,37</sup> were followed during the study. All data were stored and managed using a REDCap database hosted at the University of Illinois at Chicago.<sup>44</sup> Study protocols were reviewed and approved by institutional review boards at each site. All individuals provided written consent prior to participation.

### Participants

Study flyers were used to recruit LLP users from local prosthetic clinics in Chicago and Seattle. A convenience sampling method was used. Inclusion criteria were 18 years of age or older; unilateral amputation at the transtibial or transfemoral level; amputation due to trauma, dysvascular complications, tumor, or infection; use of a prosthesis for at least 1 year (ie, established lower limb prosthesis users); use of a comfortable prosthesis to ambulate; able to walk 10 m without an upper extremity mobility aid (eg, cane); and able to read, write, and speak English. Participants were excluded from participation if they had an amputation of a second limb (eg, arm or contralateral leg), contralateral complications (eg, hip replacement, ulcers, or infections), advanced neurologic disorder (eg, stroke, multiple sclerosis, Parkinson's disease), or severe cardiopulmonary disease that would prevent them from completing study procedures.

An a priori sample size of 38 LLP users was estimated based on a desired 95% CI with a width  $\geq 0.10$  around an ICC of 0.80.<sup>45</sup> Psychometric standards, however, recommend a minimum sample of 50 participants to estimate reliability statistics.<sup>19</sup> Based on the more conservative estimate (ie, recommended psychometric standards) and planning for modest attrition (ie, 20%) between test and retest sessions,<sup>46,47</sup> a sample of 60 LLP users was targeted.

### Procedures

At their first visit, all participants were administered demographic and characterization measures as well as 4 performance-based clinical tests. At the end of the first test session, each participant was scheduled a retest session 3 to 9 days later. This test-retest interval was chosen to provide sufficient time to limit familiarity with the tests and to avoid the possibility of changes in health or prosthetic fit. Participants wore their desired prosthesis and footwear, both of which were held constant between test sessions.

## Measurements

### Participant demographic and characterization measures.

Sociodemographic information including age, height, weight, sex, race, and ethnicity was collected via self-report. Amputation characteristics (ie, level, etiology, and time since amputation), prosthetic-related factors (ie, components and time since delivery), and mobility (ie, Medicare Functional Classification Level,<sup>48</sup> Prosthetic Limb Users Survey of Mobility,<sup>49</sup> and typical prosthesis use per day) were collected via self-report, interview, and/or clinical inspection. The number of comorbid conditions was collected via self-report using the Charlson Comorbidity Index.<sup>50</sup> Comfort of the prosthesis worn by each participant was evaluated using the Socket Comfort Score (SCS) scored from 0 to 10.<sup>51,52</sup> At retest, participants were asked whether they had experienced changes in their prosthesis and/or health since the first testing session. The SCS was readministered at retest to confirm that participants' scores were within the MDC<sub>90</sub> of the SCS in LLP users (ie,  $\pm 2.82$  points).<sup>52</sup>

### Administration of performance-based clinical tests.

The TUG,<sup>40</sup> FSST,<sup>41</sup> 10MWT,<sup>42</sup> and NBWT<sup>39</sup> were administered in a randomized order and scored according to standardized instructions (Appendix). The same test sequence was used in both test sessions. Each balance test was selected because it could be performed in less than 5 minutes, had been designed or used to assess fall risk in other populations, and had evidence of validity among LLP users.<sup>35,39,53,54</sup>

Standardized protocols were established to ensure fidelity in test administration and scoring between study sites and raters. All raters were instructed in test procedures using a common set of training documents. All raters practiced administration of the tests under the supervision of study (or site) principal investigators. Test equipment was purchased or fabricated using explicit written instructions and verified with photographs and videos during a training session. Verbal instructions were standardized to help ensure consistency in data collection across the 2 study sites.<sup>55</sup> Further, no communication was allowed between concurrent raters during testing to limit their awareness of each other's assessments, which could influence scoring.<sup>56</sup> Finally, data collected on standardized paper forms were double entered by research staff to minimize data-entry errors.<sup>57</sup> Participants were provided 2 minutes rest between tests, or more if requested.

A small pool of qualified raters was consistently used at each study site. Specifically, across study sites, raters included 2 researchers (ie, PhD), 3 PhD students (1 certified prosthetist-orthotist, 1 physical therapist, and 1 bioengineer), and 3 prosthetic and orthotic graduate students. These individuals are representative of those most likely to administer clinical tests to lower limb

prosthesis users and thus well represent raters who would use these tests, thereby contributing to the generalizability of the subsequent reliability indices.<sup>20</sup> Each rater was trained to administer the tests according to established standardized protocols (Appendix). Testing during session 1 was performed by 2 raters from the pool of raters at each site. The same raters were not used for every participant to improve generalizability of the study results. Testing during session 2 was performed by 1 of the raters at each site who had participated in session 1.

### Data Analysis

Continuous variables were evaluated for departures from normality using the Shapiro-Wilk test.<sup>58</sup> Measures of central tendency and dispersion were calculated for each variable based on their distribution. Paired 2-sided *t* tests were used to test for differences in test scores between raters and between test sessions (ie, effect of time). A Bonferroni correction ( $\alpha = .0125$ ) was used for multiple comparisons within the sample.

Interrater reliability of each clinical test was evaluated by calculating the ICC (1,1) from measurements taken simultaneously by 2 raters per participant.<sup>22,38,59</sup> Each clinical test was also evaluated for absolute test-retest reliability by calculating the ICC (2,1) from measurements taken on 2 visits by the same rater.<sup>22,38</sup> The 95% CIs were derived for each ICC using the F-distribution.<sup>60</sup> Test-retest ICCs were compared with established thresholds to determine the suitability of each test for clinical and research applications (ie,  $\geq 0.70$  for group-level comparisons,<sup>25–28</sup> and  $\geq 0.90$  for individual-level decisions<sup>19,25–28</sup>).

SEM estimates were calculated as  $SEM = SD * \sqrt{(1 - ICC)}$ , where SD is the pooled SD of test scores across both test sessions from the same rater (ie, rater A),<sup>30</sup> and ICC is the

test-retest ICC.<sup>29,33</sup> Alternative sources of the SD were explored (eg, SD from the first test session<sup>33</sup> and total sum of the squares from a 1-way ANOVA<sup>33,38,59</sup>), but neither resulted in an appreciable change in SEM values. MDC estimates were calculated as  $MDC_{90} = 1.64 * \sqrt{SEM * \sqrt{2}}$ , where 1.64 is the z-score for a 90% CI.<sup>38</sup>

### Role of the Funding Source

Research reported in this publication was supported by the US Department of Defense. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Department of Defense.

### Results

Sixty ambulatory unilateral LLP users with a mean age of 55.1 years (SD = 21.3 years), 42 of whom were male, were recruited and completed the study. Amputation, prosthesis, and mobility characteristics of the study sample are presented in Table 1. Also, 48% of participants reported no comorbid conditions, while 20%, 13%, and 10% reported 1, 2, or 3 or more comorbid conditions (eg, asthma, diabetes, HIV/AIDS), respectively. Observed test scores (Suppl. Material) for the FSST, TUG, and 10MWT were nonnormally distributed (Shapiro-Wilk  $W = 0.798–0.919$ ,  $P \leq .001$ ). In contrast, NBWT scores were normally distributed (Shapiro-Wilk  $W = 0.957–0.974$ ,  $P \geq .195$ ). Despite departures from normality, parametric 2-sided *t* tests were used to compare test scores between rater and across test sessions because of their robustness to withstand violations of normality in larger samples (ie,  $n > 30$ ).<sup>61</sup>

### Interrater Reliability

Two-sided paired *t* tests revealed no statistically significant differences in mean test scores between raters for the TUG ( $t = 1.609$ ,  $P = .113$ ), FSST ( $t = -0.753$ ,  $P = .455$ ),

**Table 1.**  
Amputation, Prosthetic, and Activity Characteristics of the Study Sample<sup>a</sup>

| Amputation Characteristics |                                                                                                 |                               | Prosthetic Characteristics                                           |                             | Activity Characteristics                 |                  |                                |
|----------------------------|-------------------------------------------------------------------------------------------------|-------------------------------|----------------------------------------------------------------------|-----------------------------|------------------------------------------|------------------|--------------------------------|
| Level                      | Etiology                                                                                        | Time Since Amputation (years) | Components Feet                                                      | Time Since Delivery (years) | MFCL                                     | PLUS-M (T-score) | Typical Prosthetic Use (hrs/d) |
|                            |                                                                                                 |                               | Knees                                                                |                             |                                          |                  |                                |
| Participants (/60)         | No. of Participants (/60)                                                                       | Mean (SD)                     | Participants (/60)                                                   | Mean (SD)                   | Participants (/60)                       | Mean (95% CI)    | Mean (SD)                      |
| Transfemoral (n = 15)      | Trauma (n = 26)<br>Infection (n = 13)<br>Dysvascular (n = 11)<br>Tumor (n = 9)<br>Other (n = 1) | 16.0 (14.8)                   | Energy storing (53)<br>Single-axis (3)<br>Multi-axis (2)<br>SACH (2) | 3.5 (3.0)                   | K2 (n = 18)<br>K3 (n = 34)<br>K4 (n = 8) | 56.1 (1.77)      | 13.9 (3.5)                     |
|                            |                                                                                                 |                               | Microprocessor (13)<br>Non-Microprocessor (2)                        |                             |                                          |                  |                                |

<sup>a</sup>CI = confidence interval; IQR = interquartile range; MFCL = Medicare Functional Classification Level (K-level); SACH = solid ankle cushioned heel; d = day.

## Clinical Test Reliability: People With Amputation

**Table 2.**

Interrater Reliability of Performance-Based Clinical Tests in People With Lower Limb Loss<sup>a</sup>

| Test              | Rater A Session 1 Score Mean (SD) | Rater B Session 1 Score Mean (SD) | Score Difference Mean (SD) | ICC (1,1) (95% CI) |
|-------------------|-----------------------------------|-----------------------------------|----------------------------|--------------------|
| NBWT (score/1.00) | 0.42 (0.21)                       | 0.41 (0.22)                       | 0.01 (0.03)                | 0.99 (0.98–0.99)   |
| TUG (time, sec)   | 10.4 (2.76)                       | 10.4 (2.79)                       | 0.05 (0.27)                | 0.99 (0.99–1.0)    |
| FSST (time, sec)  | 10.4 (5.27)                       | 10.4 (5.21)                       | –0.03 (–0.37)              | 0.99 (0.99–1.0)    |
| 10MWT (time, sec) | 8.74 (1.96)                       | 8.76 (1.96)                       | –0.02 (–0.30)              | 0.99 (0.98–0.99)   |

<sup>a</sup>10MWT = 10-Meter Walk Test; FSST = Four Square Step Test; ICC = intraclass correlation coefficient; NBWT = Narrowing Beam Walking Test; TUG = Timed “Up and Go”.

**Table 3.**

Test-Retest Reliability of Performance-Based Clinical Tests in People With Lower-Limb Loss<sup>a</sup>

| Test              | Session 1 Score Mean (SD) | Session 2 Score Mean (SD) | Score Change Mean (SD)   | ICC (2,1) (95% CI) | SEM  | MDC <sub>90</sub> (95% CI) |
|-------------------|---------------------------|---------------------------|--------------------------|--------------------|------|----------------------------|
| NBWT (score/1.00) | 0.42 (0.21)               | 0.44 (0.24)               | 0.02 (0.10)              | 0.90 (0.83–0.94)   | 0.07 | 0.16 (0.10–0.20)           |
| TUG (time, sec)   | 10.4 (2.76)               | 9.91 (2.75)               | 0.50 (1.30) <sup>b</sup> | 0.88 (0.78–0.93)   | 0.96 | 2.2 (1.5–2.9)              |
| FSST (time, sec)  | 10.4 (5.27)               | 10.3 (4.83)               | 0.14 (1.32)              | 0.97 (0.94–0.98)   | 0.87 | 2.0 (0.79–3.5)             |
| 10MWT (time, sec) | 8.74 (1.96)               | 8.58 (2.00)               | 0.16 (0.54)              | 0.96 (0.93–0.98)   | 0.39 | 0.91 (0.40–1.4)            |

<sup>a</sup>10MWT = 10-Meter Walk Test; FSST = Four Square Step Test; MDC<sub>90</sub> = minimal detectable change; NBWT = Narrowing Beam Walking Test; SEM = standard error of measurement; TUG = Timed “Up and Go”.

<sup>b</sup>Test score significantly differed between test sessions (2-sided paired *t* test) *P* = .004.

NBWT (*t* = 1.276, *P* = .207), or 10MWT (*t* = –0.518, *P* = .607) (Tab. 2). Interrater reliability ICCs (model 1,1) were high (ie, ≥0.98) for all clinical tests (Tab. 2).

### Test-Retest Reliability

Retest sessions occurred at a mean of 6.7 days (SD = 2.3 day) after the initial testing session. No changes in health, prosthetic prescription, prosthesis function, or SCS (test session 1: 7.62 [SD = 1.97], test session 2: 7.43 [SD = 1.61], *t* = 1.026, *P* = .309) were reported between test sessions by study participants as a group or individually. Statistically significant differences in test scores were observed between test sessions for the TUG (*t* = 2.968, *P* = .004) but not the 10MWT (*t* = 2.315, *P* = .024), FSST (*t* = 0.821, *P* = .415), or NBWT (*t* = –1.821, *P* = .066) (Tab. 3). Test-retest ICCs (model 2,1) varied by test, ranging from 0.88 for the TUG to 0.97 for the FSST (Tab. 3).

### SEM and MDC<sub>90</sub>

SEM and MDC<sub>90</sub> varied between 0.39 and 0.96 and between 0.91 seconds and 2.2 seconds, respectively, among the time-based tests (ie, 10MWT, TUG, and FSST) (Tab. 3). Estimates of SEM and MDC<sub>90</sub> for the NBWT, a test of normalized walking distance, were 0.07 and 0.16, respectively.

## Discussion

The objectives of this study were to assess the reproducibility of performance-based clinical tests administered to established unilateral LLP users and derive estimates of reliability indices to help clinicians and researchers select and interpret the studied tests. ICCs indicated that all studied tests are appropriate for group-level comparisons (ie, ICC ≥ 0.70), while the FSST, 10MWT, and NBWT are also suitable for individual-level applications among established unilateral LLP users (ie, ICC ≥ 0.90). With the exception of the TUG, the results supported the study hypothesis that the studied tests (ie, NBWT, FSST, TUG, and 10MWT) would demonstrate evidence of reliability matching or exceeding that required for group-level comparisons (ie, ICC ≥ 0.7) and individual decision-making (ie, ICC ≥ 0.9). Estimates of the SEM and MDC for the studied tests provide clinicians and researchers with indices to assess the precision of individual measurements (ie, SEM) and to evaluate changes in measurements over time (ie, MDC<sub>90</sub>). Our sample (*n* = 60) was larger than previous reproducibility studies of performance-based clinical balance tests with LLP users (ie, *n* = 15–44)<sup>33,34</sup> and exceeded the minimum threshold (ie, *n* = 50) recommended by Consensus-Based Standards for the Selection of Health Measurement Instruments recommendations for “good” estimates of reliability indices.<sup>37</sup> The distribution of demographic (eg, age), amputation (eg, level, etiology, and time since), and

activity characteristics (eg, Prosthetic Limb Users Survey of Mobility, Medicare Functional Classification Level) of participants in the present study were generally comparable with those reported in larger national studies of individuals with lower limb amputation (ie,  $n = 210$  to  $1568$ ),<sup>52,62–65</sup> although skewed slightly towards a higher proportion of traumatic and transtibial amputations. The overall similarity, however, with these samples suggests that the current results can be generalized to the broader population of established unilateral LLP users.

### Reliability Indices in the Literature

The magnitude of the reliability indices (ie, test-retest ICC, SEM, and MDC) for the studied tests varied with respect to previously reported values. Test-retest ICCs (Tab. 3) were consistent with those previously reported for the 10MWT (ie,  $0.93$ – $0.98$ )<sup>66</sup> as well as the TUG, as described by Resnik et al (ie,  $0.88$  [95% CI =  $0.80$ – $0.94$ ])<sup>33</sup>, but not Clemens et al (ie,  $0.98$  [95% CI =  $0.97$ – $0.99$ ])<sup>36</sup> or Jayakaran et al (ie,  $0.95$  [95% CI =  $0.85$ – $0.98$ ])<sup>34</sup>. The larger TUG test-retest ICC reported by Clemens et al.<sup>36</sup> may be explained by their experimental design. Calculating the test-retest ICC using back-to-back trials of the TUG rather than trials administered several days apart<sup>33,38</sup> excluded session-to-session variation within participants, an important component of error when calculating test-retest ICCs.<sup>38</sup> Excluding session-to-session variation likely increases the magnitude of test-retest ICCs and attenuates the corresponding indices of measurement error. The test-retest ICC for the TUG reported by Jayakaran et al<sup>34</sup> may have been affected by their atypical scoring (ie, the mean of 2 trials rather than the conventional “best” of 2 trials).<sup>40</sup> Using the mean may have stabilized the score, reduced error, and increased reliability. To date, evidence of test-retest reliability for the FSST has been limited to other patient populations (eg, Parkinson’s disease, multiple sclerosis, post-stroke)<sup>67</sup> but has yet to be reported among LLP users.

Departures from previously reported values of the SEM were found for the studied tests. The SEM for the TUG in the present study (ie,  $0.96$ ) was larger than a previously reported value,  $0.55$ ,<sup>36</sup> and smaller than another,  $1.6$ .<sup>33</sup> Since the SEM is an arithmetic function of the test-retest ICC and the SD of test scores from a study sample, the differences can be attributed to the larger test-retest ICC reported by Clemens et al<sup>36</sup> and the nearly 2-fold increase in between-participant variability reported by Resnik et al<sup>33</sup>, respectively.<sup>68</sup> To date, the SEM for the 10MWT or FSST among LLP users has not been reported.

Point estimates for the MDC<sub>90</sub> also differed from previously reported values.<sup>33,36</sup> Given that MDC is a scalar product of the SEM, these differences can be explained by the earlier discussion of observed differences in SEM. Despite the disparities in MDC<sub>90</sub> point estimates in this study and those reported previously, the 95% CIs of

MDC<sub>90</sub> derived here overlapped with previously reported values. This implies that while diverse, there is some consistency with regard to what constitutes a true change in performance over time within each of the studied tests.

This study provides the first evidence, to our knowledge, of interrater reliability using ICCs for the TUG, FSST, and 10MWT in LLP users (Tab. 2) The reported interrater ICC of the TUG (ie,  $0.99$  [95% CI =  $0.99$ – $1.0$ ]) is consistent with what has been reported using a different approach (ie,  $0.96$  with Pearson’s  $r$ ).<sup>35</sup>

### Application and Interpretation of Reliability Indices

The reliability indices reported here can serve to assist clinicians and researchers in selecting and interpreting performance-based clinical tests.<sup>38</sup> ICCs provide an indication of how confident users should be in their measurements, namely, what percentage of the observed score variance is due to true score variance versus that attributable to error. All of the studied tests were found to have interrater and test-retest ICCs between  $0.98$  and  $0.99$  and between  $0.88$  and  $0.97$ , respectively (Tab. 2 and 3). This suggests that 98% to 99% of the observed score variance between raters and 88% to 97% of the observed score variance between test sessions is due to true score variance, with the remaining observed score variance attributable to error (ie, 1%–12%). Trained users familiar with these tests can therefore be confident in the individual measurements made with these performance-based tests. Beyond establishing a level of confidence in an instrument, the magnitude of a test-retest ICC also suggests whether a test may be suitable for group-level comparisons (ie, ICC  $\geq 0.70$ ),<sup>25–28</sup> and/or individual-level applications (ie, ICC  $\geq 0.90$ ).<sup>19,25,28</sup> With the exception of the TUG, studied tests (ie, FSST, NBWT, and 10MWT) had test-retest ICCs  $\geq .90$ , indicating that they are each suitable for group as well as individual-level applications. Because the TUG test-retest ICC ( $.88$ ) is below, albeit slightly, the recommended threshold of  $0.9$ , users should be aware that this test has greater error relative to the other performance tests studied here and may therefore be better suited to group-level comparisons rather than making decisions about individual patients. This does not, however, suggest that the TUG lacks reliability altogether but simply that the observed level of reliability in this study warrants caution when using the TUG for individual applications like evaluating an individual LLP user over time.

An additional consideration when interpreting test-retest ICC values is the ICC model used to estimate reliability. Model selection, which is governed both by study design and whether systematic error is included in the calculation of the reliability coefficient,<sup>69</sup> determines if reliability results can be generalized beyond the participants and

raters included in the study.<sup>38</sup> Consistent with most reproducibility studies conducted to date,<sup>33,34,36</sup> test-retest ICCs in the present study were calculated using ICC (2,1), which includes both random and systematic error.<sup>22,38</sup> Model 2 permits the generalization of reliability results to other raters with similar participants or patients.<sup>38</sup> The reliability results of the present study can therefore be generalized to other unilateral LLP users and raters.

The SEM and MDC of the studied tests provide clinicians with indices they can use to determine precision of their patients' scores or determine whether changes have occurred in their patients.<sup>29,38</sup> The SEM provides an estimate of precision in a score. The CI derived with the SEM can be used, for example, in cross-sectional applications to assess an individual's score relative to a cutoff score (eg, a threshold established to indicate those at risk for falls). In contrast, MDC provides an index for use in longitudinal applications. MDC can be used to assess whether change over time in an individual patient's performance (ie, repeated assessments)<sup>70,71</sup> is above and beyond changes that may be attributable to measurement error.<sup>31</sup> Any change in a participant's score, either above or below the previous score, greater than the MDC is considered a real change. For example, if a LLP user recorded an initial NBWT score of 0.38, completed a 4-week balance-training program, and was then reevaluated on the NBWT with a score of 0.55, the difference would exceed the reported MDC<sub>90</sub> of .16, indicating that a real change in balance ability had been observed.

### Limitations

The present study only included established unilateral LLP users who had been using a prosthesis for at least 1 year (mean  $\pm$  SD: 16  $\pm$  15 years). Additional research is required to determine whether these results extend to acute LLP users (ie, <1 year post amputation), bilateral LLP users, or other populations. Similarly, additional research focusing on individuals with dysvascular amputations or individuals in specific age categories (eg, the US Medicare population) may be required. A larger sample size would be required to determine whether reliability and validity indices vary between subgroups of participants (ie, LLP users with different levels of amputation, causes of amputation, or mobility levels).

The present study only evaluated the reproducibility of 4 performance-based clinical tests. Given the population-specific nature of validity and reliability indices, future research is required to evaluate other psychometric properties of these tests (eg, minimally important clinical difference) as well as those of other performance-based tests that have proven useful in other patient populations (eg, Functional Reach Test, Fullerton Advanced Balance Scale, Five-times Sit to Stand, etc).

Significant session-to-session differences were observed in the TUG. Interestingly, the direction of these differences (ie, time decreased, performance improved) differed from those observed in a previous study, which found differences of a similar magnitude but in the opposite direction (ie, times increased, performance decreased).<sup>33</sup> This difference may represent the natural variation in TUG performance among LLP users. It is also possible that differences in sample characteristics could contribute. For example, there appear to be differences in age (55 vs 66 years), level of amputation (25% vs 50% transfemoral), the length of the test-retest period, and potentially cause of amputation between the 2 studies. Additional research is, however, required to verify these between-session differences and offer potential explanations. Regardless, these "time effects" suggest that practice effects may persist between test sessions for the TUG when it is applied with LLP users.<sup>38,72,73</sup> While not evaluated in the present study, within-session trial-to-trial practice effects may also exist among a host of performance-based clinical tests. These practice effects may limit not only the reliability of performance-based tests but also their validity (ie, diagnostic accuracy).<sup>73</sup> Previous research on the NBWT has shown that changes to administration and scoring procedures (ie, providing additional practice trials) can accommodate practice effects and improve psychometric performance.<sup>73</sup>

It may therefore be prudent to examine performance-based clinical test administration and scoring procedures to determine if they are susceptible to practice effects or other sources of systematic error.<sup>38</sup> If they are, modifications to administration and scoring procedures may prove effective in accommodating within- and between-session practice effects.

### Conclusion

Results of this study provide valuable evidence of interrater and test-retest reliability, measurement error, and MDC for performance-based clinical tests among established unilateral LLP users. These reliability metrics provide clinicians with actionable evidence they can use to select tests for use with individual patients, interpret the resultant test scores, and effectively use the test results to make informed clinical decisions. All studied tests demonstrated sufficient reliability to be considered appropriate for group-level comparisons (ie, ICC  $\geq$  0.70), while the FSST, 10MWT, and NBWT were also found to be suitable for individual-level applications among established unilateral LLP users (ie, ICC  $\geq$  0.90). SEM and MDC estimates of the studied tests provide clinicians and researchers with the information necessary to assess differences or changes in balance ability and/or fall risk among individual LLP users. In conjunction with published validity indices,<sup>53,74</sup> clinicians and researchers can use the calculated values to better select these clinical tests for LLP users and more effectively interpret the

resulting scores to make informed, evidence-based clinical decisions.

## Author Contributions

Concept/idea/research design: A. Sawers, B.J. Hafner  
Writing: A. Sawers  
Data collection: A. Sawers, J. Kim, G.S. Balkman, B.J. Hafner  
Data analysis: A. Sawers, B.J. Hafner  
Project management: A. Sawers  
Fund procurement: A. Sawers  
Providing participants: A. Sawers, B.J. Hafner  
Providing facilities/equipment: A. Sawers, B.J. Hafner  
Consultation (including review of manuscript before submitting): B.J. Hafner, J. Kim, G.S. Balkman

## Ethics Approval

All study procedures were reviewed and approved by the University of Illinois at Chicago and a University of Washington institutional review board.

## Funding

Research reported in this publication was supported by the US Department of Defense under award number W81XWH-17-1-0547 (A.S.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Department of Defense.

## Disclosures

The authors completed the ICMJE Form for Disclosure of Potential Conflicts of Interest and reported no conflicts of interest. B.J. Hafner reported previously serving as a paid consultant on integration of outcome measure in clinical practices for Medical Media Technologies, LLC.

DOI: 10.1093/ptj/pzaa063

## References

- Heinemann AW, Connelly L, Ehrlich-Jones L, Fatone S. Outcome instruments for prosthetics: clinical applications. *Phys Med Rehabil Clin N Am*. 2014;25:179–198.
- Wedge FM, Braswell-Christy J, Brown CJ, Foley KT, Graham C, Shaw S. Factors influencing the use of outcome measures in physical therapy practice. *Physiother Theory Pract*. 2012;28:119–133.
- Jette DU, Halbert J, Iverson C, Miceli E. Use of standardized outcome measures in physical therapist practice: perceptions and applications. *Phys Ther*. 2009;89:125–135.
- Bowman J. Challenges to measuring outcomes in occupational therapy: a qualitative focus group study. *Br J Occup Ther*. 2006;69:464–472.
- 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;12:96–104.
- Gaunaud 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;39:314–321.
- Hafner BJ, Spaulding SE, Salem R, Morgan SJ, Gaunaud I, Gailey R. Prosthetists' perceptions and use of outcome measures in clinical practice: long-term effects of focused continuing education. *Prosthet Orthot Int*. 2016;41:1–8.
- Stapleton T, McBrearty C. Use of standardized assessments and outcome measures among a sample of Irish occupational therapists working with adults with physical disabilities. *Br J Occup Ther*. 2009;72:55–64.
- Belazi D, Goldfarb NI, He H. Measuring health-related quality of life in the clinical setting. *Expert Rev Pharmacoecon Outcomes Res*. 2014;2:109–117.
- 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*. 2019;11:344–363.
- Kulkarni J, Wright S, Toole C, Morris J, Hirons R. Falls in patients with lower limb amputations: prevalence and contributing factors. *Physiotherapy*. 1996;82:130–136.
- 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;80:706–713.
- 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;47:80–86.
- 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;82:1031–1037.
- 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;48:80–85.
- 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;41:564–570.
- Chihuri S, Wong CK. Factors associated with the likelihood of fall-related injury among people with lower limb loss. *Inj Prev*. 2018;5:1–8.
- Roach KE. Measurement of health outcomes: reliability, validity and responsiveness. *J Prosthet Orthot*. 2006;18:8–12.
- Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med*. 2000;30:1–15.
- Kottner J, Audigé L, Brorson S, et al. Guidelines for reporting reliability and agreement studies (GRRAS). *J Clin Epidemiol*. 2011;64:96–106.
- Balk EM, Gazula A, Markozannes G, et al. Lower limb prostheses: measurement instruments, comparison of component effects by subgroups, and long-term outcomes. Comparative effectiveness review no. 213. AHRQ Publication No.18-EHC017-EF. Rockville, MD, USA: Agency for Healthcare Research and Quality; 2018.
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420–428.
- Hayen A, Dennis RJ, Finch CF. Determining the intra- and inter-observer reliability of screening tools used in sports injury research. *J Sci Med Sport*. 2007;10:201–210.
- Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A Practical Guide to Their Development and Use*. 5th ed. Oxford, England, UK: Oxford University Press; 2014.
- Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess*. 1998;2:1–74.
- Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34–42.
- Reeve BB, Wyrwich KW, Wu AW, et al. ISOQOL recommends minimum standards for patient-reported

- outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res*. 2013;22:1889–1905.
- 28 Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health*. 2007;10:S94–S105.
- 29 Harvill LM. Standard error of measurement. *Educ Meas*. 1991;10:33–41.
- 30 Vincent WJ, Weir JP. *Statistics in Kinesiology*. Champaign, IL, USA: Human Kinetics; 2012.
- 31 Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol*. 1999;52:861–873.
- 32 Stratford PW, Binkley J, Solomon P, Finch E, Gill C, Moreland J. Defining the minimum level of detectable change for the Roland-Morris questionnaire. *Phys Ther*. 1996;76:359–365.
- 33 Resnik L, Borgia M. Reliability of outcome measures for people with lower-limb amputations: distinguishing true change from statistical error. *Phys Ther*. 2011;91:555–565.
- 34 Jayakaran P, Johnson GM, Sullivan SJ. Reliability and concurrent validity of the step quick turn test in older persons with a unilateral transtibial amputation. *Am J Phys Med Rehabil*. 2011;90:798–804.
- 35 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;80:825–828.
- 36 Clemens SM, Gailey RS, Bennett CL, Pasquina PF, Kirk-Sanchez NJ, Gaunaud 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;32:388–397.
- 37 Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19:539–549.
- 38 Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19:231–240.
- 39 Sawers A, Hafner B. Validation of the Narrowing Beam Walking Test in lower limb prosthesis users. *Arch Phys Med Rehabil*. 2018;99:1491–1498.
- 40 Podsiadlo D, Richardson S. The Timed “Up & Go”: a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc*. 1991;39:142–148.
- 41 Dite W, Temple VA. A clinical test of stepping and change of direction to identify multiple falling older adults. *Arch Phys Med Rehabil*. 2002;83:1566–1571.
- 42 Bohannon RW. Comfortable and maximum walking speed of adults aged 20–79 years: reference values and determinants. *Age Ageing*. 1997;26:15–19.
- 43 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:e296.
- 44 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;42:377–381.
- 45 Doros G, Lew R. Design based on intra-class correlation coefficients. *Am J Biostat*. 2010;1:1–8.
- 46 Fewtrell MS, Kennedy K, Singhal A, et al. How much loss to follow-up is acceptable in long-term randomized trials and prospective studies? *Arch Dis Child*. 2008;93:458–461.
- 47 Dumville JC, Torgerson DJ, Hewitt CE. Reporting attrition in randomized controlled trials. *BMJ*. 2006;332:969–971.
- 48 Palmento Government Benefits Administrators. *Lower limb prostheses. DMERC Medicare Advis December*. 1994: 99–105.
- 49 Hafner BJ, Gaunaud 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;98:277–285.
- 50 Chaudhry S, Jin L, Meltzer D. Use of a self-report-generated Charlson Comorbidity Index for predicting mortality. *Med Care*. 2005;43:607–615.
- 51 Hanspal RS, Fisher K, Nieveen R. Prosthetic socket fit comfort score. *Disabil Rehabil*. 2003;25:1278–1280.
- 52 Hafner BJ, Morgan SJ, Askew RL, Salem R. Psychometric evaluation of self-report outcome measures for prosthetic applications. *J Rehabil Res Dev*. 2016;53:797–812.
- 53 Dite W, Connor HJ, Curtis HC. Clinical identification of multiple fall risk early after unilateral transtibial amputation. *Arch Phys Med Rehabil*. 2007;88:109–114.
- 54 Deathe AB, Miller WC. The L test of functional mobility: measurement properties of a modified version of the Timed “Up & Go” test designed for people with lower-limb amputations. *Phys Ther*. 2005;85:626–635.
- 55 Nanda R, Gupta S, Kanapathipillai P, Liow RYL, Rangan A. An assessment of the inter examiner reliability of clinical tests for subacromial impingement and rotator cuff integrity. *Eur J Orthop Surg Traumatol*. 2008;18:495–500.
- 56 Defloor T, Schoonhoven L. Inter-rater reliability of the EPUAP pressure ulcer classification system using photographs. *J Clin Nurs*. 2004;13:952–959.
- 57 Paulsen A, Overgaard S, Lauritsen JM. Quality of data entry using single entry, double entry, and automated forms processing: an example based on a study of patient-reported outcomes. *PLoS One*. 2012;7:e35087.
- 58 Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika*. 1965;52:591–611.
- 59 Stratford PW, Goldsmith CH. Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Phys Ther*. 1997;77:745–750.
- 60 McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1:30–46.
- 61 Nunnally JC, Bernstein IH. *Psychometric Theory*. New York, NY USA: McGraw-Hill; 1994.
- 62 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;85:723–729.
- 63 Ziegler-Graham K, MacKenzie EJ, Ephraim PL, Trivison TG, Brookmeyer R. Estimating the prevalence of limb loss in the United States: 2005 to 2050. *Arch Phys Med Rehabil*. 2008;89:422–429.
- 64 Ephraim PL, Wegener ST, MacKenzie EJ, Dillingham TR, Pezzin LE. Phantom pain, residual limb pain, and back pain in amputees: results of a national survey. *Arch Phys Med Rehabil*. 2005;86:1910–1919.
- 65 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;97:782–788.
- 66 Boonstra AM, Fidler V, Eisma WH. Walking speed of normal subjects and amputees: aspects of validity of gait analysis. *Prosthet Orthot Int*. 1993;17:78–82.
- 67 Moore M, Barker K. The validity and reliability of the four square step test in different adult populations: a systematic review. *Syst Rev*. 2017;6:1–9.
- 68 Beninato M, Portney LG. Applying concepts of responsiveness to patient management in neurologic physical therapy. *J Neurol Phys Ther*. 2011;35:75–81.

- 69 Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15:155–163.
- 70 Ottenbacher KJ, Johnson MB, Hojem M. The significance of clinical change and clinical change of significance: issues and methods. *Am J Occup Ther.* 1988;42:156–163.
- 71 Schmitt JS, Di Fabio RP. Reliable change and minimum important difference (MID) proportions facilitated group responsiveness comparisons using individual threshold criteria. *J Clin Epidemiol.* 2004;57:1008–1018.
- 72 Sawers A, Hafner BJ. Narrowing beam-walking is a clinically feasible approach for assessing balance ability in lower-limb prosthesis users. *J Rehabil Med.* 2018;50: 457–464.
- 73 Sawers A, Hafner BJ. Conventional administration and scoring procedures suppress the diagnostic accuracy of a performance-based test designed to assess balance ability in lower limb prosthesis users. *Prosthet Orthot Int.* 2019;43:402–408.
- 74 Sawers A, Hafner BJ. Using clinical balance tests to assess fall risk among established unilateral lower limb prosthesis users: cutoff scores and associated validity indices. *PM R.* 2020;12: 16–25.

## Appendix.

### Timed “Up and Go” (TUG) Protocol

Explanation to participant: The goal of this test is to rise from the chair, walk around the cone, walk back to the chair, and sit down again. You should walk at your normal, comfortable pace. I will time you while you perform the test.

Demonstration: Demonstrate the test 1 time.

Test instructions: Begin the test sitting with your back against the back of the chair and your arms resting on the armrests. When I say go, please stand up and walk around the cone, walk back to the chair, and sit down again. Please walk at your normal, comfortable pace.

Practice: Administer 1 practice trial. Do not time the practice trial.

Administration: Administer the test 2 times. Begin timing when you say go. Stop timing when the participants’ buttocks touch the chair.

Scoring: Select faster of the 2 timed trials as the TUG score.

### Four Square Step Test (FSST) Protocol

Explanation to participant: The goal of this test is to step over the canes in a specific sequence as quickly as possible. I will time you while you perform this test.

Demonstration: Demonstrate the test 1 time. Demonstrate starting in square 1 and stepping in squares 2, 3, 4, 1, 4, 3, 2, and 1.

Test instructions: When I say go, please step in the sequence I demonstrated. Try to complete the sequence as fast as possible without touching the canes. Both feet must make contact with the floor in each square. Face forward during the entire sequence.

Practice: Administer 1 practice trial. Do not time the practice trial.

Administration: Administer the test 2 times. Begin timing when the first foot contacts square 2. Stop timing when the last foot contacts square 1. Repeat the trial if the participant does not complete the sequence, loses balance, or contacts a cane.

Scoring: Select the faster of the 2 timed trials as the FSST score.

### Narrowing Beam Walking Test (NBWT) Protocol

Explanation to participant: The goal of this test is to walk as far as possible along the beam. Speed is not being evaluated. Begin the test by standing with 1 foot on the wide end of the beam and the other foot on the ground to the side. You may choose which foot to put on the beam and which to put on the ground. Please cross both your arms across your chest.

Demonstration: Demonstrate the test 1 time.

Test instructions: When I say go, please walk along the beam as far as you can. Please walk at a comfortable speed. Remember to keep your arms crossed over your chest as you walk. Once you move your arms away from your body or step off the beam, I will ask you to stop.

Practice: Do not administer a practice trial.

Administration: Administer the test 5 times. Stop the trial when a participant walks the length of the beam, steps off the beam, or moves their arms away from their body.

Scoring: Average the distances walked during trials 3 through 5, and divide by 22. This normalized distance (ie, 0–1) is recorded as the NBWT score. The average distance is divided by 22 and not 24 (the total length of the beam), because participants begin the test with 1 foot on the beam and must walk past the 0.61 m (2.0 ft) mark to receive a score above 0.0.

### 10 Meter Walk Test (10MWT) Protocol

Explanation to participant: The goal of this test is to walk a short distance at your preferred comfortable walking speed. I will time you while you perform this test.

Demonstration: Demonstrate the test 1 time.

Test instructions: When I say go, please walk at your normal, comfortable pace until I say stop.

Practice: Administer 1 practice trial.

Administration: Prepare a 14-m walkway in a hallway or other unobstructed area. Place lines at 0, 2, 12, and 14 m. Have the participant start in a standing position on the 0-m line. Inform them that, on the word “go,” they are to walk at a comfortable speed until you say “stop.” Begin timing when the participant crosses the 2-m line. Stop timing when the participant crosses the 12-m line. Inform the participant to stop when they cross the 14-m line. Repeat the test 2 times.

Scoring: Select the faster of the 2 timed trials as the TUG score.

Original Research

# Characterizing Practice Effects in Performance-Based Tests Administered to Users of Unilateral Lower Limb Prostheses: A Preliminary Study

Andrew Sawers, CPO, PhD , Brian J. Hafner, PhD

---

## Abstract

**Background:** Performance-based tests are viewed as a gold standard for measuring physical capability. Practice effects, however, may threaten their predictive, discriminative, and evaluative applications. Despite these potential consequences, practice effects have received limited attention in users of lower limb prostheses (LLP).

**Objective:** To perform an initial characterization of the occurrence, time-course, and magnitude of practice effects in three performance-based tests administered to users of LLP.

**Design:** Secondary analysis of data from a multisite repeated-measures study.

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

**Participants:** Convenience sample of established ambulatory users of unilateral transtibial and transfemoral prostheses ( $n = 60$ ).

**Intervention:** Not applicable.

**Main Outcome Measures:** Practice effects were identified as significant changes in slope of participants' cumulative trial-by-trial records. The occurrence, time-course, and magnitude of practice effects were computed for the Timed Up and Go (TUG), Four Square Step Test (FSST), and the 10-m Walk Test (10mWT).

**Results:** Across tests, practice effects were observed in 45% to 76% of participants. The proportion of participants with practice effects (ie, occurrence) was significantly greater for the FSST than the 10mWT ( $P = .008$ ). The median number of trials (ie, time-course) required for participants to reach a consistent level of performance was not significantly different between tests (FSST: 4 trials, TUG: 4 trials; 10mWT: 3.5 trials;  $P = .481$ ). Practice effect magnitude (ie, difference between the mean of trials during the plateau and best performance over the first two trials) was significantly greater than zero for the FSST (1.6 s; 16%) and TUG (1.4 s; 13%) ( $P < .05$ ).

**Conclusion:** Results indicate that the FSST, TUG, and 10mWT were susceptible to practice effects in this sample of users of LLP. Practice effects may obscure significant differences in walking and balance ability, and thereby, in the absence of modifications, limit the use of these tests for making individual patient clinical decisions and analyzing group-level data.

---

## Introduction

Performance-based tests are often the principal means by which clinicians estimate users' potential to use a lower limb prosthesis (LLP), classify users of LLP by functional level to determine their candidacy for specific prosthetic componentry, and assess outcomes associated with use of a prosthesis or other rehabilitation interventions to determine their effectiveness.<sup>1,2</sup> Performance-based tests are also important to researchers because they act as primary outcomes in clinical trials.<sup>3</sup> The accuracy with which

performance-based tests can be applied for predictive, discriminative, and evaluative purposes among users of LLP, however, may be threatened by their susceptibility to practice effects.<sup>4,5</sup>

Practice effects are systematic changes in performance that occur between repeated assessments of a test in absence of any intervention.<sup>6,7</sup> Practice effects may confound changes in performance because of treatment or intervention with changes in performance due to the tested individual learning or adapting to the test.<sup>8</sup> Results may therefore overestimate an intervention's benefit or

obscure the true extent of a person's recovery following a health event.<sup>9</sup> For example, despite randomizing the order in which two prosthetic feet were tested, a recent comparative effectiveness study found that participants' performance generally improved during the second test. The authors attributed this result to a practice effect.<sup>10</sup> Additionally, practice effects were found to suppress the diagnostic accuracy of the Narrowing Beam Walking Test, a performance-based balance test designed to predict fall risk in users of LLP.<sup>5</sup> Suppression of diagnostic accuracy due to practice effects would serve to reduce the ability of clinicians to identify users of LLP at greatest risk of falling. It is possible that practice effects could also be present in other clinical tests frequently administered to users of LLP, including the Timed Up and Go (TUG), Four Square Step Test (FSST), and 10-m Walk Test (10mWT). These tests are often used in clinical practice and research studies because they can be administered in less than 5 minutes with minimal equipment and possess evidence of validity and reliability in users of LLP.<sup>11-16</sup> If present, practice effects may exert a substantial influence on how results from these tests are interpreted, affecting not only individual clinical patient measurements but also the data collected in research studies that are used to determine reimbursement policies for prosthetic componentry.

Despite these potential consequences, practice effects in performance-based tests administered to users of LLP have received limited attention. Only two studies have explicitly sought to quantify practice effects in performance-based tests administered to users of LLP. Kristensen et al, found no evidence of within-session practice effects over five trials of the single-limb stance test, although 44% of participants achieved their best performance during the last trial.<sup>17</sup> In contrast, Sawers and Hafner reported that practice effects were observed in nearly 40% of participants performing a narrowing beam walking task.<sup>18</sup> Differences in findings between these two studies may be attributed to the types of tasks studied (ie, static standing balance vs dynamic walking balance) but may also be because of the mode of analysis. Kristensen et al employed a repeated-measures group-level analysis, rather than examining individual participant's performance.<sup>17</sup> Conversely, Sawers and Hafner examined each participant's trial-by-trial record to assess presence of practice effects.<sup>18,19</sup> Additional evidence of potential practice effects in performance-based tests administered to users of LLP can be inferred from evaluations of test reliability. Although not universal, a number of test-retest reliability studies have, in addition to large intraclass correlation coefficients (ICCs), reported statistically significant improvements in performance between test sessions.<sup>16,20,21</sup> Evidence of reliability may therefore not be sufficient proof that practice effects are not present, as large ICCs can be obtained if most or all participants change in the same direction (ie, improve performance) across test sessions.<sup>7,22,23</sup>

Additionally, it remains unknown whether the level of performance achieved on a test when practice effects have dissipated (ie, someone's stable performance) differs significantly from how performance on a test would otherwise be scored (ie, someone's performance while they are learning the test). In the absence of such data the potential clinical and scientific consequences associated with practice effects remain largely unexplored. The current literature therefore offers little guidance on the issue of practice effects among users of LLP.

The objective of this study was therefore to perform an initial characterization of the occurrence, time-course, and magnitude of practice effects in three performance-based tests routinely administered to LLP users. We hypothesized that unilateral users of LLP would exhibit practice effects on the TUG, FSST, and 10mMWT.

## Methods

### Study Design

A secondary analysis of data from a multisite repeated-measures study (Chicago and Seattle) was conducted to characterize practice effects in the TUG,<sup>24</sup> FSST,<sup>25</sup> and 10mWT<sup>26</sup> when administered to users of LLP.<sup>16</sup> Original study protocols were reviewed and approved by institutional review boards at each site. All individuals provided written informed consent prior to participation.

### Participants

Users of LLP were recruited from local prosthetic clinics in Chicago and Seattle using convenience sampling. Inclusion criteria were 18 years of age or older; unilateral amputation at the transtibial or transfemoral level; amputation due to trauma, dysvascular complications, tumor, or infection; use of a prosthesis for at least 1 year; use of a comfortable prosthesis to ambulate; able to walk a short distance (eg, 10 m) without an upper extremity mobility aid (eg, cane); and able to read, write, and speak English. Exclusion criteria were amputation of a second limb (eg, arm or contralateral leg); contralateral complications (eg, hip replacement, ulcers, or infections), neurologic disorder; or severe cardiopulmonary disease that would prevent the individual from completing the study protocol.

### Procedures

Participants were administered demographic, amputation, and mobility characterization measures and one of three performance-based tests. Tests were randomly assigned using block randomization. Participants were scheduled to return approximately 1 week later for a second test session, at which time the same performance-based test was readministered. This test-retest interval was chosen to minimize familiarity with the tests and to avoid the

possibility of changes in health, prosthetic proscriptio, or prosthetic fit. Participants' prosthesis and desired footwear were held constant between test sessions. Upon returning to the lab for the second test session, participants were asked if there had been any change in their prosthesis and/or health since the first test session.

## Measurements

### *Participant Demographic, Amputation, and Mobility Characterization Measures*

Sociodemographic data including age, height, weight, and gender and amputation characteristics (eg, level, etiology, and time since amputation) were collected through self-report and standardized interview. Mobility-related outcomes were quantified using the Prosthetic Limb User Survey - Mobility (PLUS-M),<sup>27</sup> Medicare Functional Classification Level (MFCL),<sup>28</sup> and typical hours of prosthesis use per day. Participants also completed the Socket Comfort Score<sup>29</sup> at the start of each test session.

### *Administration of Performance-Based Clinical Tests*

One of three performance tests was randomly selected and administered to each participant; the TUG, designed to evaluate functional mobility,<sup>24</sup> the FSST, designed to assess dynamic balance,<sup>25</sup> or the 10mWT, designed to measure gait speed.<sup>26</sup> Only one test was administered to each participant to mitigate effects of fatigue. The candidate tests were chosen because of their minimal administrative burden and evidence of psychometric properties in users of LLP.<sup>11-16</sup> During each test session, and following a single unscored practice trial, the same trained rater administered 10 trials of the selected test to each participant according to standardized instructions and protocols (Appendix S1).<sup>24-26</sup> Ten trials were administered based on previous research that found performance by users of LLP on a narrowing beam-walk task reached a plateau within 10 trials.<sup>18</sup> One-minute rest breaks were provided between trials to minimize risk of fatigue. Acu-Split A610X stopwatches (Pleasanton, CA) were used to record participants' time taken to complete each trial. Time was used to quantify FSST and TUG performance, and 10mWT performance was quantified via gait speed (m/s).

## Data Analysis

### *Occurrence, Time-Course, and Magnitude of Practice Effects*

Practice effects were identified as statistically significant changes in the slope of each participant's cumulative trial-by-trial record.<sup>18,19,30</sup> The occurrence of practice effects for each test during each session was quantified as the percentage of participants who experienced one or more significant changes in the slope of their cumulative 10-trial record.<sup>18,19,31</sup> An occurrence threshold of 15% was designated as that above which a test

would be considered susceptible to practice effects. The threshold of 15% was selected owing to its use in identifying ceiling and floor effects in health instruments.<sup>32,33</sup>

The time-course of practice effects was quantified as the number of trials required for participants' performance to plateau (ie, reach a consistent level indicative of their stable performance on a test).<sup>18,19</sup> Each participant's performance was considered consistent once the cumulative trial-by-trial record within a test session reached a terminal slope (ie, the participant exhibited no further changes in performance between trials).<sup>18,19,30</sup> Performance was considered consistent across all trials (ie, no practice effects) if no changes in slope were identified in the cumulative trial-to-trial data record. The number of trials required for 80% and 100% of participants to reach a stable and consistent level of performance on each test was determined for both sessions.

Participants' stable performance was estimated as the mean of trials during the plateau period. Performance was also scored as recommended by the test developers (ie, participants' best performance [ie, fastest time] across the first two scored trials following a demonstration and practice trial).<sup>11,24-26</sup> The magnitude of practice effects was quantified as the difference between best performance over the first two scored trials and stable performance. The magnitudes of practice effects were also expressed as a percent change from scores based on the best of the first two score trials. Shapiro-Wilk's tests were used to assess the normality of the data. Based on the results of the Shapiro-Wilk's tests (ie, normal or non-normally distributed), parametric or non-parametric tests were run to test for differences in occurrence, time-course, and practice effect magnitude between tests and between sessions. All statistical tests were conducted using SPSS (V.24; SPSS, Inc., Chicago, IL), with an alpha value of .05.

## Results

Sixty users of unilateral LLP were recruited and randomized to one of three performance-based tests, the FSST (n = 21), TUG (n = 19), and 10mWT (n = 20). Demographic, amputation-, and mobility-related characteristics of the study sample, stratified by test, are presented in Table 1. Among all three of the test groups age and PLUS-M T-scores were both normally distributed ( $W \geq 0.952$ ,  $P \leq .274$ ), whereas time since amputation was non-normally distributed ( $W \leq .815$ ,  $P \leq .03$ ) (Table 1). A one-way analysis of variance revealed that neither age nor PLUS-M T-scores were significantly different between the FSST, TUG, and 10mWT groups (Age:  $F(2, 57) = .340$ ,  $P = 0.713$ ; PLUS-M:  $F(2, 57) = .579$ ,  $P = .564$ ). In contrast, a Kruskal-Wallis test found that time since amputation was significantly different between test groups ( $H[2] = 7.51$ ,  $P = .023$ ). Pairwise comparisons performed with a Bonferroni correction for multiple comparisons revealed that time since amputation was significantly greater among the 10mWT group than the

**Table 1**

Demographic, amputation, and activity characteristics stratified by test group

| Test              | Demographic Characteristics   |                      |                          |                         | Amputation Characteristics |                                                                 |                                                  | Activity Characteristics      |                              |
|-------------------|-------------------------------|----------------------|--------------------------|-------------------------|----------------------------|-----------------------------------------------------------------|--------------------------------------------------|-------------------------------|------------------------------|
|                   | Gender<br>Participants<br>(n) | Age (y)<br>Mean (SD) | Weight (kg)<br>Mean (SD) | Height (m)<br>Mean (SD) | Level<br>Participants (n)  | Etiology<br>Participants (n)                                    | Time Since A<br>mputation (y)<br>Median (Q1, Q3) | PLUS-M (T-score)<br>Mean (SD) | MFCL<br># Participants       |
| FSST<br>(n = 21)  | Male (14)<br>Female (7)       | 56.7 (11.1)          | 86.1 (17.1)              | 1.74 (0.07)             | TT (16)<br>TF (5)          | Trauma (9)<br>Dysvascular (5)<br>Cancer (2)<br>Infection (5)    | 7 (2, 19)                                        | 55.2 (7.16)                   | K1(0) K2(6)<br>K3(13) K4(2)  |
| TUG<br>(n = 19)   | Male (15)<br>Female (4)       | 52.6 (17.8)          | 88.3 (21.6)              | 1.74 (0.09)             | TT (12)<br>TF (7)          | Trauma (7)<br>Dysvascular (4)<br>Cancer (3)<br>Infection (5)    | 8 (3, 16)*                                       | 55.9 (5.65)                   | K1(0) K2(6)<br>K3(11) K4(2)  |
| 10mWT<br>(n = 20) | Male (13)<br>Female (7)       | 55.7 (19.1)          | 79.4 (18.3)              | 1.74 (0.1)              | TT (17)<br>TF (3)          | Trauma (10)<br>Dysvascular (3)<br>Cancer (4)<br>Infection (3)   | 22 (11, 33)*                                     | 57.4 (7.80)                   | K1(0) K2(6)<br>K3(10) K4(4)  |
| Total             | Male (42)<br>Female (18)      | 55.1 (16.2)          | 84.6 (19.2)              | 1.74 (.09)              | TT (45)<br>TF (15)         | Trauma (26)<br>Dysvascular (12)<br>Cancer (9)<br>Infection (13) | 12 (4, 26)                                       | 56.1 (6.87)                   | K1(0) K2(18)<br>K3(34) K4(8) |

FSST, Four Square Step Test; MFCL, Medicare Functional Classification Level; PLUS-M, Prosthetic Limb User's Survey of Mobility; Q1, first quartile; Q3, third quartile; TF, Transfemoral; TT, Transtibial; TUG, Timed Up and Go; 10mWT, 10 m Walk Test.

\*Time since amputation was significantly greater for participants assigned to the 10mWT group than the TUG group ( $P \leq .05$ ).

TUG group ( $H[2] = 7.44$ ,  $P = .019$ ) (Table 1). Level of amputation (ie, transtibial vs transfemoral), cause of amputation (ie, dysvascular vs non-dysvascular), and MFCL (ie, K1 and K2 vs K3 and K4) were compared between the FSST, TUG, and 10mWT groups using 2xC chi-square tests of homogeneity. None of these categorical variables were found to differ between test groups ( $\chi^2[2] \leq 3.79$ ,  $P \geq .150$ ) (Table 1). Session 2 occurred at a mean of 6.7 days (SD = 2.3 days, minimum = 3 days, maximum = 9 days) after session 1. Study participants did not report any changes in health, prosthetic prescription, or prosthetic fit between sessions. A paired-samples  $t$ -test revealed that Socket Comfort scores were not significantly different between sessions (session 1: 7.62 (1.97), session 2: 7.53 (1.53),  $t(58) = .590$ ,  $P = .557$ ).

### Occurrence of Practice Effects on the FSST, TUG, and 10mWT

Practice effects (ie, significant changes in performance between consecutive trials) were observed in all administered tests and during both sessions (Figure 1). Within session 1, practice effects in the FSST, TUG, and 10mWT groups were observed among 76% (16/21), 68% (13/19), and 45% (9/20) of participants, respectively. In session 2, practice effects were observed in 57% (12/21), 73% (14/19), and 55% (11/20) of participants in the FSST, TUG, and 10mWT groups, respectively. Cochran's Q tests were run to determine if the occurrence of practice effects differed between performance tests within each session. The proportion

of participants who exhibited practice effects was significantly different between the FSST, TUG, and 10mWT groups during session 1 ( $\chi^2[2] = 10.57$ ,  $P = .003$ ) but not session 2 ( $\chi^2[2] = 4.67$ ,  $P = .222$ ). Post hoc McNemar exact tests revealed that during session 1 the proportion of participants with practice effects was significantly greater for the FSST group than the 10mWT group ( $P = .008$ ) but did not differ between the TUG and the 10mWT ( $P = .06$ ), or the TUG and FSST ( $P = .125$ ).

McNemar's tests were run to determine if the proportion of participants who exhibited practice effects differed between sessions. The proportion of participants who exhibited practice effects did not differ across test sessions for the FSST ( $\chi^2[1] = 2.25$ ,  $P = .125$ ), TUG ( $\chi^2[1] = 0.005$ ,  $P > .999$ ), or 10mWT groups ( $\chi^2[1] = .500$ ,  $P = .480$ ).

### Time-Course of Practice Effects on the FSST, TUG, and 10mWT

Performance plateaued for all participants, on each test, during both sessions (Figure 2). The time-course (ie, number of trials) for participants' performance to plateau was non-normally distributed for all tests across both sessions ( $W \leq 0.901$ ,  $P \leq .036$ ), except for the TUG during session 2 ( $W = 0.917$ ,  $P = .088$ ). Kruskal-Wallis tests were run to determine if the median number of trials required for performance to plateau differed between tests. The number of trials (median, Q1-Q3) required for participants to reach a

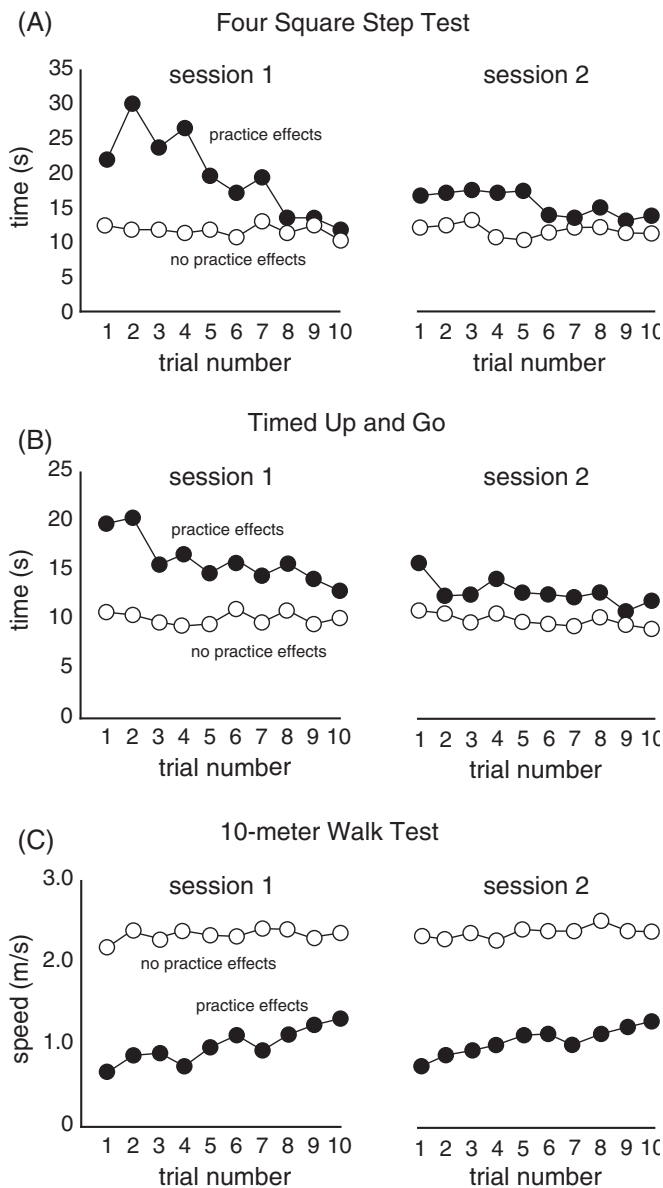


**Figure 1.** Trial-by-trial data from two participants with unilateral lower limb amputation who did (closed circles) and did not (open circles) exhibit practice effects across two test sessions of the Four Square Step Test (A), Timed Up and Go (B), and 10-m Walk Test (C). The 10 trials were preceded by a single unscored practice trial.

consistent level of performance was not significantly different between test groups during session 1 (FSST: 4.0, 2-6; TUG: 3.5, 0-5; 10mWT: 0, 0-6,  $H[2] = 1.46$ ,  $P = .482$ ) or session 2 (FSST: 2, 0-5; TUG: 4.0, 0-5; 10mWT: 3, 0-6;  $H[2] = 1.19$ ,  $P = .551$ ).

Wilcoxon signed-rank tests were run to determine if the number of trials required for performance to plateau differed between sessions 1 and 2. The median number of trials required for performance to plateau was not significantly different between session 1 and 2 for any test groups (FSST:  $z = -1.72$ ,  $P = .086$ ; TUG:  $z = -0.518$ ,  $P = .60$ ; 10mWT:  $z = 0.631$ ,  $P = .53$ ).

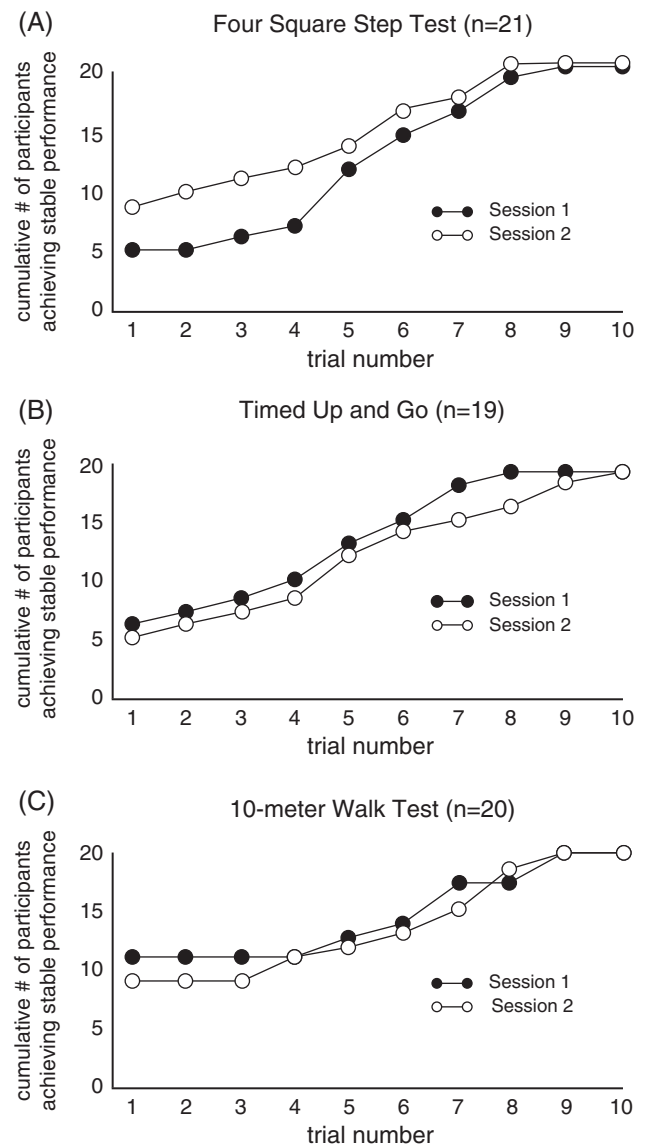


**Figure 2.** The time-course of performance stabilization on the Four Square Step Test (A), Timed Up and Go (B), and 10-m Walk Test (C) among participants with unilateral lower limb amputation, across two test sessions (session 1: closed circles, session 2: open circles). Performance stabilized for each participant, during both test sessions, on all three performance-based tests. For 80% of participants, performance stabilized within 6 to 8 trials depending on the test and session. The 10 trials were preceded by a single unscored practice trial.

### Magnitude of Practice Effects on the FSST, TUG, and 10mWT

The magnitude of the observed practice effects (ie, the difference between best of the first two scored trials and stable performance) was non-normally distributed for the FSST during session 1 and 2 (session 1:  $W = 0.784$ ,  $P = .001$ , session 2:  $W = 0.732$ ,  $P < .001$ ) as well as for the TUG ( $W = 0.713$ ,  $P < .001$ ) and 10mWT ( $W = 0.851$ ,  $P = .005$ ) during session 1 but not for session 2 ( $W = 0.926$ ,  $P = .132$  and  $W = 0.918$ ,  $P = .092$ ).

**Table 2**

Conventional and true performance test scores

|                                    | Session 1                                      |                       |                                            | Session 2                                      |                       |                                            |
|------------------------------------|------------------------------------------------|-----------------------|--------------------------------------------|------------------------------------------------|-----------------------|--------------------------------------------|
|                                    | Best of First<br>Two Trials<br>Median (Q1, Q3) | Stable<br>Performance | Practice Effect<br>Magnitude<br>Median (%) | Best of First<br>Two Trials<br>Median (Q1, Q3) | Stable<br>Performance | Practice Effect<br>Magnitude<br>Median (%) |
| Participants with practice effects |                                                |                       |                                            |                                                |                       |                                            |
| FSST<br>(time, s)                  | 9.7 (7.1,12)                                   | 8.1 (6.0,11)          | 1.6 (16%) <sup>†</sup>                     | 9.1 (6,12)                                     | 8.1 (5.5,10)          | 1.0 (11%)*                                 |
| TUG<br>(time, s)                   | 11 (8.9,12)                                    | 9.6 (8.6,12)          | 1.4 (13%)*                                 | 9.8 (8.7,12)                                   | 9.6 (8.3,12)          | 0.20 (2%)                                  |
| 10mWT<br>(speed, m/s)              | 1.2 (.98,1.5)                                  | 1.1 (1.0,1.5)         | 0.10 (8%)                                  | 1.1 (1.0,1.3)                                  | 1.1 (1.0,1.3)         | 0.00 (0%)                                  |
| All participants                   |                                                |                       |                                            |                                                |                       |                                            |
| FSST<br>(time, s)                  | 9.9 (8.4,12)                                   | 8.9 (7.5,11)          | 1.01 (10%) <sup>†</sup>                    | 9.3 (7.3,12)                                   | 8.7 (7.4,11.2)        | 0.55 (6%)                                  |
| TUG<br>(time, s)                   | 10.4 (8.6,12)                                  | 9.6 (8.5,11)          | 0.80 (8%) <sup>†</sup>                     | 9.6 (7.9,12)                                   | 9.4 (8.1,11)          | 0.20 (2%)                                  |
| 10mWT<br>(speed, m/s)              | 1.3 (1.0,1.5)                                  | 1.3 (1.1,1.5)         | 0.01 (1%)                                  | 1.2 (1.1,1.4)                                  | 1.3 (1.1,1.5)         | 0.01 (1%)                                  |

FSST, Four Square Step Test; m/s, meters per second; Q1, first quartile; Q3, third quartile; s, seconds; TUG, Timed Up and Go; 10mWT, 10 m Walk Test.

\*Practice effect magnitude is significantly greater than zero ( $P \leq .05$ ).<sup>†</sup>Practice effect magnitude is significantly greater than zero ( $P \leq .002$ ).

respectively). Wilcoxon signed-rank tests revealed that during session 1 the difference between the best of the first two scored trials and stable performance (ie, magnitude of practice effects) was significantly greater than zero among participants in the FSST group that exhibited practice effects (best of first two: 9.7 s; stable: 8.1 s,  $z = 3.52$ ,  $P < .001$ ) as well as the FSST group as a whole (ie, all participants) (best of first two: 9.9 s; stable: 8.9 s,  $z = 4.02$ ,  $P < .001$ ) (Table 2). During session 2, FSST practice effect magnitude was significantly greater than zero only among participants who exhibited practice effects (best of first two: 9.1 s; stable: 8.1 s,  $z = 2.12$ ,  $P = .044$ ) (Table 2). In the TUG group the difference between the best of first two scored trials and true performance was significantly greater than zero during session 1 for participants with practice effects (best of first two: 11 s; stable: 9.6 s,  $z = -2.41$ ,  $P = .016$ ) as well as the TUG group as a whole (best of first two: 10.4 s; stable: 9.6 s,  $z = 3.08$ ,  $P = .002$ ) (Table 2). Practice effect magnitude was not significantly greater than zero in the TUG group during session 2 ( $z \leq 0.597$ ,  $P \geq .50$ ) (Table 2). Practice effect magnitudes in the 10mWT group were not significantly greater than zero during session 1 or 2 ( $z \leq 1.78$ ,  $P \geq .075$ ) (Table 2).

## Discussion

The objective of this study was to perform an initial characterization of the occurrence, time-course, and magnitude of practice effects in three performance-based tests commonly administered to users of unilateral LLP. Results support the study hypothesis that the FSST, TUG, and 10mWT are susceptible to practice effects in

this clinical population. The occurrence of practice effects exceeded a threshold of 15% in each of the test groups, during both sessions. Further, the median number of trials required for participants to achieve a consistent level of performance exceeded the number of trials stipulated by recommended administration and scoring procedures (ie, two trials).<sup>11,13,24-26</sup> The best of the first two scored trials on the FSST, TUG, and 10mWT (Table 2) were consistent with values previously reported among comparable samples.<sup>11,13,34,35</sup> The difference between the best of the first two scored trials and stable performance in the FSST and TUG groups were significantly greater than zero not only among participants who exhibited practice effects but also among the groups as a whole. The persistence of practice effects (ie, mainly occurrence and time-course) across session 1 and 2 suggests that any consistency in performance achieved through practice within one session may be at least partially lost within as little as 1 week. Clinically, performance-based tests are often administered weeks or months apart. One may therefore anticipate varying degrees of erosion in performance acquired through previous administration over these different retest periods, further complicating comparisons over time. All participants did however achieve a consistent level of performance during both sessions, suggesting that revisions to administration and scoring procedures may be able to accommodate practice effects (ie, more trials and a modified scoring procedure).<sup>4,18,19</sup> These results indicate that among ambulatory, moderately active, established unilateral users of LLP, practice effects may obscure significant differences in balance and walking ability. Similar studies are required to determine whether practice

effects are prevalent among older, inexperienced, less active users of LLP as well as those participating in rehabilitation. We would speculate however, that such individuals are equally if not more likely to exhibit practice effects, as they may have lower initial test scores, which provide more room for improvement with repeated administration of a test.

### ***Occurrence of Practice Effects in Performance-Based Tests Administered to Users of Unilateral LLP***

Across each of the test groups and sessions, the occurrence of practice effects (ie, percentage of participants demonstrating practice effects) exceeded the 15% threshold that is generally considered evidence of possible psychometric issues (eg, a floor or ceiling effect).<sup>32,33</sup> Although no specific thresholds have been established to infer whether a clinical test is susceptible to practice effects, 45%-75% of study participants responding with significant changes in performance may be considered convincing initial evidence of practice effects in each of these tests.

Previous efforts to characterize practice effects in performance-based tests administered to people with lower limb amputation<sup>17</sup> or other health conditions<sup>22,36-38</sup> were unable to quantify the occurrence of practice effects (ie, percentage of participants who exhibited practice effects) because they adopted a repeated-measures group-level analysis, rather than examine each individual's trial-by-trial performance. Although a group-level analysis can provide insight into the overall magnitude of practice effects, it cannot provide information about how often practice effects are likely to affect test results. Further, a group-level analysis will likely obscure individual variations in the time-course of any practice effects, an important element to understanding and accommodating practice effects through administration and scoring procedures.<sup>9,23,39</sup> These results therefore represent an important first step in ascertaining how frequently practice effects may influence test results in users of LLP.

### ***Time-Course of Practice Effects in Performance-Based Tests Administered to Users of Unilateral LLP***

Quantifying the time-course of practice effects on a test is an important step toward decreasing the likelihood that test scores will over- or underestimate an individual's performance.<sup>4,18,19,23</sup> Efforts to establish the time-course of practice effects in performance-based tests administered to users of LLP have however been limited.<sup>17-20</sup> Kristensen et al reported that, despite no evidence of group-level practice effects on the single-leg stance test, 44% of users of unilateral LLP achieved their maximum standing time on the last of five trials,<sup>17</sup> suggesting that a consistent level of performance might not have been achieved in a portion of the sample

because a sufficient number of trials was not administered. In the current study, performance on all three tests stabilized but often only after having performed more trials than are traditionally administered in practice.<sup>24-26</sup> This suggests that administration and scoring procedures for these tests do not accommodate practice effects and may therefore underestimate walking and balance ability in users of unilateral LLP. This may cause clinicians and researchers to interpret post-intervention improvements in test performance as a true change in walking or balance ability, when some participants only "learned the test,"<sup>10</sup> with no change in physical capacity. When the practice effects are not fully accommodated, erroneous conclusions may therefore be drawn about the effects of treatment.<sup>40</sup>

Performance stabilized for all participants on all tests in the current study, suggesting that practice effects can be accommodated by administering trials until a plateau in performance is achieved.<sup>23</sup> Administering a sufficiently large number of trials so that a plateau in performance is reached may, however, be impractical and impose undue burden on clinicians as well as patients. A more practical solution may be to develop administration and scoring procedures that use as few trials to estimate later stable performance.<sup>4,18</sup> Such a procedure would reduce the number of trials that need to be performed in order to estimate stable performance, thereby limiting patient burden because of potential fatigue or pain that could arise with extended testing. This approach has been successfully implemented to accommodate practice effects in users of LLP on the Narrowing Beam Walking Test.<sup>18</sup> Importantly, these revisions to the administration and scoring procedures served to improved test validity.<sup>5</sup> Such an approach may, however, be warranted only if practice effects affect interpretation of test results (ie, diagnostic accuracy, minimal detectable change).

### ***Susceptibility of Performance-Based Tests to Practice Effects among Users of LLP***

Among the tests assessed in this preliminary study, the FSST appears to be more susceptible to practice effects than either the TUG or 10mWT. This was confirmed by the more frequent occurrence and larger magnitude of practice effects during session 1 as well as their persistence in session 2. This susceptibility to practice effects does not appear to be based on differences in amputation, mobility, or age between each of the test groups. The reasons for practice effects on any specific performance-based test, or why certain tests may be more susceptible to practice effects than others, remain largely unknown.<sup>18</sup> Additional research is therefore required to confirm these initial results in a larger sample and elucidate possible causes. Practice effects may be attributable to personal factors<sup>41,42</sup> or elements within the tests themselves. Understanding the basis for practice effects is likely to

help determine how best to accommodate them. Memory of previous performance, a desire to show improvement,<sup>20</sup> and movement strategy selection<sup>5</sup> may each play a role. These potential sources of practice effects may be best accommodated through revisions to test instructions. Alternatively, if specific motor tasks present in performance-based tests were consistently found to be susceptible to practice effects, they may only be accommodated by eliminating them from existing tests.

Lastly, many of the validity and reliability indices (eg, intraclass correlation coefficients, minimal detectable changes, cutoff thresholds) used to select, administer, and interpret tests in clinical practice have been derived using data collected according to recommended administration and scoring procedures (ie, the best of two trials). As these procedures do not appear to accommodate practice effects or reflect users of LLP stable performance ability on a test,<sup>5</sup> there is the potential that there could be errors in these indices of validity and reliability. Research to determine the impact of practice effects on these important clinical indices is therefore required.

## Limitations

The current results are those of a preliminary study, limited to the assessment of three performance-based tests, in a single clinical population, and among small samples. Additional studies, powered a priori, are therefore required to replicate and extend the current findings to other performance-based tests and populations. This may include an examination of whether practice effects vary with level and cause of amputation. The proportion of participants with transtibial and transfemoral amputations within the 10mWT and FSST groups was skewed towards a higher percentage of individuals with transtibial amputation compared to larger national studies,<sup>43-47</sup> potentially limiting the generalization of those results to users of transfemoral prostheses. There are several methodological choices that may have influenced the results (ie, increased or decreased practice effects). For example, more time between testing sessions may have increased practice effects during session two, whereas less time between sessions may have decreased practice effects. Similarly, "stable" performance may have increased or decreased had a larger or smaller number of trials been performed, thereby potentially increasing or decreasing practice effects. However, given the plateau observed in participants' trial-by-trial performance, additional trials are unlikely to have altered the characterization of stable performance in this sample. And finally, had "best" or "worst" performance been selected in lieu of "stable" performance, practice effects may also have increased or decreased. The choice of "best," "worst," or "stable" performance in the scoring of performance-based tests requires further examination. Finally, the collection of additional variables, including cognitive functioning, mood, effort, and

movement strategies, are needed in future studies to help discern the basis of practice effects and ultimately limit their impact on test scores and interpretation.

## Conclusion

Statistically significant differences in balance and walking ability appear to be obscured when the best of the first two scored trials are used with the FSST, TUG, and 10mWT among users of LLP. Practice effects may therefore hinder use of these tests for making individual patient clinical decisions (eg, monitoring change over time) and for analyzing group-level data (eg, evaluating effectiveness of an intervention in a clinical trial). Until more is known regarding the occurrence, time-course, and magnitude of practice effects in this clinical population, researchers and clinicians should be aware that these effects might influence interpretation of test scores. Future research is needed to confirm the current results in a larger sample, assess additional performance-based tests, and study practice effects across different time intervals (eg, 1 hour, day, week, month, and year). An evaluation of the clinical significance of practice effects by determining whether they influence test applications (ie, discrimination, prediction, and evaluation) and associated validity and reliability indices (eg, cutoff scores, likelihood ratios, and minimal detectable change) is central to establishing their impact on clinical care. Finally, establishing the psychomotor basis of practice effects to aid in the development of suitable modifications to accommodate them (eg, alternative test procedures, scoring system, alternative tests and/or tasks) will be an important next step.

## Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

## References

1. Gaunaud 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.
2. Borrenpohl D, Kaluf B, Major MJ. Survey of U.S. practitioners on the validity of the Medicare functional classification level system and utility of clinical outcome measures for aiding K-level assignment. *Arch Phys Med Rehabil*. 2016;97(7):1053-1063.
3. Wu G, Sanderson B, Bittner V. The 6-minute walk test: how important is the learning effect? *Am Heart J*. 2003;146(1):129-133.
4. Ashendorf L, Horwitz JE. Abbreviating the finger tapping test. *Arch Clin Neuropsychol*. 2015;30(2):99-104.
5. Sawers A, Hafner BJ. Conventional administration and scoring procedures suppress the diagnostic accuracy of a performance-based test designed to assess balance ability in lower limb prosthesis users. *Prosthet Orthot Int*. 2019;43(4):402-408.
6. Bartels C, Wegrzyn M, Wiedl A, Ackermann V, Ehrenreich H. Practice effects in healthy adults: a longitudinal study on frequent repetitive cognitive testing. *BMC Neurosci*. 2010;11(1):118.

7. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med.* 2000;30(1):1-15.
8. Portney LG, Watkins MP. *Foundations of Clinical Research.* Philadelphia, PA: Davis Company; 2013.
9. Del Rossi G, Malaguti A, Del Rossi S. Practice effects associated with repeated assessment of a clinical test of reaction time. *J Athl Train.* 2014;49(3):356-359.
10. Halsne EG, McDonald CL, Morgan SJ, Cheever SM, Hafner BJ. Assessment of low- and high-level task performance in people with transtibial amputation using crossover and energy-storing prosthetic feet: a pilot study. *Prosthet Orthot Int.* 2018;42(6):583-591.
11. 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.
12. 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.
13. Resnik L, Borgia M. Reliability of outcome measures for people with lower-limb amputations: distinguishing true change from statistical error. *Phys Therapy.* 2011;91(4):555-565.
14. Deathe AB, Miller WC. The L test of functional mobility: measurement properties of a modified version of the timed "up & go" test designed for people with lower-limb amputations. *Phys Therapy.* 2005;85(7):626-635.
15. 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.
16. Sawers A, Kim J, Balkman G, Hafner BJ. Interrater and test-retest reliability of performance-based clinical tests administered to established users of lower limb prostheses. *Phys Therapy.* 2020;100(7):1206-1216.
17. Kristensen MT, Nielsen AØ, Topp UM, et al. Number of test trials needed for performance stability and interrater reliability of the one leg stand test in patients with a major non-traumatic lower limb amputation. *Gait Posture.* 2014;39(1):424-429.
18. Sawers A, Hafner BJ. Narrowing beam-walking is a clinically feasible approach for assessing balance ability in lower-limb prosthesis users. *J Rehabil Med.* 2018;50(5):457-464.
19. Sawers A, Hafner BJ. A study to assess whether fixed-width beam walking provides sufficient challenge to assess balance ability across lower limb prosthesis users. *Clin Rehabil.* 2017;32(4):483-492.
20. 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.
21. Lin S-J, Bose NH. Six-minute walk test in persons with transtibial amputation. *Arch Phys Med Rehabil.* 2008;89(12):2354-2359.
22. Gill S, McBurney H. Reliability of performance-based measures in people awaiting joint replacement surgery of the hip or knee. *Physiother Res Int.* 2008;13(3):141-152.
23. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res.* 2005;19(1):231-240.
24. 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.
25. 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.
26. Bohannon RW. Comfortable and maximum walking speed of adults aged 20-79 years: reference values and determinants. *Age Ageing.* 1997;26(1):15-19.
27. Hafner BJ, Gaunaud 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.* 2017;98(2):277-285.
28. Palmento Government Benefits Administrators. Lower limb prostheses. DMERC Medicare Advis December 1994;(9):99-105.
29. Hanspal RS, Fisher K, Nieveen R. Prosthetic socket fit comfort score. *Disabil Rehabil.* 2003;25(22):1278-1280.
30. Gallistel CR, Fairhurst S, Balsam P. The learning curve: implications of a quantitative analysis. *Proc Natl Acad Sci USA.* 2004;101(36):13124-13131.
31. Jenkins S, Cecins NM. Six-minute walk test in pulmonary rehabilitation: do all patients need a practice test? *Respirology.* 2010;15(8):1192-1196.
32. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res.* 1995;4(4):293-307.
33. Andresen EM. Criteria for assessing the tools of disability outcomes research. *Arch Phys Med Rehabil.* 2000;81:S15-S20.
34. Sions JM, Manal TJ, Horne JR, Sarlo FB, Pohlig RT. Balance-confidence is associated with community participation, perceived physical mobility, and performance-based function among individuals with a unilateral amputation. *Physiother Theor Pract.* 2018;28:1-8.
35. Clemens SM, Gailey RS, Bennett CL, Pasquina PF, Kirk-Sanchez NJ, Gaunaud 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.
36. Solari A, Radice D, Manneschi L, Motti L, Montanari E. The multiple sclerosis functional composite: different practice effects in the three test components. *J Neurol Sci.* 2005;228(1):71-74.
37. Adsett J, Mullins R, Hwang R, et al. Repeat six-minute walk tests in patients with chronic heart failure: are they clinically necessary? *Eur J Cardiovasc Prev Rehabil.* 2011;18(4):601-606.
38. Kristensen MT, Ekdahl C, Kehlet H, Bandholm T. How many trials are needed to achieve performance stability of the timed up & go test in patients with hip fracture? *Arch Phys Med Rehabil.* 2010;91(6):885-889.
39. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med.* 1998;26(4):217-238.
40. Duff K. Practice effects, test-retest stability, and dual baseline assessments with the California verbal learning test in an HIV sample. *Arch Clin Neuropsychol.* 2001;16(5):461-476.
41. Guyatt GH, Pugsley SO, Sullivan MJ, et al. Effect of encouragement on walking test performance. *Thorax.* 1984;39(11):818-822.
42. Liu J, Drutz C, Kumar R, et al. Use of the six-minute walk test post-stroke: is there a practice effect? *Arch Phys Med Rehabil.* 2008;89(9):1686-1692.
43. Ziegler-Graham K, MacKenzie EJ, Ephraim PL, Trivison 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.
44. 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.
45. 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.
46. 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.
47. 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.

## Disclosure

**A.S.** Department of Kinesiology, University of Illinois at Chicago, Chicago, IL.  
Address Correspondence to: A.S., Department of Kinesiology University of Illinois at Chicago 1919 West Taylor Street, Rm. 646 Chicago, IL, 60612; e-mail: asawers@uic.edu

**B.J.H.** Department of Rehabilitation Medicine, University of Washington, Seattle, WA

### Author Contribution:

Study Concept and Design: A. Sawers, B. Hafner.

Acquisition of Data: A. Sawers, B. Hafner.

Analysis and Interpretation of Data: A. Sawers, B. Hafner.

Drafting of Manuscript: A. Sawers.

Critical Revision of Manuscript of Important Intellectual

Content: B. Hafner.

Statistical Analysis: A. Sawers.

Study Supervision: A. Sawers.

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

**Funding Support:** Research reported in this publication was supported by the Department of Defense (DoD) under award number W81XWH-17-1-0547 (AS). The content is solely the responsibility of the authors and does not necessarily represent the official views of the DoD.

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

**Institutional Review:** All study procedures were reviewed and approved by University of Illinois at Chicago and University of Washington institutional review boards.

The authors declare no competing interest.

**Clinical Trials Registration:** Not applicable.

Submitted for publication June 23, 2020; accepted October 5, 2020.

---

# Prosthetics and Orthotics International

## Recalled number of falls in the prior year, combined with perceived mobility, predicts the incidence of future falls in established unilateral lower limb prosthesis users --Manuscript Draft--

|                                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|---------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Manuscript Number:</b>                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| <b>Full Title:</b>                                | Recalled number of falls in the prior year, combined with perceived mobility, predicts the incidence of future falls in established unilateral lower limb prosthesis users                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <b>Article Type:</b>                              | Original Research Report                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| <b>Manuscript Classifications:</b>                | Amputee rehabilitation; Lower limb prosthetics; Rehabilitaion of prostheses users; rehabilitation of amputees; rehabilitation of prostheses users                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| <b>Keywords:</b>                                  | amputation; accidental falls; rehabilitation; amputee                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| <b>Corresponding Author:</b>                      | Andrew Sawers, PhD<br>University of Illinois at Chicago<br>Chicago, IL UNITED STATES                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| <b>Corresponding Author E-Mail:</b>               | asawers@uic.edu                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| <b>First Author:</b>                              | Moaz Tobaigy, MPT, MPO                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| <b>Other Authors:</b>                             | Moaz Tobaigy, MPT, MPO                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|                                                   | Brian J Hafner, PhD                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| <b>Order of Authors (with Contributor Roles):</b> | Moaz Tobaigy, MPT, MPO                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|                                                   | Brian J Hafner, PhD                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|                                                   | Andrew Sawers, PhD                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| <b>Abstract:</b>                                  | <p>Introduction</p> <p>Falls are a frequent and costly concern for lower limb prosthesis (LLP) users. At present, there are no models clinicians can use to predict the incidence of future falls in established LLP users. Assessing who is at risk for falls, therefore, remains a challenge. The objective of this study was to test whether a combination of easily accessible clinical attributes and measurements can predict the incidence of future falls in established unilateral LLP users.</p> <p>Methods</p> <p>A secondary analysis of data from a prospective observational study was conducted. LLP users (n=60) reported the number of falls they recalled experiencing over the past year (i.e., recalled falls) before prospectively reporting falls over a 6-month observation period. Additional potential fall-risk factors were recorded at baseline. Negative binomial regression was used to develop a model intended to predict the incidence of future falls.</p> <p>Results</p> <p>The final model, which included the number of recalled falls (<math>p=.035</math>) and PLUS-M T-scores (<math>p=.078</math>), was significantly better than a null model at predicting the number of falls over the next 6-months (<math>X^2(2)=9.76</math>, <math>p=.008</math>), and fit the observed prospective fall count data (<math>X^2(56)=54.78</math>, <math>p=.20</math>).</p> <p>Conclusion</p> <p>The number of recalled falls, combined with PLUS-M T-scores, predicted the incidence of falls over the next 6 months in established, unilateral LLP users. The success and simplicity of the final model suggests it may serve as a screening tool for clinicians to assess fall risk. Additional research to validate the proposed model of fall risk in LLP users is needed.</p> |

|                                     |               |
|-------------------------------------|---------------|
| <b>Suggested Reviewers:</b>         |               |
| <b>Opposed Reviewers:</b>           |               |
| <b>Manuscript Region of Origin:</b> | UNITED STATES |

**Running Head:** Predicting falls in lower limb prosthesis users

**Title:** Recalled number of falls in the prior year, combined with perceived mobility, predicts the incidence of future falls in established unilateral lower limb prosthesis users

**Authors:** Moaz Tobaigy, MPT, MPO<sup>1,3</sup>, Brian J. Hafner, PhD<sup>2</sup>, Andrew Sawers, CPO, PhD<sup>1</sup>

**Institutions and Affiliations:**

<sup>1</sup> Department of Kinesiology, University of Illinois at Chicago, Chicago, IL 60612

<sup>2</sup> Department of Rehabilitation Medicine, University of Washington, Seattle, WA 98195

<sup>3</sup> Faculty of Medical Rehabilitation Science, King Abdulaziz University, Jeddah, SA 21589

**Corresponding Author:**

Andrew Sawers, CPO, PhD

Department of Kinesiology

University of Illinois at Chicago

1919 West Taylor Street, Rm. 646

Chicago, IL, 60612

Phone: 312-996-1427

Fax: 312-413-0319

Email: [asawers@uic.edu](mailto:asawers@uic.edu)

**Funding Support:** Research reported in this publication was supported by the Department of Defense (DoD) under award number W81XWH-17-1-0547 (AS). The content is solely the responsibility of the authors and does not necessarily represent the official views of the DoD.

**Acknowledgements:** The authors would like to acknowledge Andre Kajlich, Marrita Johnson, Helenlyn Popescu, and Janis Kim for their assistance with data collection and data entry.

**Financial Disclosure / Conflicts of Interest:** The authors declare no competing interest.

**Word Count:** 3819/3000 words (introduction through conclusion)

**Institutional Review:** All study procedures were reviewed and approved by University of Illinois at Chicago and University of Washington institutional review boards.

**Title:** Recalled number of falls in the prior year, combined with perceived mobility, predicts the incidence of future falls in established unilateral lower limb prosthesis users

**Abstract** (247/250 words)

**Introduction:** Falls are a frequent and costly concern for lower limb prosthesis (LLP) users. At present, there are no models clinicians can use to predict the incidence of future falls in established LLP users. Assessing who is at risk for falls, therefore, remains a challenge. The objective of this study was to test whether a combination of easily accessible clinical attributes and measurements can predict the incidence of future falls in established unilateral LLP users.

**Methods:** A secondary analysis of data from a prospective observational study was conducted. LLP users (n=60) reported the number of falls they recalled experiencing over the past year (i.e., recalled falls) before prospectively reporting falls over a 6-month observation period. Additional potential fall-risk factors were recorded at baseline. Negative binomial regression was used to develop a model intended to predict the incidence of future falls.

**Results:** The final model, which included the number of recalled falls ( $p=.035$ ) and PLUS-M T-scores ( $p=.078$ ), was significantly better than a null model at predicting the number of falls over the next 6-months ( $X^2(2)=9.76$ ,  $p=.008$ ), and fit the observed prospective fall count data ( $X^2(56)=54.78$ ,  $p=.20$ ).

**Conclusion:** The number of recalled falls, combined with PLUS-M T-scores, predicted the incidence of falls over the next 6 months in established, unilateral LLP users. The success and simplicity of the final model suggests it may serve as a screening tool for clinicians to assess fall risk. Additional research to validate the proposed model of fall risk in LLP users is needed.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

**Key Words:** amputation, accidental falls, rehabilitation, amputee

3819/3000 words (introduction through conclusion)

**Acronyms and Abbreviations:**

ABC: Activities-specific Balance Confidence scale

CCI = Charlson Comorbidity Index

IQR = Interquartile range

IRR = Incidence rate ratio

LLP = Lower limb prosthesis

MFCL: Medicare Functional Classification Level

PLUS-M: Prosthetic Limb Users Survey – Mobility

SCS: Socket Comfort Score

STARD: Standards for Reporting Diagnostic Accuracy Studies

STROBE: Strengthening The Reporting of Observational studies in Epidemiology

## **Background**

Falls are a frequent and costly health concern for lower limb prosthesis (LLP) users.<sup>1,2</sup> More than 50% of LLP users report having fallen in the past year,<sup>1,3,4</sup> many more than once.<sup>3,5,6</sup> Existing evidence suggests that the prevalence of falls in LLP users places them at a high risk for adverse health outcomes that include activity limitations and participation restrictions,<sup>1,7,8</sup> injuries,<sup>1,9</sup> and reduced quality of life.<sup>10</sup> The ability to identify LLP users at greatest risk for falls, and thus most likely to require preventative interventions, is therefore essential to mitigate the incidence of future falls in LLP users.

Identifying LLP users at risk for falls, however, remains a challenge. Contemporary performance-based balance tests used to screen LLP users for fall risk are prone to ceiling effects,<sup>11–14</sup> are often unable or limited in their ability to discriminate between LLP users with and without a history of falls,<sup>11,12,14,15</sup> and have rarely been validated prospectively.<sup>15</sup> Additionally, performance-based tests are not routinely administered in clinical practice.<sup>16–18</sup>

Owing to the challenges with using performance-based balance tests to assess fall risk, clinicians may turn to commonly held beliefs regarding attributes thought to imply fall risk in LLP users. These may include demographic (e.g., age),<sup>1,19,20</sup> health (e.g., number of co-morbidities),<sup>1,4</sup> amputation (e.g., level, etiology),<sup>1,21</sup> activity (e.g., MFCL, step count),<sup>3,19</sup> and/or balance-related (e.g., prior falls, balance confidence) risk factors. While appealing due to the ease with which they can be assessed, many of these factors either lack support in the literature,<sup>1,4,19,20</sup> have yet to be studied in LLP users despite predicting falls in other clinical

populations,<sup>22-27</sup> or have been evaluated using cross-sectional, rather than the prospective study designs that are needed to establish a temporal relationship with fall status.<sup>1,3,21</sup> Evidence to support the association between these assumed risk factors and falls remains limited among established LLP users.<sup>28,29</sup> There is thus a need to evaluate prospectively the degree to which personal attributes or clinical measurements may predict falls among established LLP users.<sup>28,29</sup>

The purpose of this secondary analysis was therefore to test whether easily-accessible attributes and/or measurements clinicians might use to assess fall risk (e.g., level of amputation, recalled fall incidence), could predict the future incidence of falls in established unilateral LLP users. We hypothesized that a model consisting of easily-accessible fall risk factors, would predict the number of falls in the next 6-months among established, unilateral LLP users, as evidenced by a statistically significant omnibus test and factors in the final model with statistically significant contributions to fall prediction.

## **Methods**

### **Study Design**

A secondary analysis of falls data collected in a multi-site prospective observational study between February 2018 and November 2018, for the purpose of establishing the psychometric properties of several performance-based tests,<sup>30</sup> was performed. Best practices in conducting observational studies (i.e., STROBE: Strengthening The Reporting of Observational studies in Epidemiology),<sup>31</sup> as well as the collection and reporting of diagnostic and prognostic accuracy studies (i.e., Standards for Reporting of Diagnostic Accuracy Studies (STARD))<sup>32</sup> were followed. All data were stored and managed using a REDCap database hosted at XXXX.<sup>33</sup>

Original study protocols were reviewed and approved by institutional review boards at the XXXX and XXXX. All individuals provided written informed consent prior to participation.

## Participants

Study flyers, placed in local prosthetic clinics in two major metropolitan areas in different regions of the country with different climates (i.e., XXXX and XXXX), were used to recruit LLP users via convenience sampling. Established LLP users were targeted because they experience a high number of falls,<sup>1,5,21</sup> even beyond the period of in-patient rehabilitation. The post-rehabilitation phase also represents the period in which LLP users spend most of their lives. Inclusion criteria were 18 years of age or older; unilateral amputation at the transtibial or transfemoral level; amputation due to trauma, dysvascular complications, tumor, or infection; use of a prosthesis for at least one year; use of a comfortable prosthesis to ambulate; able to walk 10 meters without an upper extremity mobility aid (e.g., cane); and able to read, write, and speak English. Exclusion criteria included an amputation of a second limb (e.g., arm or contralateral leg); contralateral complications (e.g., hip replacement, ulcers or infections), advanced neurologic disorder; or severe cardiopulmonary disease that would prevent them from completing study procedures. A priori power analysis was not performed for this secondary analysis. The sample size (n=60) was determined a priori for a separate, but related, research question.<sup>30</sup>

## Procedures

Demographic, health, amputation, mobility, and balance characterization measures were obtained at baseline from all study participants. The number of falls experienced over the

1  
2  
3  
4 116 previous 12 months was self-reported at baseline, and participants were then followed  
5  
6 117 prospectively over the next 6 months. New fall events were reported to study team members  
7  
8  
9 118 during monthly telephone calls.  
10  
11  
12 119

## 13 120 **Measurements**

### 15 121 *Participant characterization measures*

16 122 Demographic information, including age and sex, along with amputation-related  
17  
18 123 characteristics (e.g., etiology, level, and time since amputation) were collected via self-report and  
19  
20  
21 124 interview at baseline. Height and weight were measured by an investigator. Presence and number  
22  
23  
24  
25 125 of co-morbidities were assessed using the Charlson Comorbidity Index (CCI),<sup>34</sup> while mobility  
26  
27  
28 126 was assessed via self-report, interview, and/or clinical inspection (i.e., Prosthetic Limb Users  
29  
30  
31 127 Survey (PLUS-M),<sup>35</sup> hours of prosthesis use per day, and Medicare Functional Classification  
32  
33 128 Level (MFCL))). Balance confidence was assessed using the 5-point ordinal scale (i.e., 0-4)  
34  
35 129 version of the Activities-specific Balance Confidence (ABC) scale.<sup>13,36</sup> Comfort of the prosthesis  
36  
37  
38 130 worn by each participant was evaluated using the Socket Comfort Score (SCS).  
39  
40 131

### 42 132 *Falls assessment*

43 133 The number of falls in the past 12 months was determined by interview. Participants were  
44  
45  
46  
47 134 asked “*in the past year have you lost your balance and landed on the ground or lower level,*  
48  
49  
50 135 *other than as a result of a loss of consciousness, a violent blow, stroke, or epileptic seizure?*”.<sup>37–</sup>  
51  
52 136 <sup>39</sup> The number of falls over the next 6 months was assessed via telephone interview. Participants  
53  
54  
55 137 were called once a month and asked “*I would like to ask you about the time since we last spoke.*  
56  
57 138 *We last saw you/spoke to you on [date]”. Since then, have you lost your balance and landed on*  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 139 *the ground or lower level?”*. For categorical analyses of fall data, participants were classified as  
5  
6 140 non-fallers (0 falls), non-recurrent fallers (1 fall), or recurrent fallers ( $\geq 2$  falls).<sup>4,40,41</sup>  
7  
8  
9 141

## 11 142 **Statistical Analysis**

13  
14 143 Continuous variables were evaluated for normality with the Shapiro-Wilk test.<sup>42</sup> The study  
15  
16 144 sample and fall count data were characterized using descriptive analyses. Falls data were  
17  
18 145 summarized as number of falls; number of non-fallers, non-recurrent fallers, recurrent fallers;  
19  
20  
21 146 and fall rate per person year.<sup>37</sup> Mann-Whitney U and Fisher’s exact (2xC) tests were run to test  
22  
23 147 for differences in continuous and dichotomous demographic (e.g., age, sex), amputation (e.g.,  
24  
25 148 time since, level, socket comfort score), health (e.g., number of co-morbidities), and mobility-  
26  
27 149 related characteristics (e.g., PLUS-M) across faller categories established using the prospective  
28  
29 150 fall data. A Mann-Whitney U test was run to determine whether any difference in the number of  
30  
31 151 falls over the 6-month prospective period was statistically significantly across fall categories  
32  
33 152 established using falls in the past 12 months. The level of significance for these analyses was  
34  
35 153 adjusted to  $\alpha \leq .0167$  to account for multiple comparisons between the three fall categories.  
36  
37  
38  
39  
40  
41 154

42  
43 155 Negative binomial regression was used to develop a model intended to predict the future  
44  
45 156 incidence of falls in LLP users and address the study hypothesis. Negative binomial regression  
46  
47 157 has been identified,<sup>43–45</sup> previously used,<sup>27,46–48</sup> and was selected as the most suitable method for  
48  
49 158 analyzing fall data due to the type of data (i.e., counts), the recurrent and non-independent nature  
50  
51 159 of falls in any one person, the non-normal distribution (i.e., Poisson) and accompanying  
52  
53 160 overdispersion of fall data, as well as its ability to accommodate for missing data and adjust for  
54  
55 161 confounding variables (i.e., covariates). The initial model included factors considered to be: i)  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

easily-accessible attributes and/or measurements clinicians might use to assess fall risk (e.g., level of amputation, number of recalled falls), ii) associated with previous falls in published retrospective studies,<sup>1,3,4</sup> and iii) available in the current dataset. Risk factors were also restricted to those related to falls in established LLP users in the community, as other phases of rehabilitation or care settings have been associated with different risk factors.<sup>28,29</sup> The initial model therefore, included six factors: i) number of falls recalled over the past 12 months, ii) level of amputation, iii) cause of amputation, iv) time since amputation, v) perceived socket comfort (i.e., Socket Comfort Score),<sup>49,50</sup> and vi) perceived mobility (i.e., PLUS-M T-score).<sup>35</sup> We reduced the initial model by removing independent variables in a backward order,<sup>51-53</sup> preserving those variables with p-values  $\leq .15$  to ensure borderline relationships would not be overlooked.<sup>1</sup> Likelihood ratio chi-squared omnibus tests, and Pearson chi-squared goodness-of-fit tests,<sup>54</sup> were run on each iteration of the model to determine if it was significantly better than a null model (i.e., one with no predictors) at predicting the incidence of future falls, and to assess the fit of each model, respectively. These steps were repeated until a statistically significant model (i.e., omnibus test  $p < .05$ ) was identified, which contained factors whose contributions to predicting the incidence of future falls had p-values  $\leq .15$ . Regression coefficients ( $\beta$ ) of factors making statistically significant contributions to the prediction of future falls in the final model ( $p \leq .05$ ) were exponentiated to calculate incidence rate ratios (IRR), and translated into the estimated number of falls associated with that factor while holding the remaining factors in the final model constant at their means.<sup>55</sup> All statistical analyses were performed using SPSS v.27 (Chicago, IL).

## **Results**

Sixty established unilateral LLP users participated in the parent study (Table 1).<sup>30</sup> A total of 91 falls were reported over the past 12 months. 42% (25/60), 28% (17/60), and 30% (18/60) of study participants reported zero, one, or two or more falls in the past 12 months, respectively. The annual fall rate per participant over the 12-month recall period was 1.52 falls. A total of 53 falls were recorded over the next 6 months. 53% (32/60), 28% (17/60), and 19% (11/60) of study participants reported no falls, falling once, or falling at least twice over the prospective 6-month observation period, respectively. The annual fall rate per participant over the prospective 6-month period was 1.76 falls, and the median time to first fall was 3 months (IQR: 3.25 months). All continuous study variables, with the exception of PLUS-M T-scores ( $W=.986$ ,  $p=.729$ ), were non-normally distributed ( $W\leq .956$ ,  $p\leq .029$ ). Mann-Whitney U and Fisher's exact (2xC) tests revealed that among the differences in continuous and dichotomous demographic, amputation, health, perceived mobility, and fall-related factors across prospective faller categories, only PLUS-M T-scores were statistically significant (Table 1).

[INSERT TABLE 1 HERE]

Number of falls recorded during the 6-month prospective period were non-normally distributed ( $W.607$ ,  $p\leq .001$ ), and overdispersed (i.e., variance exceeded the mean), confirming the suitability of negative binomial regression. After reduction, the final negative binomial regression models retained two of its original six factors, number of falls recalled, and PLUS-M T-scores (see supplemental table). A likelihood ratio chi-squared omnibus test revealed that the final model was significantly better than a null model (i.e., one with no factors) at predicting the number of falls likely to occur over the next 6 months ( $X^2(2)=9.76$ ,  $p=.008$ ), while a chi-squared

goodness-of-fit test revealed that there was no statistically significant difference between the expected and observed number of falls in the prospective period ( $X^2(56)=54.78$ ,  $p=.20$ ), implying the model fit the prospective fall data.

Based on the final model, and holding PLUS-M T-scores at their sample mean, the modeled relationship between the number of falls recalled and the incidence of future falls, predicts a statistically significant increase in the number of falls over the next 6 months as the number of falls recalled in the past 12 months goes up ( $p=.035$ ) (Table 2; Figure 1A). In contrast, when holding the number of falls recalled over the past year at its mean, the estimated relationship between PLUS-M T-scores (i.e., perceived mobility) and future falls, predicts a protective, but not statistically significant decrease ( $p=.078$ ) in the number of future falls as PLUS-M T-scores rise (Table 2, Figure 1B). Viewed jointly within the final model [EQ.1], the combined effect of the number of falls recalled and perceived mobility on predicting the incidence of future falls was statistically significant (Table 2).

$$\# \text{ future falls} = \exp((0.126)(\# \text{ falls past year}) + (-0.052)(\text{PLUS-M T-score}) + 2.452) \quad [\text{EQ.1}]$$

[INSERT TABLE 2 HERE]

[INSERT FIGURE 1 HERE]

## **Discussion**

1  
2  
3  
4 229 The purpose of this secondary analysis was to test whether fall risk factors, frequently easily-  
5  
6 230 accessible attributes and/or measurements clinicians might use to assess fall risk, could predict  
7  
8  
9 231 the incidence of future falls in established unilateral LLP users. The resulting negative binomial  
10  
11  
12 232 regression model supported the hypothesis that a combination of easily-accessible fall risk  
13  
14 233 factors would predict the number of future falls in established unilateral LLP users. Consistent  
15  
16 234 with studies in other clinical populations,<sup>22–27</sup> a recalled number of falls contributed to the  
17  
18  
19 235 prediction of the number of falls over the next 6 months in established unilateral LLP users.  
20  
21 236 Owing to the simplicity of the final model, querying the number of falls recalled over the past  
22  
23  
24 237 year and administering PLUS-M, may serve as a suitable method for clinicians to obtain an  
25  
26 238 initial assessment of fall risk, and determine whether additional testing is warranted in  
27  
28  
29 239 established unilateral LLP users. Likewise, researchers may use these results as a baseline from  
30  
31 240 which to develop and validate more detailed predictive models of fall risk in LLP users.  
32  
33  
34 241

35  
36 242 Despite the high prevalence of falls in LLP users,<sup>1,3–5</sup> and relative simplicity of inquiring  
37  
38 243 about the number of falls recalled over the past year, this is – to the best of our knowledge – the  
39  
40  
41 244 first study to examine recalled falls as a predictor of future falls in LLP users. Expressing the  
42  
43 245 regression coefficient ( $\beta$ ) for fall recall in the final model as an incidence rate ratio (IRR), and  
44  
45  
46 246 holding all other factors (i.e., PLUS-M T-scores) at their mean, for every additional fall in the  
47  
48 247 past year, there is a compounding 1.13 times (95% CI: 1.01-1.28) or 13% (95% CI: 1-28%)  
49  
50  
51 248 increase in the likelihood of another fall over the next 6 months. For example, a LLP user who  
52  
53 249 recalls one fall in the past year would have a 13% likelihood of falling again in the next six  
54  
55  
56 250 months, a LLP user who recalls 2 falls in the past year would have a 28% likelihood of falling  
57  
58 251 again in the next 6 months, while a LLP user who recalls 3 falls in the past year would have a  
59  
60  
61  
62  
63  
64  
65

45% likelihood of falling again in the next 6 months. Among established unilateral LLP users, the number of falls recalled in the past year therefore informs the probability of future falls. The relationship between recalled and future falls should be viewed however, in combination with the remaining factor in the final model, PLUS-M T-scores. For example, the final predictive model (i.e., EQ. 1) applied to an established LLP user who recalls two falls in the past year, and a PLUS-M T-score of 44 would predict 1.52 falls for that individual over the next 6 months;

$$\# \text{ of predicted falls in next 6 months} = \exp ((0.126)(2) + (-0.052)(44) + 2.452)$$

$$\# \text{ of predicted falls in next 6 months} = \exp (0.42)$$

$$\# \text{ of predicted falls in next 6 months} = 1.52$$

The results of the current analysis address a notable clinical gap, a prospectively developed fall prediction model clinicians can implement with relative ease to assess fall risk in established, unilateral LLP users. Clinicians may adopt this model as a simple means with which to quantitatively characterize, and monitor, fall risk in patients with unilateral lower limb amputation. The adoption of this model in clinical practice may help address the persistent prevalence of falls in LLP users by helping clinicians identify those individuals at risk, and most likely to need and benefit from intervention.

Previous retrospective studies identified demographic, amputation, and health-related characteristics that were each associated with the number of falls recalled over the past year in LLP users.<sup>28,29</sup> Among established LLP, these include level of amputation,<sup>1,3</sup> time since amputation,<sup>1</sup> hours of prosthetic wear,<sup>21</sup> residual limb and socket problems,<sup>1</sup> and the number of

co-morbidities.<sup>1,15</sup> In the current study however, these risk factors did not significantly contribute to the prediction of future falls in established, unilateral LLP users. The absence of any statistically significant contributions from these historical fall risk factors to the prediction of future falls in the current study may be due to differences in sample size or composition, cross-sectional versus prospective study designs,<sup>56</sup> or limitations associated with previous categorical analyses of falls data (e.g., faller versus non-faller).<sup>43,44,46</sup> The sample in the current study was notably smaller (i.e., n=60 versus, n=164-435), slightly younger (i.e., 58.5 years old versus approximately 65 years old), and composed of a smaller proportion of participants with dysvascular (i.e., 40% versus 53%-78%) and transfemoral (i.e., 27% versus 27%-45%) amputations.<sup>1,3,15,21</sup> Fall risk factors based on health, level and time since amputation, as well as hours of prosthesis wear may therefore be better predictors of falls in older, transfemoral, dysvascular LLP users. While expedient and appropriate for initial testing, cross-sectional studies cannot establish a temporal relationship between current health, fall status, and the incidence of future falls.<sup>56</sup> Prospective study designs such as that employed here, may therefore be better suited to identify true fall risk factors. In addition to study design, methods of data analysis may influence the interpretation of fall risk factors. Specifically, limitations associated with treating falls data as categorical variables (e.g., non-faller, non-recurrent faller, and recurrent faller) rather than count data have recently been described.<sup>43-45</sup> Combined with the paucity of prospective fall studies in LLP users, the discrepancies in fall risk factors between the current study and previously published results suggest that additional prospective investigations with larger samples are warranted to examine LLP user-specific fall risk factors across a host of demographic, amputation, health, and mobility-related characteristics.<sup>29</sup>

## **Limitations**

There are a number of limitations in this study that need to be considered when interpreting the results and planning future research. The sample size was determined a priori, but for a different research question.<sup>30</sup> Furthermore, the sample composition (e.g., age, level of amputation, and cause of amputation) was slightly skewed towards people younger, transtibial, and non-dysvascular LLP users.<sup>31,32,4657-60</sup> Consequently, the proposed fall prediction model may not apply to all established LLP users. Larger studies, powered *a priori*, that include participants representative of the general LLP user population, are therefore required to verify and extend the current findings.

There was also no control for exposure, or amount of time participants were active each day. Participants in the current study with a higher frequency of falling may have had more opportunity to fall if they were more active. However, procedures for collection and reporting falls data in the current study were based on a consensus statement that recommends against adjusting for physical activity when presenting and analyzing fall data, suggesting instead that it is better to report the absolute risk of falls.<sup>37</sup> Collecting activity-related data may prove useful in identifying those at risk of falls, and should therefore be considered in future predictive fall risk models for LLP users.

Due to recall bias or decay,<sup>61,62</sup> participants may have been challenged to accurately recall the precise number of falls they experienced over the past year (i.e., it is possible that LLP users may have over- or underestimated the true number of falls that occurred). The accuracy of recall does not, however, alter whether the number of falls recalled is able to predict future falls.

Whether the number of falls recalled accurately reflects fall history (i.e., the true number of fall events that occurred), or whether it predicts future falls are related but separate questions. Further, asking LLP users to recall the number of falls over the past year as a means to gauge future fall risk, reflects how this information would typically be collected in day-to-day clinical practice, where the results of this study could be applied. It is also consistent with existing research protocols (i.e., it is the gold standard by which such data is gathered), and has been used in many other published studies, across multiple clinical populations, to test the relationship between past and future falls.<sup>22–24,26,27,46,47</sup> A longer recall timeframe (i.e. one year) was chosen in the current study because it has been shown to minimize recall bias when compared to shorter periods (i.e. 3 or 6 months) in older adults.<sup>62,63</sup> To date, no fall recall accuracy studies have been published in LLP users. As such, it is unclear whether LLP users generally over- or under-report the number of prior falls. Future research exploring this topic is therefore warranted.

The predictive model developed in this study does not differentiate between injurious and non-injurious falls. Many of the factors that were included in the initial model (e.g., cause, level, time since amputation, and age) are not associated with increased risk or incidence of injurious falls in LLP users,<sup>41</sup> and therefore may not be well suited to predicting the risk of injurious falls. Given the financial, behavioral, and psychological costs of fall-related injuries,<sup>2</sup> future studies should include alternative, and ideally modifiable, factors to develop a robust predictive model for injurious fall risk in LLP users. Such studies will likely need to extend their prospective data collection period beyond that of the current study to ensure a sufficient number of injurious falls are captured. Similarly, improved methods to collect injury-related falls data in LLP users is required.

Results of the current study are limited to predicting the number of falls over the subsequent 6 months. A longer prospective follow-up period (e.g., 1 year rather than 6 months) may be warranted in future research as seasonal differences may alter participants' activities and corresponding fall risk. Whether the model in the current study could predict falls over a longer period is unknown.

This secondary analysis was limited to a select number of possible fall risk factors, and a single phase of rehabilitation (i.e., post-rehabilitation), and therefore may not apply to other (e.g., acute) phases of rehabilitation. Expansion of the predictive model requires collection and prospective evaluation of additional potential fall risk factors across each phase of post-amputation rehabilitation.<sup>28,29</sup>

## **Conclusion**

The combination of the number of falls recalled over the prior year, and PLUS-M T-scores, was found to predict the number of falls over the next 6 months in established, unilateral LLP users. Notably, a host of other demographic, health, and amputation-related factors previously found to be related to a past falls, did not significantly contribute to the prediction of future falls in LLP users. Additional prospective studies are required to further validate the proposed model, and expand upon these results. Additional factors including scores from performance-based balance tests; measures of strength, sensation, and reaction time may increase predictive power. The results of this secondary analysis suggest that the prospectively developed model could be

used by clinicians as a simple method to screen for fall risk, and determine if additional testing of balance deficits is warranted in established, unilateral LLP users.

## **References**

1. 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.
2. Mundell B, Kremers HM, Visscher S, Hoppe K, Kaufman K. Direct medical costs of accidental falls for adults with transfemoral amputations. *Prosthet Orthot Int Sage Publ Ltd*. 2017;41(6):564-570.
3. Gauthier-Gagnon C, Grisé M-C, Potvin D. Enabling factors related to prosthetic use by people with transtibial and transfemoral amputation. *Arch Phys Med Rehabil*. 1999;80(6):706-713.
4. 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.
5. 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*. 2019;11(4):344-353.
6. Ülger Ö, Tpouz S, Bayramlar K, Erbahçeci F, Sener G. Risk factors, frequency, and causes of falling in geriatric persons who has had a limb removed by amputation. *Top Geriatr Rehabil*. 2010;26(2):156-163.
7. Miller WC, Deathe AB. The influence of balance confidence on social activity after discharge from prosthetic rehabilitation for first lower limb amputation. *Prosthet Orthot Int*. 2011;35(4):379-385.
8. Velzen JM van, van Bennekom CAM, Polonski W, Slootman JR, van der Woude LHV, Houdijk H. Physical capacity and walking ability after lower limb amputation: a systematic review. *Clin Rehabil*. 2006;20(11):999-1016.
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):521-545.
10. 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.

11. 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.e1.
12. 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.
13. 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.
14. 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.
15. 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.
16. Gaunaud 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.
17. 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.
18. Wedge FM, Braswell-Christy J, Brown CJ, Foley KT, Graham C, Shaw S. Factors influencing the use of outcome measures in physical therapy practice. *Physiother Theory Pract*. 2012;28(2):119-133.
19. Hordacre B, Barr C, Crotty M. Community activity and participation are reduced in transtibial amputee fallers: a wearable technology study. *BMJ Innov*. 2015;1:10-16.
20. 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.
21. 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.
22. Mackintosh SF, Hill KD, Dodd KJ, Goldie PA, Culham EG. Balance score and a history of falls in hospital predict recurrent falls in the 6 months following stroke rehabilitation. *Arch Phys Med Rehabil*. 2006;87(12):1583-1589.
23. Bloem BR, Grimbergen YAM, Cramer M, Willemsen M, Zwinderman AH. Prospective assessment of falls in Parkinson's disease. *J Neurol*. 2001;248(11):950-958.

24. Swinkels A, Newman JH, Allain TJ. A prospective observational study of falling before and after knee replacement surgery. *Age Ageing*. 2009;38(2):175-181.
25. Forster A, Young J. Incidence and consequences of falls due to stroke: a systematic inquiry. *Br Med J*. 1995;310(6997):83.
26. Cameron MH, Thielman E, Mazumder R, Bourdette D. Predicting falls in people with Multiple Sclerosis: fall history is as accurate as more complex measures. *Mult Scler Int*. 2013;496325-496327.
27. Mänty M, Mänty M, Heinonen A, et al. Self-reported preclinical mobility limitation and fall history as predictors of future falls in older women: prospective cohort study. *Osteoporos Int*. 2010;21(4):689-693.
28. Hunter S, Batchelor F, Hill K, Hill A, 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.e1.
29. Steinberg N, Gottlieb A, Siev-Ner I, Plotnik M. Fall incidence and associated risk factors among people with a lower limb amputation during various stages of recovery – a systematic review. *Disabil Rehabil*. 2019;41(15):1778-1787.
30. Blinded
31. Elm E von, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med*. 2007;147(8):573.
32. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*. 2016;6(11):e012799.
33. 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.
34. Chaudhry S, Jin L, Meltzer D. Use of a self-report-generated Charlson Comorbidity Index for predicting mortality. *Med Care*. 2005;43(6):607-615.
35. Hafner BJ, Gaunard 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*. 2017;98(2):277-285.
36. Powell LE, Myers AM. The Activities-specific Balance Confidence (ABC) Scale. *J Gerontol A Biol Sci Med Sci*. 1995;50A(1):M28-M34.
37. Lamb SE, Jorstad-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 JAGS*. 2005;53(9):1618-1622.

38. Lord SR, Ward JA, Williams P, Anstey KJ. Physiological factors associated with falls in older community-dwelling women. *J Am Geriatr Soc JAGS*. 1994;42(10):1110-1117.
39. Askham J, Glucksman E, Owens P, Swift C. 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.
40. Sawers A, Hafner BJ. Using clinical balance tests to assess fall risk among established unilateral lower limb prosthesis users: cutoff scores and associated validity indices. *PM&R*. 2020;12(1):16-25.
41. Wong CK, Chihuri ST, 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.
42. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika*. 1965;52(3-4):591-611.
43. Donaldson MG, Sobolev B, Cook WL, Janssen PA, Khan KM. Analysis of recurrent events: a systematic review of randomised controlled trials of interventions to prevent falls. *Age Ageing*. 2009;38(2):151-155.
44. Robertson MC, Campbell AJ, Herbison P. Statistical analysis of efficacy in falls prevention trials. *J Gerontol A Biol Sci Med Sci*. 2005;60(4):530-534.
45. Aeberhard WH, Cantoni E, Heritier S. Robust inference in the negative binomial regression model with an application to falls data. *Biometrics*. 2014;70(4):920-931.
46. Simpson LA, Miller WC, Eng JJ. Effect of Stroke on fall rate, location and predictors: a prospective comparison of older adults with and without Stroke. *PloS One*. 2011;6(4):e19431.
47. Davis JC, Best JR, Khan KM, et al. Slow processing speed predicts falls in older adults with a falls history: 1- year prospective cohort study. *J Am Geriatr Soc*. 2017;65(5):916-923.
48. Barnett A. Community-based group exercise improves balance and reduces falls in at-risk older people: a randomised controlled trial. *Age Ageing*. 2003;32(4):407-414.
49. Hanspal RS, Fisher K, Nieveen R. Prosthetic socket fit comfort score. *Disabil Rehabil*. 2003;25:1278-1280.
50. Hafner BJ, Morgan SJ, CPO, Askew RL, Salem R. Psychometric evaluation of self-report outcome measures for prosthetic applications. *J Rehabil Res Dev*. 2016;53(6):797-812.
51. Sun G, Shook T, Kay G. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J Clin Epidemiol* 1996; 49: 907-916. 29.

52. Harrell F Jr., Lee K, Mark D. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; 15: 361–387. 30.
53. Steyerberg E, Eijkemans M, Harrell F Jr., Habbema J. Prognostic modeling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med* 2000; 19: 1059–1079.
54. Sukhatme, P.V. On the distribution of 2 in samples of the poisson series. *J. Royal Statist. Soc. B* 1938, 5, 75–79.
55. Long JS, Freese J. Regression models for categorical dependent variables using stata, 2nd ed. *College Station: Stata Press*; 2006.
56. Muir S, Speechley M. Establishing predictive validity of the Fullerton Advanced Balance Scale. *Arch Phys Med Rehabil.* 2010;91(7):1147.
57. 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.
58. Ephraim PL, Wegener ST, MacKenzie EJ, Dillingham TR, Pezzin LE. Phantom pain, residual limb pain, and back pain in amputees: results of a national survey. *Arch Phys Med Rehabil.* 2005;86(10):1910-1919.
59. 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.
60. Wurdeman S, Stevens P, Campbell J. Mobility analysis of Amputees II: comorbidities and mobility in lower limb prosthesis users. *Am J Phys Med Rehabil.* 2018;97(11):782-788.
61. Feldman S F, Robinovitch. Recalling the mechanics of falls: young adults cannot accurately describe the sites of impact immediately after a fall occurs. *Osteoporos Int.* 2006;17:955-5.
62. Cummings SR, Nevitt MC, Kidd S. Forgetting falls. The limited accuracy of recall of falls in the elderly. *J Am Geriatr Soc JAGS.* 1988;36(7):613-616.
63. Hale WA, Delaney MJ, Cable T. Accuracy of patient recall and chart documentation of falls. *J Am Board Fam Med.* 1993; 6(3); 239-242

## **List of Tables**

Table 1. Characteristics of the study sample by faller category established using the prospective fall data

Table 2. Negative binomial regression model to predict falls in unilateral lower limb prosthesis users

### **Figure Captions**

Figure 1 – Estimated relationship between factors retained in the final model and the predicted number of falls in the next 6 months for unilateral lower limb prosthesis users. Each plot displays predicted number of falls at different levels of the predictor while holding the remaining variable in the model at its mean. The predicted number of falls are plotted for values of the final model predictors, along with a 95% confidence interval (A: number of falls recalled in past 12 months; B: PLUS-M T-scores). Sample mean is indicated by the arrow (recalled falls: 1.52; PLUS-M T-score: 56.1).

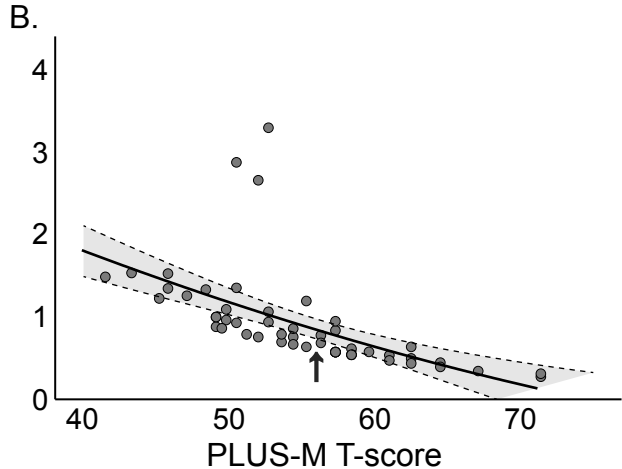
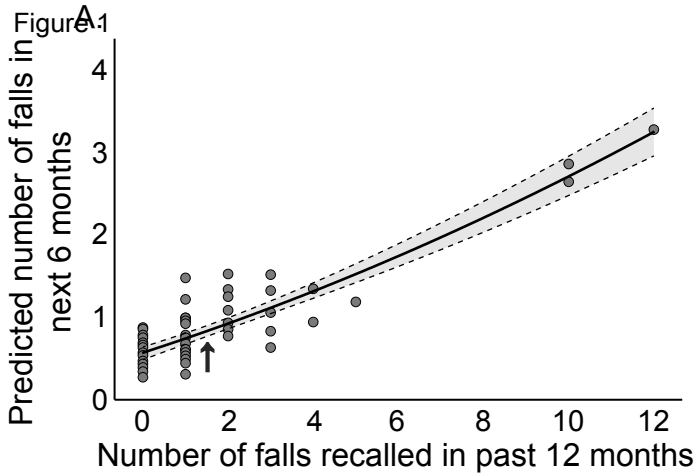


Table 1. Characteristics of the study sample by faller category established using the prospective fall data

|                               |                 | Non-fallers<br>(n=32) | Non-recurrent fallers<br>(n=17) | Recurrent fallers<br>(n=11) | p-value            |
|-------------------------------|-----------------|-----------------------|---------------------------------|-----------------------------|--------------------|
|                               |                 | # of participants (%) |                                 |                             |                    |
| Level                         | Transtibial     | 25 (42%)              | 13 (22%)                        | 7 (12%)                     | 0.660              |
|                               | Transfemoral    | 7 (12%)               | 4 (7%)                          | 4 (7%)                      |                    |
| Etiology                      | Dysvascular     | 15 (25%)              | 3 (5%)                          | 6 (10%)                     | 0.080              |
|                               | Non-Dysvascular | 17 (28%)              | 14 (23%)                        | 5 (8%)                      |                    |
| MFCL                          | K1 and K2       | 9 (15%)               | 5 (8%)                          | 4 (7%)                      | 0.870              |
|                               | K3 and K4       | 23 (38%)              | 12 (20%)                        | 7 (12%)                     |                    |
| Sex                           | Male            | 24 (40%)              | 11 (18%)                        | 7 (12%)                     | 0.660              |
|                               | Female          | 8 (13%)               | 6 (10%)                         | 4 (7%)                      |                    |
| Median (IQR)                  |                 |                       |                                 |                             |                    |
| Time since amputation (years) |                 | 9.0 (24.0)            | 19.0 (21.0)                     | 12.0 (28.0)                 | 0.872              |
| Prosthesis use (hours/day)    |                 | 15.0 (4.0)            | 14.0 (4.5)                      | 16.0 (8.0)                  | 0.769              |
| Socket Comfort Score (0-10)   |                 | 8.0 (2.0)             | 8.0 (2.0)                       | 7.0 (1.0)                   | 0.226              |
| PLUS-M (T-score)              |                 | 57.3 (11.6)           | 57.3 (8.95)                     | 51.2 (3.60)                 | 0.015 <sup>a</sup> |
| ABC Scale (0-4)               |                 | 3.38 (1.09)           | 3.25 (0.84)                     | 2.63 (1.44)                 | 0.502              |
| Age (years)                   |                 | 58.5 (25.2)           | 57.0 (24.5)                     | 59.0 (16.0)                 | 0.928              |
| CCI (0-10)                    |                 | 0 (2.0)               | 1.0 (2.0)                       | 1.0 (20)                    | 0.344              |

ABC: Activities-specific Balance Confidence scale; CCI: Charleston Co-morbidity Index; IQR: Interquartile range; MFCL: Medicare Functional Classification Level; PLUS-M: Prosthetic Limb Users Survey of Mobility

<sup>a</sup> Recurrent fallers significantly lower than non-fallers (p=.007) and non-recurrent fallers (p=.004)

Table 2

Table 2. Negative binomial regression model to predict falls in unilateral lower limb prosthesis users

|                                                                                                                        |         |                    |         |
|------------------------------------------------------------------------------------------------------------------------|---------|--------------------|---------|
| Final model ( $\beta_0 = 2.452$ )                                                                                      |         |                    |         |
| Variable                                                                                                               | $\beta$ | IRR (95% CI)       | p-value |
| Number falls recalled                                                                                                  | 0.126   | 1.13 (1.01, 1.28)  | .035    |
| PLUS-M T-score                                                                                                         | -0.052  | 0.949 (.895, 1.01) | .078    |
| Predicted # falls = $\exp((0.126) * (\text{falls recalled past 12 mo.}) + (-0.052) * (\text{PLUS-M T-score}) + 2.452)$ |         |                    |         |
| $\beta$ : regression coefficient; $\beta_0$ : regression intercept; IRR: incidence rate ratio                          |         |                    |         |

Supplementary Table. Model development (reduction)

| Initial model: Fall history, PLUS-M, Socket Comfort Score, cause, level, and time since amputation |                                  |                                     |                            |                |        |
|----------------------------------------------------------------------------------------------------|----------------------------------|-------------------------------------|----------------------------|----------------|--------|
| Variable                                                                                           | Wald $X^2$ test of model effects | Likelihood ratio $X^2$ omnibus test | Goodness-of-fit            |                |        |
|                                                                                                    |                                  |                                     | $X^2$ goodness-of-fit      | Log Likelihood | AIC    |
| Number recalled falls                                                                              | $X^2(1)=9.73$<br>$p=.002$        | $X^2(6)=6.72$<br>$p=.348$           | $X^2(52)=78.93$<br>$p<.05$ | -74.62         | 165.25 |
| PLUS-M                                                                                             | $X^2(1)=3.03$<br>$p=.082$        |                                     |                            |                |        |
| Socket Comfort Score                                                                               | $X^2(1)=1.29$<br>$p=.257$        |                                     |                            |                |        |
| Cause of amputation                                                                                | $X^2(1)=.550$<br>$p=.458$        |                                     |                            |                |        |
| Level of amputation                                                                                | $X^2(1)=.056$<br>$p=.812$        |                                     |                            |                |        |
| <i>Time since amputation</i>                                                                       | $X^2(1)=.014$<br>$p=.907$        |                                     |                            |                |        |
| 2nd model: Fall history, PLUS-M, Socket Comfort Score, cause, and level of amputation              |                                  |                                     |                            |                |        |
| Variable                                                                                           | Wald $X^2$ test of model effects | Likelihood ratio $X^2$ omnibus test | Goodness-of-fit            |                |        |
|                                                                                                    |                                  |                                     | $X^2$ goodness-of-fit      | Log Likelihood | AIC    |
| Number recalled falls                                                                              | $X^2(1)=9.89$<br>$p=.002$        | $X^2(5)=6.64$<br>$p=.249$           | $X^2(53)=78.88$<br>$p<.05$ | -74.66         | 163.32 |
| PLUS-M                                                                                             | $X^2(1)=3.16$<br>$p=.075$        |                                     |                            |                |        |
| Socket Comfort Score                                                                               | $X^2(1)=1.28$<br>$p=.258$        |                                     |                            |                |        |
| Cause of amputation                                                                                | $X^2(1)=.592$<br>$p=.442$        |                                     |                            |                |        |
| <i>Level of amputation</i>                                                                         | $X^2(1)=.047$<br>$p=.828$        |                                     |                            |                |        |
| 3rd model: Fall history, PLUS-M, Socket Comfort Score, and cause of amputation                     |                                  |                                     |                            |                |        |
| Variable                                                                                           | Wald $X^2$ test of model effects | Likelihood ratio $X^2$ omnibus test | Goodness-of-fit            |                |        |
|                                                                                                    |                                  |                                     | $X^2$ goodness-of-fit      | Log Likelihood | AIC    |
| Number recalled falls                                                                              | $X^2(1)=10.0$<br>$p=.002$        | $X^2(4)=6.50$<br>$p=.165$           | $X^2(54)=79.47$<br>$p<.05$ | -74.73         | 161.47 |
| PLUS-M                                                                                             | $X^2(1)=3.14$<br>$p=.077$        |                                     |                            |                |        |
| Socket Comfort Score                                                                               | $X^2(1)=1.30$<br>$p=.255$        |                                     |                            |                |        |
| <i>Cause of amputation</i>                                                                         | $X^2(1)=.655$<br>$p=.418$        |                                     |                            |                |        |
|                                                                                                    |                                  |                                     |                            |                |        |
| 4th model: Fall history, PLUS-M, Socket and Comfort Score                                          |                                  |                                     |                            |                |        |
| Variable                                                                                           | Wald $X^2$ test of model effects | Likelihood ratio $X^2$ omnibus test | Goodness-of-fit            |                |        |
|                                                                                                    |                                  |                                     | $X^2$ goodness-of-fit      | Log Likelihood | AIC    |
| Number recalled falls                                                                              | $X^2(1)=4.28$<br>$p=.039$        | $X^2(3)=10.6$<br>$p=.014$           | $X^2(55)=54.87$<br>$p<.20$ | -72.69         | 155.37 |
| PLUS-M                                                                                             | $X^2(1)=2.26$<br>$p=.133$        |                                     |                            |                |        |
| <i>Socket Comfort Score</i>                                                                        | $X^2(1)=.826$<br>$p=.363$        |                                     |                            |                |        |

**Final model: Fall history and PLUS-M**

| Variable              | Wald $X^2$ test of<br>model effects | Likelihood ratio $X^2$<br>omnibus test | Goodness-of-fit          |                |        |
|-----------------------|-------------------------------------|----------------------------------------|--------------------------|----------------|--------|
|                       |                                     |                                        | $X^2$ goodness-of-fit    | Log Likelihood | AIC    |
| Number recalled falls | $X^2(1)=4.44$<br>p=.035             | $X^2(2)=9.76$<br>p=.008                | $X^2(56)=54.78$<br>p=.20 | -73.10         | 154.21 |
| PLUS-M                | $X^2(1)=3.10$<br>p=.078             |                                        |                          |                |        |

**Running Head:** Predicting falls in lower limb prosthesis users

**Title:** Performance-based balance tests, combined with the number of falls recalled in the past year, predicts the incidence of future falls in established unilateral transtibial prosthesis users.

**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.

**Clinical Trials Registration:** Not applicable

**Abstract**

**Introduction:** Falls are a frequent and costly concern for lower limb prosthesis (LLP) users. At present, there is no prospectively collected evidence of predictive validity for contemporary performance-based balance tests administered to established LLP users. Whether performance-based balance tests can predict future falls in established LLP users remains unknown. The objective of this study was to test whether common performance-based balance tests could predict the number of future falls in established unilateral LLP users.

**Methods:** A multi-site prospective observational study was conducted. At baseline LLP users (n=60) reported the number of falls they recalled experiencing over the past year (i.e., recalled falls), and were administered three performance-based balance tests; the Timed Up and Go, Four Square Step Test, and Narrowing Beam Walking Test, before prospectively reporting falls over a 6-month observation period. Additional potential fall-risk factors were recorded at baseline. Negative binomial regression was used to develop a model intended to predict the incidence of future falls.

**Results:** The final model, which included Timed Up and Go ( $p=.044$ ) and Four Square Step Test ( $p=.149$ ) times, as well as the number of recalled falls ( $p=.009$ ), was significantly better than a null model at predicting the number of falls over the next 6-months ( $X^2(3)=5.39$ ,  $p=.009$ ) in established, unilateral transtibial prosthesis users, and fit the observed prospective fall count data ( $X^2(41)=36.12$ ,  $p=.20$ ). No combination of factors significantly predicted the number of future falls in participants with transfemoral amputation.

**Conclusion:** The number of recalled falls, combined with Timed Up and Go and Four Square Step Test scores, predicted the incidence of falls over the next 6 months in established, unilateral, but not transfemoral prosthesis users. Given the ease with which fall recall, FSST, and TUG times can be collected, the resulting predictive model may serve as a suitable method for clinicians to predict the number of falls over the next 6 months in established unilateral transtibial prosthesis users.

**Key Words:** amputee, accidental falls, postural balance, patient outcomes assessment, rehabilitation.

47    **Abbreviations:**

48    ABC = Activities-specific Balance Confidence scale

49    FSST = Four Square Step Test

50    LLP = lower limb prosthesis

51    MFCL = Medicare Functional Classification Level

52    NBWT = Narrowing Beam Walking Test

53    SD = standard deviation

54    TUG = Timed Up and Go

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

## **Introduction**

Falls remain a common<sup>1-3</sup> and costly<sup>4</sup> healthcare problem for lower limb prosthesis (LLP) users<sup>4</sup>. The ability to identify LLP users at greatest risk for falls, and thus most likely to need and benefit from rehabilitation services and prosthetic interventions, is essential to lowering the incidence of falls in LLP users. Effective screening of fall risk in LLP users remains elusive however, due to a paucity of psychometric evidence to support the use of specific risk factors and/or performance-based tests<sup>5-7</sup>. In the absence of valid thresholds, indices, or models to help clinicians predict who is at risk of falling, the prescription and delivery of prosthetic interventions and rehabilitation services intended to reduce the incidence of falls by LLP users will remain a challenge.

Existing evidence to guide the assessment of fall risk in LLP users is largely based on retrospective or cross-sectional study designs<sup>2,7-16</sup>. While suitable for identifying potential fall risk factors<sup>2,10,11,16,17</sup>, evaluating the degree of content and construct validity of performance-based tests<sup>7,9,18</sup>, or deriving initial validity indices or thresholds to discriminate LLP users with and without a history of fall(s)<sup>7,8,13</sup>, cross-sectional study designs are unable to establish the temporal relationship between balance ability or risk factor, and fall status<sup>19</sup>. Prospective studies, where balance ability or risk factors are ascertained at baseline, and future fall events are tracked over an ensuing reporting period, are required to properly assess the predictive validity of fall risk factors and performance-based balance tests<sup>5</sup>.

Evidence of predictive validity for common performance-based balance tests, and hence clinicians' ability to predict future falls in LLP users, is limited by the number and scope of

prospective studies that have investigated fall-related events in LLP users<sup>5,7,13</sup>. To date, prospective, fall-related studies in LLP users have identified risk factors associated with injurious falls (i.e., sex and race)<sup>20</sup>; inpatient falls (i.e., age greater than 70)<sup>21</sup>, as well as falls by acute transtibial prosthesis users (i.e., number of co-morbidities, activity)<sup>22</sup>, and established LLP users (i.e., number of recalled falls, mobility)<sup>23</sup>. Prospective validation of performance-based balance tests administered to LLP users have received even less attention. In the lone study thus far, thresholds of  $\geq 19$ s and 24s on the Four Square Step Test (FSST) and Timed Up and Go (TUG) respectively, provided good identification of acute (i.e.,  $< 6$  months post amputation) unilateral transtibial prosthesis users at risk for multiple falls<sup>22</sup>. Whether performance-based balance tests can predict future falls in established LLP users remains unknown. As a result, there is little in the way of thresholds, indices, or models to guide the assessment of fall risk in established LLP users. Clinicians therefore, have a limited ability to identify LLP users most likely to need and benefit from intervention intended to reduce falls.

The primary objective of this prospective study was to assess whether common performance-based balance tests could predict the number of future falls in established unilateral LLP users. Based on their construct and discriminant validity<sup>7,13</sup>, we hypothesized that a combination of performance-based balance tests would predict the number of future falls in unilateral LLP users. We further hypothesized that the addition of performance-based test scores to the number of falls recalled over the past year would improve the prediction of future falls in LLP users compared to fall recall, previously found to predict future falls in established, unilateral LLP users<sup>23</sup>. We also sought to derive predictive regression equations (i.e., models), and incidence rate ratios for those factors found to make statistically significant contributions to the prediction of future falls. These

models and indices would allow clinicians and researchers to interpret performance-based balance test scores, and assess an individual LLP users' fall risk.

## **Methods**

### **Study Design**

A two-site (XXXX and XXXX) prospective observational study was conducted between February 2018 and December 2018 to assess the predictive validity of the Narrowing Beam Walking Test (NBWT)<sup>7</sup>, Timed Up and Go (TUG)<sup>24</sup>, and Four Square Step Test (FSST)<sup>25</sup> in established unilateral LLP users. Study protocols were reviewed and approved by an institutional review board at each site. All individuals prior to participation provided written informed consent.

### **Participants**

Individuals with a unilateral lower limb amputation were recruited from nearby prosthetic clinics in XXXX and XXXX using convenience sampling. To participate, individuals were required to be 18 years of age or older; have a unilateral transtibial or transfemoral amputation due to trauma, dysvascular complications, cancer, or infection, for which they had worn a prosthesis for at least one year; be able to walk 10 meters without use of a cane or walker; and possess the ability to read, write, and speak English. Participants were excluded if they had an amputation of a second limb, contralateral complications, or a neuromusculoskeletal or cardiopulmonary condition that would preclude them from completing testing procedures.

An a priori power analysis indicated that a sample of 30 LLP users (i.e., 15 fallers, 15 non-fallers) would be sufficient to detect a statistically significant difference in performance-based test scores between fallers and non-fallers. To maximize the likelihood that at least 15 fallers were identified over the 6-month prospective reporting period, and to ensure minimum sample size requirements were met for additional planned psychometric analyses (e.g., test-retest reliability)<sup>26</sup>, a conservative sample of 60 LLP users was targeted<sup>27</sup>.

## **Procedures**

Sociodemographic, amputation, health, balance, and mobility measures were administered at baseline, along with three performance-based balance tests. The number of falls over the ensuing 6 months was recorded for each participant during monthly telephone calls.

## **Measurements**

### *Participant characterization measures*

Age and sex was collected via self-report, while amputation-related characteristics, as well as Medicare Functional Classification Level (MFCL)<sup>28</sup> were collected via interview. The number of co-morbidities was assessed using the Charlson Comorbidity Index (CCI)<sup>29</sup>, while perceived mobility and balance confidence were assessed using the Prosthetic Limb Users Survey of Mobility (PLUS-M)<sup>30</sup>, and Activities-specific Balance Confidence (ABC) scale<sup>31,32</sup> respectively. Comfort of the prosthesis worn by each participant was evaluated using the Socket Comfort Score (SCS)<sup>33,34</sup>.

### *Performance-based balance tests*

The TUG<sup>24</sup>, FSST<sup>25</sup>, and NBWT<sup>35</sup> were administered to participants in a randomized order, and scored according to standardized instructions (Appendix 1). Standardized procedures were created to ensure conformity in test administration and scoring between study sites. Each test was selected because it has been designed or used to assess fall risk in LLP users, and has evidence of construct and discriminant validity<sup>7,22,36,37</sup>, as well as reliability<sup>26,38,39</sup> in LLP users.

#### *Falls assessment*

The number of falls over the past 12 months was determined at baseline by asking participants, “*In the past year have you had any falls including a slip or trip in which you inadvertently lost your balance and landed on the ground or lower level?*”<sup>40</sup>. The number of falls over the next 6 months was assessed via monthly telephone interviews. At each call participants were asked “*Since we last saw you/spoke to you on [date], have you lost your balance and landed on the ground or lower level?*”.

#### **Statistical Analysis**

Departures from normality among continuous variables were evaluated using the Shapiro-Wilk test<sup>41</sup>. Descriptive analyses were performed to characterize the study sample and fall count data. Prospective falls data were summarized as the total number of falls in the sample, fall rate per person year, time to first fall, as well as the number and percentage of non-fallers, non-recurrent fallers, and recurrent fallers<sup>40</sup>. Kruskal-Wallis and Fisher’s exact (2xC) tests were run to test for differences in the distributions of continuous variables, and the frequency of dichotomous variables, between prospective fall categories (i.e., non-faller, non-recurrent faller,

and recurrent faller) respectively. The level of significance for these analyses was adjusted to  $\alpha \leq .0167$  to account for multiple comparisons between the three fall categories.

Negative binomial regression was used to develop predictive models, and test study hypotheses. Negative binomial regression has been recognized<sup>42-45</sup>, applied<sup>23, Manty:2010ev; 46-48</sup>, and chosen as the most suitable method for analyzing fall data due to the recurrent and dependent nature of falls in each participant, the type of data (i.e., counts), the non-normal distribution (i.e., Poisson), and the accompanying overdispersion of fall count data. Three predictive negative binomial regression models, a *performance-based test (PBT) model*, a *fall recall (FR) model*, and a *performance-based test plus fall recall (PBT+FR) model*, were developed to determine whether performance-based tests could predict future falls in unilateral LLP users, and if they improved fall prediction over fall recall alone. The initial *PBT+FR model* included TUG and FSST times, NBWT scores, the number of falls recalled in the past 12 months, and PLUS-M T-scores. We reduced the initial *PBT+FR model* by removing factors in a backward order<sup>49,50</sup>, preserving those variables with p-values  $\leq .15$  to ensure borderline relationships would not be overlooked<sup>2</sup>. Similar procedures, minus the inclusion of fall recall, were followed to develop the *PBT model*. The *FR model*, which included the number of falls LLP users recalled experiencing in the past 12 months, and PLUS-M T-scores, was based on prior research that found this model to significantly predict the number of future falls in established, unilateral LLP users<sup>23</sup>. Fall risk factors previously associated with a history of falls in LLP users<sup>5,6</sup> including age, level<sup>2,51</sup>, etiology, and time since amputation<sup>2</sup>, as well as hours of prosthetic wear<sup>1</sup>, and number of comorbidities<sup>2,22</sup> were excluded from the current analysis based on prior research that found they did not significantly contribute to the prediction of *future falls* in established, unilateral LLP

users<sup>23</sup>. Likelihood ratio chi-squared omnibus tests, and Pearson chi-squared goodness-of-fit tests (REF) were run on each iteration of the PBT, PBT+FR, and PBT+FR models to determine if they were significantly better than a null model (i.e., one with no predictors) at predicting the incidence of future falls, and to assess the fit of each model, respectively. These steps were repeated with each iteration until a statistically significant model (i.e., omnibus test  $p < .05$ ) was identified, which also contained factors whose contributions to predicting the incidence of future falls had  $p$ -values  $\leq .15$ . A likelihood ratio test was also run to compare the goodness-of-fit of, and select between the *FR* and *PBT+FR models*<sup>45</sup>. Regression coefficients ( $\beta$ ) of factors making statistically significant contributions to the prediction of future falls ( $p < .05$ ) in each of the final models were exponentiated to calculate incidence rate ratios (IRR), and translated into the estimated number of falls associated with that factor while holding all other factors in the final model constant at their means<sup>52</sup>. All statistical analyses were performed using SPSS v.27 (Chicago, IL).

## **Results**

Sixty unilateral LLP users were recruited and participated in the study (Table 1). 53 prospective falls were recorded over 6 months. The prospective annual fall rate per participant was 1.76 falls, and the median time to first prospective fall was 3 months (IQR: 3.25 months). 53% (32/60), 28% (17/60), and 19% (11/60) of study participants reported no falls, falling once, or falling at least twice during the prospective 6 month period, respectively. The negative binomial regression model found to best predict future falls in the full sample (i.e., transtibial and transfemoral prosthesis users) was equally effective at predicting future falls in transtibial prosthesis users alone. In contrast, when applied to participants with transfemoral amputation,

the model was no better than a null model (i.e., one with no predictors) at predicting the number of future falls. Further analysis revealed that no combination of independent variables significantly predicted the number of future falls in participants with transfemoral amputation. Mann-Whitney U and Fisher's exact (2xC) tests revealed that with the exception of PLUS-M T-scores, no significant differences in continuous or dichotomous demographic, amputation, mobility, balance, or health-related characteristics were identified between participants with transtibial and transfemoral amputation (Table 1). The remainder of the results will therefore focus on participants with unilateral transtibial amputation.

[INSERT TABLE 1 HERE]

Forty-five individuals with unilateral transtibial amputation were recruited and participated in the study (Table 2). Age and PLUS-M T-scores were normally distributed ( $W \geq 0.953$ ,  $p \geq .068$ ), while the remaining amputation, demographic, mobility, and health-related characteristics were non-normally distributed ( $W \leq .815$ ,  $p \leq .03$ ) (Table 2). 38 prospective falls were recorded over 6 months. The prospective annual fall rate per participant with a transtibial amputation was 1.69 falls, and the median time to first prospective fall was 3 months (IQR: 4 months). 55% (25/45), 29% (13/45), and 16% (7/45) of study participants with unilateral transtibial amputation reported no falls, falling once, or falling at least twice during the 6-month prospective period, respectively. Demographic, amputation, mobility, balance, and health-related characteristics were not significantly different across non-fallers, non-recurrent fallers, and recurrent fallers (Table 2).

[INSERT TABLE 2 HERE]

NBWT scores (mean $\pm$ 95% CI) (0.47 $\pm$ 0.13) were normally distributed ( $W \geq .961$ ,  $p \geq .133$ ), while TUG (8.75 $\pm$ 3.36 seconds) and FSST times (9.1 $\pm$ 2.3 seconds), as well as the number of falls in the past 12-months, were non-normally distributed ( $W = .713-.910$ ,  $p < .008$ ). Kruskal-Wallis tests revealed that the distribution of performance-based test scores, and the number of falls recalled in the past 12 months were not significantly different across prospective fall categories ( $U \geq .823$ ,  $p \geq .148$ ) (Table 3).

[INSERT TABLE 3 HERE]

Prospective 6-month fall counts were non-normally distributed ( $W = .552$ ,  $p \leq .001$ ), and overdispersed, confirming the suitability of negative binomial regression. After reduction, the final *performance-based test plus fall recall (PBT+FR) model* retained TUG and FSST times, as well as the number of falls recalled over the past 12 months (Supplemental Table 2). In contrast, the final *performance-based test (PBT) model* retained only PLUS-M (Supplemental Table 1), while the *fall recall (FR) model* retained the number of falls recalled in the past 12 months (Supplemental Table 3). Likelihood ratio chi-squared omnibus test revealed that the *FR model* and *PBT+FR model* were significantly better than a null model (i.e., one with no factors) at predicting the number of falls over the next 6 months, while the *PBT model was not (FR model:  $X^2(1) = 7.26$ ,  $p = .007$ , AIC=111.21; *PBT+FR model:  $X^2(3) = 11.6$ ,  $p = .009$ , AIC=110.87; *PBT model:  $X^2(1) = 3.83$ ,  $p = .05$ , AIC=114.64)*. Chi-squared goodness-of-fit tests revealed that there were no significant difference between the observed and expected number of falls in the prospective period for any of the final models, implying that each model fit the prospective fall**

count data (*FR model*:  $X^2(43) = 45.45$ ,  $p = .20$ ; *PBT model*:  $X^2(43) = 52.18$ ,  $p = .10$ ; *PBT+FR model*:  $X^2(41) = 36.12$ ,  $p = .20$ ). Within the final *PBT+FR model*, TUG times and fall recall, but not FSST times, were statistically significant predictors for the number of falls over the next 6-months in unilateral transtibial prosthesis users (TUG:  $p = .044$ ; fall recall:  $p = .009$ ) (Table 4). In the final *FR model*, the number of falls recalled in the past year was also a statistically significant predictor of the number of falls over the next 6 months (fall history:  $p = .014$ ) (Table 4). The likelihood ratio test revealed that the larger *PBT+FR model* provided a significant improvement in fit to the prospective fall count data over the smaller *FR model*  $X^2(1) = 4.342$ ,  $p < .05$ .

[INSERT TABLE 4 HERE]

For both the final *FR* and *PBT+FR models*, the modeled relationship predicts a minor increase in the number of future falls as the number of falls recalled over the past 12 months increases, and all other factors are held at their sample means (Figure 1A and B). In comparison, an increase in TUG time within the *PBT+FR model*, while holding all other factors at their sample means, predicts a notable increase in the number of future falls (Figure 1C). While holding all other factors in the *PBT+FR model* at their sample means<sup>52</sup>, a TUG score greater than 10.97 seconds was associated with 1 or more falls in the next 6 months among unilateral transtibial prosthesis users (Figure 1C). Predicted falls increased sharply as the TUG score increased beyond 11.5 seconds. An increase in TUG scores from 7.51 to 13.46 seconds (10<sup>th</sup> to 90<sup>th</sup> percentile), when FSST and fall history were held at their means, resulted in a predicted increase of 2.25 falls in the next 6 months (Figure 1C). Viewed jointly within the final *PBT+FR model* [EQ. 1], the combined effect of the number of falls recalled, TUG times, and FSST times

on predicting the number of future falls was statistically significant (Table 4).

[INSERT FIGURE 1 HERE]

$$\# \text{ Future falls} = \exp [(0.202) * (\# \text{ falls recalled past year}) + (0.370) * (\text{TUG}) + (-0.154) * (\text{FSST}) - 2.90] \text{ [EQ.1]}$$

## **Discussion**

The objective of this study was to test the hypothesis that a combination of performance-based balance tests could predict the number of future falls in established unilateral LLP users. Results failed to support the primary hypothesis. No combination of performance-based tests included in this study (i.e., the *PBT model*) could alone predict the number of future falls in established unilateral transtibial prosthesis users. Rather, a combination of FSST and TUG times, along with the number of falls recalled from the past 12 months (i.e., *PBT+FR model*) was required to predict the number of future falls in established unilateral transtibial prosthesis users. Further, and in support of the secondary study hypothesis, the *PBT+FR model* provided a significantly better fit to the prospective falls data than the *FR model*, suggesting that the addition of performance-based test scores to fall recall improved the prediction of future falls in established, unilateral transtibial prosthesis users over fall recall alone<sup>23</sup>. Given the ease with which fall recall, FSST, and TUG times can be collected, the resulting predictive model may serve as a suitable method for clinicians to predict the number of falls over the next 6 months in established unilateral transtibial prosthesis users. The accompanying incidence rate ratios (IRR) and estimated relationship curves provide researchers with indices and visualizations to understand the theoretical association between the studied factors and future falls in established unilateral transtibial prosthesis users. This is the first study, to our knowledge, to examine

whether performance-based balance tests can predict future falls in established unilateral transtibial prosthesis users.

A mismatch between the task(s) assessed by a performance-based test, and the nature of falls recorded over a study period, may influence the observed level of discriminant and/or predictive validity for a test. Where previous retrospective studies found the NBWT, FSST, and TUG to successfully discriminate between LLP users with and without a history of falls<sup>7,13</sup>, the TUG was the only one to serve as a statistically significantly predictor of future falls in the current study. Beyond differences in sample characteristics, study design<sup>19</sup>, methods<sup>53</sup>, and data analysis<sup>42-44</sup>, the effectiveness of the TUG in both the current prospective and earlier retrospective studies<sup>7,13</sup> may be due the range of tasks included in the TUG (i.e., transfer, walking, and turning), alongside the collection and prediction of “all-cause” falls (i.e., no restrictions were placed on fall direction, source of instability, activity at the time of a fall, or severity of the fall). By including a range of tasks, the TUG may be better suited to assess fall risk across this broad range of falls, (i.e., “all-cause falls”). In contrast, the NBWT and FSST focus on specific aspects of balance control, which may make them better suited for predicting specific types of falls. The NBWT for example, assesses medial-lateral control of the center-of-mass<sup>35,54</sup>, and may therefore be better suited to predict the incidence of lateral falls attributed to intrinsic sources of center-of-mass instability, such as in older transtibial prosthesis users<sup>55</sup>. Additional research is warranted to determine whether specific performance-based tests are better suited to predict specific types of falls versus all-cause falls in LLP users<sup>3,55</sup>.

The current results provide new clinical insight into the assessment of fall risk in established unilateral transtibial prosthesis users. Consistent with our findings, Dite et al., (2007) reported that the TUG provided good identification of unilateral transtibial prosthesis users at risk for falls<sup>22</sup>. The two studies differ however in sample composition, analysis of fall events, as well as their clinical application and interpretation. Dite et al., (2007) focused on older, newer transtibial prosthesis users (i.e.,  $\leq 6$  months post amputation), where as the current study concentrated on established and slightly younger transtibial prosthesis users. Between the two studies, it could be inferred that the TUG has broad appeal to predict “all-cause falls” in unilateral transtibial prosthesis users regardless of time since amputation and possibly age. Dite et al., (2007) treated falls as a categorical binary outcome (i.e., multiple faller or non-multiple faller), while the current study treated falls as counts, yielding a continuous outcome (i.e., number of falls in the next 6 months). Furthermore, Dite et al., (2007) studied the predictive validity of the TUG independent of other potentially confounding variables, while it was part of a multivariate predictive model in the current study. Therefore, where Dite et al., (2007) offered a discrete cut-off threshold of  $\geq 19$  seconds on the TUG alone as a means for clinicians to make a binary determination of whether a unilateral transtibial prosthesis user is at risk for multiple falls in the next 6 months, the *PBT+FR model* in the current study enables clinicians to predict a continuous and more specific outcome, the number of expected falls over the ensuing 6 months. By administering the TUG and FSST, and asking about the number of falls recalled over the past year, clinicians can use the *PBT+FR model* (i.e., equation 1) to make predictions about the number of fall(s) that might be expected in the next 6 months for individual unilateral transtibial prosthesis users. For example, an established, unilateral transtibial prosthesis user with a 11.8 s

TUG time, a 9.7 s FSST time, and who recalls having fallen twice in the past year, would be predicted to fall 1.46 times over the next 6 months.

$\# \text{ Future falls} = \exp [(0.202) * (\# \text{ falls recalled past year}) + (0.370) * (\text{TUG}) + (-0.154) * (\text{FSST}) - 2.90]$

$\# \text{ Future falls} = \exp [(0.202) * (2) + (0.370) * (11.8) + (-0.154) * (9.7) - 2.90]$

$\# \text{ Future falls} = \exp [(0.202) * (2) + (0.370) * (11.8) + (-0.154) * (9.7) - 2.90]$

$\# \text{ Future falls} = \exp [.377]$

$\# \text{ Future falls} = 1.46$

The current results provide researchers with indices and visualizations to understand relationships between potential fall risk factors and future falls in unilateral transtibial prosthesis users. Exponentiating the TUG regression coefficient ( $\beta$ ) in the *PBT+FR model* (i.e.,  $\beta$ : .370) to express it as an incidence rate ratio (i.e., IRR: 1.45) indicates that for every 1 unit increase in TUG (i.e., 1 second), and holding all other factors in the model at their sample mean, there is a compounding 1.45 times or 45% increase in the likelihood of a fall over the next 6-months. For example, a unilateral transtibial prosthesis user with a TUG score of 12.25 s would have a 45% greater likelihood of falling in the next 6 months compared a unilateral transtibial prosthesis user with a TUG score of 11.25 seconds, and a 110% greater likelihood of falling in the next 6 months compared to a unilateral transtibial prosthesis user with a TUG score of 10.25 s, all other factors being equal. In contrast, the IRR for fall history, in either the *FR model* (i.e., 1.19) or the *PBT+FR model* (i.e., 1.22), indicates that for every additional fall in the past year, there is a compounding 1.19, or 19%, and 1.22 or 22% increase in the likelihood of another fall over the next 6 months, respectively. For example, a unilateral transtibial prosthesis user reporting one fall in the past year would have a 19% likelihood of falling in the next 6 months, whereas a

history of two falls in the past year would increase the likelihood that they would fall again in the next 6 months to 1.42, or 42%. These differences in the compounding rates at which fall risk are expressed for TUG times and fall history are reflected in the slopes of their respective estimated relationship curves (Figure 1). The steeper slope of the TUG curve suggests that fall risk is more susceptible to a change in TUG time than fall history, providing additional support to the secondary hypothesis that performance-based tests improve fall risk assessment in unilateral transtibial prosthesis users over fall history alone.

The inability of any combination of independent variable(s) in the current study to predict the number of future falls in participants with transfemoral amputation suggests that the assessment of fall risk may need to be stratified by level of amputation. The smaller sample of transfemoral prosthesis users (n=15) may have restricted the prediction of future falls by transfemoral prosthesis users in this study, however, a misalignment between balance ability, as measured by performance-based tests, and the circumstances under which transfemoral prosthesis users fall, may limit the accuracy with which performance-based tests can assess fall risk in transfemoral prosthesis users. Lower scores on performance-based balance tests are assumed to reflect poorer balance ability, which in turn is interpreted as greater fall risk. This anticipated relationship between balance ability and fall events did not materialize among transfemoral prosthesis users in this study. Transfemoral prosthesis users performed significantly worse on all three performance-based tests than transtibial prosthesis users (Table 1), but they did not incur a significantly greater number of future falls. From this one may infer that transfemoral prosthesis users may have different balance deficits than transtibial prosthesis users, and that these deficits were not assessed by the tests administered in the current. The assessment of balance ability in

this study may not reflect the activities, causes, or reasons transfemoral prosthesis users fall. As a result, performance-based tests that are suitable for inferring fall risk in transtibial prosthesis users may not be equally suitable for transfemoral prosthesis users. Alternatively, each of the performance-based tests may assess constructs beyond balance ability alone, and as a result, “worse performance” on these tests may speak to deficits in other outcomes, rather than balance ability alone. Additional research probing for differences in epidemiological, clinical, and biomechanical correlates of balance and falls between transtibial and transfemoral prosthesis users is therefore warranted to help identify tasks that better align measured balance ability with fall risk in transfemoral prosthesis users.

#### *Study Limitations*

The results of this study should be interpreted and applied in light of several limitations. First, the sample composition was slightly skewed towards younger, and non-dysvascular transtibial prosthesis users who had been using a prosthesis for at least one year<sup>56-59</sup>. While previous research has suggested that future falls are not associated with amputation etiology<sup>2,8,23</sup>, additional research is required to determine if these results extend to older, dysvascular, bilateral, and acute LLP users (i.e., < 1 year post amputation).

Next, fall recall, used to ascertain fall history, may be susceptible to memory decay over time, and cause LLP users to either under or over-report falls<sup>60,61</sup>. The accuracy of recall does not, however, affect whether the number of falls recalled in the past year can serve as a predictor future falls. Additionally, querying LLP users about the number of falls they recall over the past year, as a way to estimate future fall risk, is consistent with how this information would be

collected in clinical practice. We also selected a longer 1-year recall period, because it has been shown to minimize fall recall decay compared to shorter timeframes (i.e., 3-6 month) in older adults<sup>53,62</sup>.

We did not control for exposure (i.e., the amount of time each participant was active per day). Physical activity levels may have therefore influenced prospective fall counts. Methods for reporting fall data in the current study were however based on a consensus statement that recommends against adjusting for physical activity when presenting and analyzing fall data<sup>40</sup>. Nonetheless, measures of activity and participation should be considered and explored in future studies as ways to assess fall risk, and/or stratify study samples.

Additional fall risk factors in LLP users including strength<sup>10</sup>, protective stepping<sup>63,64</sup>, number of medications<sup>12</sup>, and sense of vibration<sup>11</sup> may improve model performance. The current study only evaluated three performance-based tests. Future research to assess the predictive validity of additional performance-based tests that have proven useful in other clinical populations (e.g., Functional Reach Test, Fullerton Advanced Balance Scale, Five-times Sit to Stand) is warranted.

This study sought to predict all-cause falls, regardless of severity. The prediction of injurious falls is an important next step<sup>4,17,20</sup>, likely to require additional factors beyond performance-based test scores (e.g., strength, endurance, reaction time etc), as well as a longer prospective reporting period to ensure a sufficient number of injurious falls are recorded. Finally, additional research to validate the efficacy of the *PBT+FR model* in an independent sample of unilateral transtibial prosthesis users is recommended.

459

460 **Conclusion**

461       The primary objective of this study was to determine whether select performance-based  
462 balance tests could predict future falls in unilateral LLP users. Using a prospective design to  
463 develop a predictive model, this study found that the combination of TUG and FSST times, plus  
464 the number of falls recalled over the past year could predict the number of falls over the next 6  
465 months in established, unilateral transtibial prosthesis users. Additional prospective studies are  
466 required to validate the proposed model, and evaluate additional fall-risk factors and  
467 performance-based tests. Given the limited options available to predict falls in LLP users, the  
468 resulting predictive model may serve as a suitable method for clinicians to predict the number of  
469 falls over the next 6 months in established unilateral transtibial prosthesis users.

470

471 Main text word count: 4750/4500

472

473 **Declaration of Conflicting Interest**

474 The authors declare that there is no conflict of interest.

475

476

477

478

479

480

481

## 482 **References**

- 483 1. Kulkarni J, Wright S, Toole C, Morris J, Hiron R. Falls in patients with lower limb amputations: prevalence  
484 and contributing factors. *Physiotherapy*. 1996;82:130–6.
- 485 2. Miller WC, Speechley M, Deathe B. The prevalence and risk factors of falling and fear of falling among  
486 lower extremity amputees. *Arch Phys Med Rehabil*. 2001;82(8):1031-1037.
- 487 3. Kim J, Major MJ, Hafner B, Sawers A. Frequency and circumstances of falls reported by ambulatory  
488 unilateral lower limb prosthesis users: a Secondary Analysis. *PM&R*. 2019;11(4):344-353.
- 489 4. Mundell B, Kremers HM, Visscher S, Hoppe K, Kaufman K. Direct medical costs of accidental falls for  
490 adults with transfemoral amputations. *Prosthet Orthot Int Sage Publ Ltd*. 2017;41(6):564-570.
- 491 5. Hunter SW, Batchelor F, Hill K, Hill A, Mackintosh S, Payne M. Risk factors for falls in people with a  
492 lower limb amputation: a systematic review. *PM&R*. 2017;9(2):170-180.
- 493 6. Steinberg N, Gottlieb A, Siev-Ner I, Plotnik M. Fall incidence and associated risk factors among people  
494 with a lower limb amputation during various stages of recovery - a systematic review. *Disabil Rehabil*.  
495 2018;16:1–10.
- 496 7. Sawers A, Hafner B. Validation of the Narrowing Beam Walking Test in lower limb prosthesis users. *Arch*  
497 *Phys Med Rehabil*. 2018;99:1491–1.
- 498 8. Wong CK, Chen CC, Blackwell WM, Rahal RT, Benoy SA. Balance ability measured with the Berg balance  
499 scale: a determinant of fall history in community-dwelling adults with leg amputation. *J Rehabil Med*.  
500 2015;47:80–6.
- 501 9. Major MJ, Fatone S, Roth EJ. Validity and reliability of the Berg Balance Scale for community-dwelling  
502 persons with lower-limb amputation. *Arch Phys Med Rehabil*. 2013;94(11):2194-2202.
- 503 10. Vanicek N, Strike S, McNaughton L, Polman R. Gait patterns in transtibial amputee fallers vs. non-fallers:  
504 Biomechanical differences during level walking. *Gait Posture*. 2009;29:415–20.
- 505 11. Quai TM, Brauer SG, Nitz JC. Somatosensation, circulation and stance balance in elderly dysvascular  
506 transtibial amputees. *Clin Rehabil*. 2005;19:668–76.
- 507 12. Pauley T, Devlin M, Heslin K. Falls sustained during inpatient rehabilitation after lower limb amputation:  
508 prevalence and predictors. *Am J Phys Med Rehabil*. 2006;85(6):521-545.
- 509 13. Sawers A, Hafner BJ. Using clinical balance tests to assess fall risk among established unilateral lower limb  
510 prosthesis users: cutoff scores and associated validity indices. *PM&R*. 2020;12(1):16-25.
- 511 14. Parker K, Hanada E, Adderson J. Gait variability and regularity of people with transtibial amputations. *Gait*  
512 *Posture*. 2013;37:269–73.
- 513 15. Yu JC, Lam K, Nettel-Aguirre A, Donald M, Dukelow S. Incidence and risk factors of falling in the  
514 postoperative lower limb amputee while on the surgical ward. *PM&R*. 2010;2:926–34.
- 515 16. Vu K, Payne MWC, Hunter SW, Viana R. Risk Factors for Falls in Individuals With Lower Extremity  
516 Amputations During the Pre-Prosthetic Phase: A Retrospective Cohort Study. *PM&R*. 2019;11:828–33.
- 517 17. Chihuri S, Wong CK. Factors associated with the likelihood of fall-related injury among people with lower  
518 limb loss. *Inj Epidemiol*. 2018;5:42.

- 519 18. Clemens SM, Gailey RS, Bennett CL, Pasquina PF, Kirk-Sanchez NJ, Gaunaud IA. The Component  
520 Timed-Up-and-Go test: the utility and psychometric properties of using a mobile application to determine  
521 prosthetic mobility in people with lower limb amputations. *Clin Rehabil*. 2018;32:388–97.
- 522 19. Muir SW, Speechley M. Establishing predictive validity of the Fullerton Advanced Balance scale.  
523 2010;91:1147–authorreply1147–8.
- 524 20. Wong CK, Chihuri ST, Li G. Risk of fall-related injury in people with lower limb amputations: A  
525 prospective cohort study. *J Rehabil Med*. 2016;48:80–5.
- 526 21. Gooday HMK, Hunter J. Preventing falls and stump injuries in lower limb amputees during inpatient  
527 rehabilitation: completion of the audit cycle. *Clin Rehabil*. 2004;18:379–90.
- 528 22. Dite W, Connor HJ, Curtis HC. Clinical identification of multiple fall risk early after unilateral transtibial  
529 amputation. *Arch Phys Med Rehabil*. 2007;88:109–14.
- 530 23. Tobaigy M, Hafner BJ, Sawers A. Recalled number of falls in the prior year, combined with perceived  
531 mobility, predicts the incidence of future falls in established unilateral lower limb prosthesis users. (*in*  
532 *review*).
- 533 24. Podsiadlo D, Richardson S. The timed “Up & Go”: a test of basic functional mobility for frail elderly  
534 persons. *J Am Geriatr Soc*. 1991;39:142–8.
- 535 25. Dite W, Temple VA. A clinical test of stepping and change of direction to identify multiple falling older  
536 adults. *Arch Phys Med Rehabil*. 2002;83:1566–71.
- 537 26. Sawers A, Kim J, Balkman G, Hafner BJ. Interrater and test-retest reliability of performance-based clinical  
538 tests administered to established users of lower limb prostheses. *Physical Therapy*. 2020;100:1206–16.
- 539 27. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Medicine*. 2000;30:1–15.
- 540 28. Palmento Government Benefits Administrators. Lower limb prsothesis. DMERC Medicare Advis.
- 541 29. Chaudhry S, Jin L, Meltzer D. Use of a self-report-generated Charlson Comorbidity Index for predicting  
542 mortality. *Medical Care*. 2005;43:607–15.
- 543 30. Hafner BJ, Gaunaud IA, Morgan SJ, Amtmann D, Salem R, Gailey RS. Construct validity of the Prosthetic  
544 Limb Users Survey of Mobility (PLUS-M) in adults with lower limb amputation. *Arch Phys Med Rehabil*.  
545 2017;98(2):277-285.
- 546 31. Sakakibara BM, Miller WC, Backman CL. Rasch analyses of the Activities-specific Balance Confidence  
547 Scale with Individuals 50 years and older with lower-limb amputations. *Arch Phys Med Rehabil*.  
548 2011;92(8):1257-1263.
- 549 32. Powell LE, Myers AM. The Activities-specific Balance Confidence (ABC) Scale. *J Gerontol A Biol Sci*  
550 *Med Sci*. 1995;50A(1):M28-M34.
- 551 33. Hanspal RS, Fisher K, Nieveen R. Prosthetic socket fit comfort score. *Disabil Rehabil*. 2003;25:1278–80.
- 552 34. Hafner BJ, Morgan SJ, Askew RL, Salem R. Psychometric evaluation of self-report outcome measures for  
553 prosthetic applications. *J Rehabil Res Dev*. 2016;53:797–812.
- 554 35. Sawers A, Hafner BJ. Narrowing beam-walking is a clinically feasible approach for assessing balance ability  
555 in lower-limb prosthesis users. *J Rehabil Med*. 2018;50:457–64.

- 556 36. Schoppen T, Boonstra A, Groothoff JW. The Timed “up and go” test: reliability and validity in persons with  
557 unilateral lower limb amputation. *Arch Phys Med Rehabil*. 1999;80:825–8.
- 558 37. Deathe AB, Miller WC. The L test of functional mobility: Measurement properties of a modified version of  
559 the timed “Up & Go” test designed for people with lower-limb amputations. *Physical Therapy*.  
560 2005;85:626–35.
- 561 38. Resnik L, Borgia M. Reliability of Outcome Measures for People With Lower-Limb Amputations:  
562 Distinguishing True Change From Statistical Error. *Physical Therapy*. 2011;91:555–65.
- 563 39. Jayakaran P, Johnson GM, Sullivan SJ. Reliability and Concurrent Validity of the Step Quick Turn Test in  
564 Older Persons with a Unilateral Transtibial Amputation. *Am J Phys Med Rehabil*. 2011;90:798–804.
- 565 40. Lamb SE, Jörstad-Stein EC, Hauer K, Becker C, on behalf of the Prevention of Falls Network Europe and  
566 Outcomes Consensus Group. Development of a common outcome data set for fall injury prevention trials:  
567 The prevention of falls network Europe consensus. *J Am Geriatr Soc*. 2005;53:1618–22.
- 568 41. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika*.  
569 1965;52:591.
- 570 42. Aeberhard WH, Cantoni E, Heritier S. Robust inference in the negative binomial regression model with an  
571 application to falls data. *Biometrics*. 2014;70:920–31.
- 572 43. Robertson MC, Campbell AJ, Herbison P. Statistical analysis of efficacy in falls prevention trials. *J*  
573 *Gerontol. A Biol Sci Med Sci*. 2005;60:530–4.
- 574 44. Donaldson MG, Sobolev B, Cook WL, Janssen PA, Khan KM. Analysis of recurrent events: a systematic  
575 review of randomised controlled trials of interventions to prevent falls. *Age and Ageing*. 2009;38:151–5.
- 576 45. Ullah S, Finch CF, Day L. Statistical modelling for falls count data. *Accid Anal Prev*. 2010;42:384–92.
- 577 46. Simpson LA, Miller WC, Eng JJ. Effect of stroke on fall rate, location and predictors: a prospective  
578 comparison of older adults with and without stroke. *PLoS ONE*. 2011;6:e19431.
- 579 47. Gill DP, Zou GY, Jones GR, Speechley M. Comparison of regression models for the analysis of fall risk  
580 factors in older veterans. *Ann Epidemiol*. 2009;19:523–30.
- 581 48. Davis JC, Best JR, Khan KM, Dian L, Lord S, Delbaere K, et al. Slow processing speed predicts falls in  
582 older adults with a falls history: 1-Year Prospective Cohort Study. *J Am Geriatr Soc*. 2017;65:916–23.
- 583 49. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating  
584 assumptions and adequacy, and measuring and reducing errors. *Statist. Med*. 1996;15:361–87.
- 585 50. Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in  
586 multivariable analysis. *J Clin Epidemiol*. 1996;49:907–16.
- 587 51. Gauthier-Gagnon C, Grise MC, Potvin D. Enabling factors related to prosthetic use by people with  
588 transtibial and transfemoral amputation. *Arch Phys Med Rehabil*. 1999;80:706–13.
- 589 52. Long JS, Freese J. Regression Models for Categorical Dependent Variables Using Stata, Second Edition.  
590 Stata Press; 2006.
- 591 53. Cummings SR, Nevitt MC, Kidd S. Forgetting falls. The limited accuracy of recall of falls in the elderly. *J*  
592 *Am Geriatr Soc*. 1988;36:613–6.

593 54. Sawers A, Ting LH. Beam walking can detect differences in walking balance proficiency across a range of  
594 sensorimotor abilities. *Gait Posture*. 2015;41:619–23.

595 55. Anderson CB, Miller MJ, Murray AM, Fields TT, So NF, Christiansen CL. Falls after dysvascular  
596 transtibial amputation: A secondary analysis of falling characteristics and reduced physical performance.  
597 *PM R*. 2020 (online ahead of print)

598 56. Pezzin LE, Dillingham TR, MacKenzie EJ, Ephraim P, Rossbach P. Use and satisfaction with prosthetic  
599 limb devices and related services. *Arch Phys Med Rehabil*. 2004;85(5):723-729.

600 57. Ephraim PL, Wegener ST, MacKenzie EJ, Dillingham TR, Pezzin LE. Phantom pain, residual limb pain,  
601 and back pain in amputees: results of a national survey. *Arch Phys Med Rehabil*. 2005;86(10):1910-1919.

602 58. Ziegler-Graham K, MacKenzie EJ, Ephraim PL, Travison TG, Brookmeyer R. Estimating the prevalence of  
603 limb loss in the United States: 2005 to 2050. *Arch Phys Med Rehabil*. 2008;89(3):422-429.

604 59. Wurdeman SR, Stevens P, Campbell J. Mobility analysis of Amputees II: comorbidities and mobility in  
605 lower limb prosthesis users. *Am J Phys Med Rehabil*. 2018;97(11):782-788.

606 60. Mackenzie L, Byles J, D'Este C. Validation of self-reported fall events in intervention studies. *Clin Rehabil*.  
607 2006;20:331–9.

608 61. Ganz DA, Higashi T, Rubenstein LZ. Monitoring Falls in Cohort Studies of Community-Dwelling Older  
609 People: Effect of the Recall Interval. *J Am Geriatr Soc*. 2005;53:2190–4.

610 62. Hale WA, Delaney MJ, Cable T. Accuracy of patient recall and chart documentation of falls. *J Am Board*  
611 *Fam Pract*. 1993;6:239–42.

612 63. Crenshaw JR, Kaufman KR, Grabiner MD. Compensatory-step training of healthy, mobile people with  
613 unilateral, transfemoral or knee disarticulation amputations: A potential intervention for trip-related falls.  
614 *Gait Posture*. 2013;38:500–6.

615 64. Kaufman KR, Wyatt MP, Sessoms PH, Grabiner MD. Task-specific fall prevention training is effective for  
616 warfighters with transtibial amputations. *Clin Orthop Relat Res*. 2014;472:3076–84.

617

618

619

620

621

622

623

624

625

## **List of Tables**

Table 1. Participant characteristics stratified by level of amputation

Table 2. Characteristics of unilateral transtibial prosthesis users stratified by prospective fall category

Table 3. Performance-based test scores among transtibial prosthesis users stratified by prospective fall category

Table 4. Negative binomial regression models to predict falls in unilateral transtibial prosthesis users

## **Figure Captions**

Figure 1 – Estimated relationship between factors retained in the final models and the predicted number of falls in the next 6 months for unilateral transtibial prosthesis users. Each plot displays predicted number of falls at different levels of the predictor while holding the remaining variable(s) in the model at their sample mean. The predicted number of falls is plotted for values of the final model predictors (A: Fall Recall Model - number of falls recalled in past 12 months; B: Performance-Based Test + Fall Recall Model - number of falls recalled in past 12 months; B: Performance-Based Test + Fall Recall Model – Timed Up and Go time). Sample mean indicated by the arrow (recalled falls: 1.52; Timed Up and Go time: 9.89s).

Table 1. Participant characteristics stratified by level of amputation

|                               |                 | All Participants<br>(n=60) | Transtibial<br>(n=45) | Transfemoral<br>(n=15) | p-value            |
|-------------------------------|-----------------|----------------------------|-----------------------|------------------------|--------------------|
|                               |                 | # of participants (%)      |                       |                        |                    |
| Etiology                      | Dysvascular     | 25 (42%)                   | 20 (33%)              | 5 (8%)                 | .450 <sup>a</sup>  |
|                               | Non-dysvascular | 35 (58%)                   | 25 (42%)              | 10 (17%)               |                    |
| MFCL                          | K1 and K2       | 18 (30%)                   | 11 (18%)              | 7 (12%)                | .104 <sup>a</sup>  |
|                               | K3 and K4       | 42 (70%)                   | 34 (57%)              | 8 (13%)                |                    |
| Sex                           | Male            | 42 (70%)                   | 33 (55%)              | 9 (15%)                | .329 <sup>a</sup>  |
|                               | Female          | 18 (30%)                   | 12 (20%)              | 6 (10%)                |                    |
|                               |                 | Mean (95% CI)              |                       |                        |                    |
| Age (years)                   |                 | 55.1 (51.0, 59.2)          | 54.7 (49.8, 59.5)     | 56.4 (47.6, 65.2)      | .516 <sup>b</sup>  |
| CCI                           |                 | 0.95 (0.65, 1.25)          | 0.96 (0.45, 1.47)     | 1.12 (0.65, 1.66)      | .599 <sup>b</sup>  |
| Time since amputation (years) |                 | 16.0 (12.2, 19.9)          | 14.0 (9.9, 18.0)      | 21.2 (12.5, 31.8)      | .179 <sup>b</sup>  |
| Prosthesis use (hours/day)    |                 | 13.9 (13.0, 14.8)          | 13.9 (12.8, 15.0)     | 13.8 (12.1, 15.5)      | .763 <sup>b</sup>  |
| SCS                           |                 | 7.6 (7.1, 8.1)             | 7.7 (7.1, 8.3)        | 7.5 (6.5, 8.6)         | .764 <sup>b</sup>  |
| PLUS-M (T-score)              |                 | 56.1 (54.3, 57.9)          | 57.2 (55.0, 59.3)     | 52.8 (50.2, 55.4)      | .038 <sup>b*</sup> |
| ABC Scale (0-4)               |                 | 3.07 (2.88, 3.25)          | 3.07 (2.84, 3.30)     | 3.06 (2.72, 3.40)      | .701 <sup>b</sup>  |
| FSST (time, s)                |                 | 10.0 (8.8, 11.1)           | 9.1 (8.0, 10.3)       | 12.6 (9.8, 15.5)       | .003 <sup>b*</sup> |
| NBWT (norm. distance)         |                 | .42 (.37, .47)             | .47 (.40, .53)        | .27 (.21, .32)         | .001 <sup>b*</sup> |
| TUG (time, s)                 |                 | 10.2 (9.6, 10.9)           | 9.8 (9.2, 10.4)       | 11.7 (9.9, 13.4)       | .019 <sup>b*</sup> |
| Falls recalled 12 months      |                 | 1.52 (0.88, 2.15)          | 1.27 (.56, 1.97)      | 2.27 (.79, 3.74)       | .063 <sup>b</sup>  |
| Prospective falls 6 months    |                 | .90 (.50, 1.30)            | .84 (.37, 1.32)       | 1.07 (.34, 1.80)       | .435 <sup>b</sup>  |

ABC: Activities-specific Balance Confidence scale; CCI: Charleston Co-morbidity Index; IQR: Interquartile range; MFCL: Medicare Functional Classification Level; PLUS-M: Prosthetic Limb Users Survey of Mobility; SCS: Socket Comfort Score

<sup>a</sup> Chi-square test were run to compare categorical variables between participants with transtibial and transfemoral amputation

<sup>b</sup> Mann-Whitney U tests were run to compare continuous variables between participants with transtibial and transfemoral amputation

\* Participants with transtibial amputation performed/scored significantly better than participants with transfemoral amputation (p < .05).

Table 2. Characteristics of unilateral transtibial prosthesis users stratified by prospective fall category

|                               |                 | Non-fallers<br>(n=25) | Non-recurrent fallers<br>(n=13) | Recurrent fallers<br>(n=7) | p-value           |
|-------------------------------|-----------------|-----------------------|---------------------------------|----------------------------|-------------------|
|                               |                 | # of participants (%) |                                 |                            |                   |
| Etiology                      | Dysvascular     | 12 (27%)              | 3 (7%)                          | 5 (11%)                    | .100 <sup>a</sup> |
|                               | Non-dysvascular | 13 (29%)              | 10 (22%)                        | 2 (4%)                     |                   |
| MFCL                          | K1 and K2       | 6 (13%)               | 3 (8%)                          | 2 (4%)                     | .961 <sup>a</sup> |
|                               | K3 and K4       | 19 (42%)              | 10 (22%)                        | 5 (11%)                    |                   |
| Sex                           | Male            | 20 (44%)              | 8 (19%)                         | 5 (11%)                    | .471 <sup>a</sup> |
|                               | Female          | 5 (11%)               | 5 (11%)                         | 2 (4%)                     |                   |
|                               |                 | Mean (95% CI)         |                                 |                            |                   |
| Time since amputation (years) |                 | 13.1 (8.4, 17.8)      | 19.6 (8.7, 30.5)                | 6.6 (2.3, 10.8)            | .195 <sup>b</sup> |
| Prosthesis use (hours/day)    |                 | 14.5 (13.2, 15.7)     | 14.2 (12.5, 16.0)               | 11.5 (6.0, 17.0)           | .536 <sup>b</sup> |
| PLUS-M (T-score)              |                 | 58.0 (55.1, 60.9)     | 57.1 (52.6, 61.6)               | 54.5 (47.4, 61.5)          | .490 <sup>b</sup> |
| ABC Scale (0-4)               |                 | 3.15 (2.83, 3.46)     | 3.07 (2.65, 3.50)               | 2.78 (2.01, 3.55)          | .462 <sup>b</sup> |
| Age (years)                   |                 | 53.4 (46.2, 60.5)     | 56.6 (46.7, 66.5)               | 55.7 (43.3, 68.2)          | .895 <sup>b</sup> |
| CCI                           |                 | 0 (0, 2)              | 1 (0, 2)                        | 1 (0, 2)                   | .599 <sup>b</sup> |
| SCS                           |                 | 8.2 (7.5, 8.8)        | 7.5 (6.4, 8.7)                  | 6.1 (5.0-8.5)              | .062 <sup>b</sup> |

ABC: Activities-specific Balance Confidence scale; CCI: Charleston Co-morbidity Index; IQR: Interquartile range; MFCL: Medicare Functional Classification Level; PLUS-M: Prosthetic Limb Users Survey of Mobility; SCS: Socket Comfort Score

<sup>a</sup> Chi-square tests run to compare categorical variables between prospective fall categories

<sup>b</sup> Kruskal-Wallis tests run to compare continuous variables across prospective fall categories

Table 3. Performance-based tests scores among transtibial prosthesis users stratified by prospective fall category

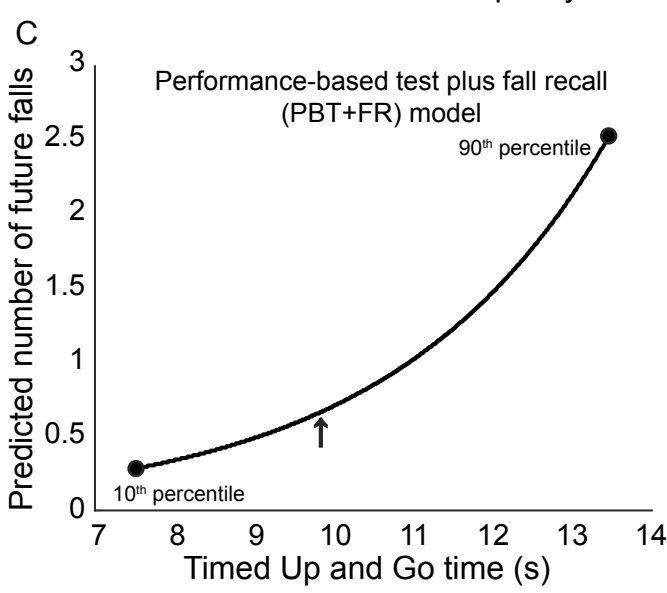
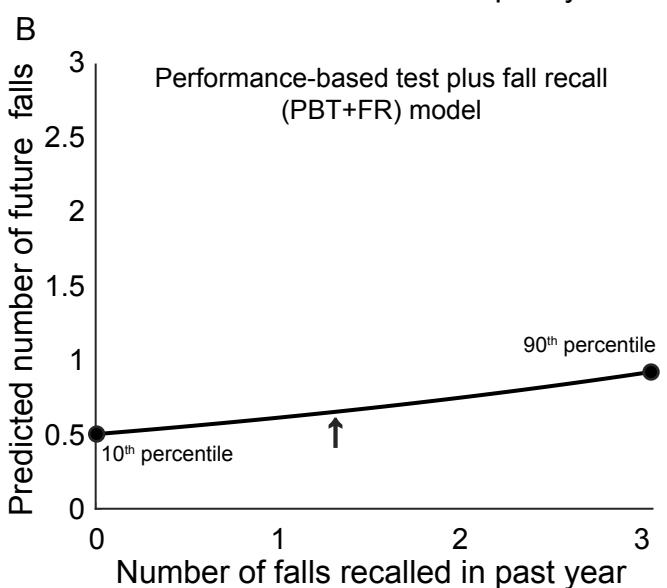
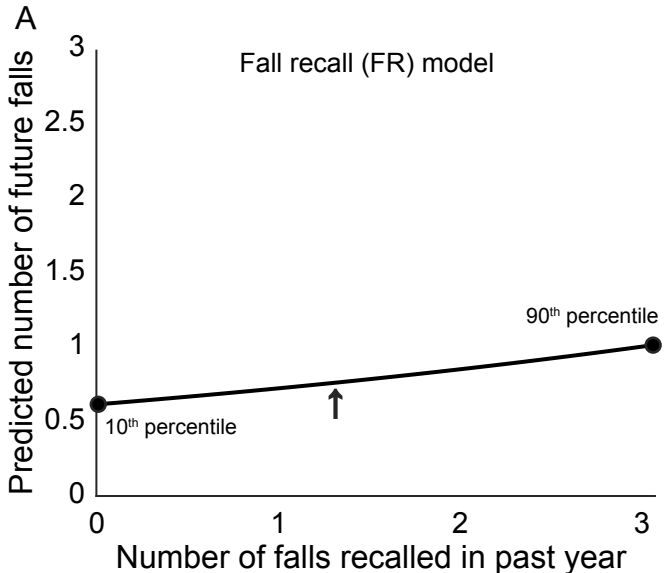
|                        | All Participants<br>n=45 |                                                   | Fall status based on 6-month prospective period |                                                |                                           | p-value |
|------------------------|--------------------------|---------------------------------------------------|-------------------------------------------------|------------------------------------------------|-------------------------------------------|---------|
|                        | Mean<br>(95% CI)         | 10 <sup>th</sup> , 90 <sup>th</sup><br>Percentile | Non-fallers<br>n=25<br>Mean (95% CI)            | Non-recurrent fallers<br>n=13<br>Mean (95% CI) | Recurrent fallers<br>n=7<br>Mean (95% CI) |         |
| NBWT score, /1.0       | .47<br>(.40, .53)        | .22, .77                                          | .49 (.40, .58)                                  | .43 (.30, .56)                                 | .47 (.34, .60)                            | .663    |
| FSST time, s           | 9.1<br>(8.0, 10.3)       | 5.2, 14.0                                         | 9.1 (7.2, 11.0)                                 | 9.4 (8.0, 10.7)                                | 8.6 (5.7, 11.6)                           | .477    |
| TUG time, s            | 9.8<br>(9.2, 10.4)       | 7.5, 13.5                                         | 9.4 (8.5, 10.3)                                 | 10.3 (9.2, 11.3)                               | 10.4 (8.3, 12.4)                          | .148    |
| 10MWT time, s          | 8.5<br>(8.0, 9.0)        | 6.5, 10.9                                         | 8.3 (7.6, 9.0)                                  | 8.8 (7.7, 9.9)                                 | 8.5 (7.3, 9.8)                            | .564    |
| # falls past 12 months | 1.27<br>(.56, 1.97)      | 0, 3                                              | .96 (.48, 1.44)                                 | .62 (.22, 1.01)                                | 3.57 (1.21, 8.34)                         | .590    |

CI: Confidence Interval; FSST = Four Square Step Test; IQR = Interquartile Range; MFCL = Medicare Functional Classification Level; NBWT = Narrowing Beam Walking Test; PLUS-M = Prosthetic Limb Users Survey of Mobility; s = seconds; TUG = Timed Up and Go.

Table 4. Negative binomial regression models to predict falls in unilateral transtibial prosthesis users

| <b>Performance-based test plus fall recall (PBT+FR) final model (<math>\beta_0 = -2.90</math>)</b>       |         |                   |         |
|----------------------------------------------------------------------------------------------------------|---------|-------------------|---------|
| Variable                                                                                                 | $\beta$ | IRR (95% CI)      | p-value |
| Number falls recalled                                                                                    | 0.202   | 1.22 (1.05, 1.42) | .009    |
| TUG                                                                                                      | 0.370   | 1.45 (1.01, 2.07) | .044    |
| FSST                                                                                                     | -0.154  | 0.86 (0.69, 1.06) | .149    |
| <i>Predicted #falls = exp ((0.202)*(number falls recalled) + (0.370)*(TUG) + (-0.154)*(FSST) - 2.90)</i> |         |                   |         |
| <b>Performance-based test (PBT) final model (<math>\beta_0 = 3.37</math>)</b>                            |         |                   |         |
| Variable                                                                                                 | $\beta$ | IRR (95% CI)      | p-value |
| PLUS-M T-score                                                                                           | -.064   | 0.94 (0.88, 1.00) | .058    |
| <i>Predicted #falls = exp ((-0.064)*(PLUS-M T-score) + 3.37)</i>                                         |         |                   |         |
| <b>Fall recall (FR) final model (<math>\beta_0 = -0.54</math>)</b>                                       |         |                   |         |
| Variable                                                                                                 | $\beta$ | IRR (95% CI)      | p-value |
| Number falls recalled                                                                                    | .175    | 1.19 (1.04, 1.37) | .014    |
| <i>Predicted #falls = exp ((0.175)*(number falls recalled) - 0.54)</i>                                   |         |                   |         |

$\beta$ : regression coefficient;  $\beta_0$ : regression intercept; FSST: Four Square Step Test; IRR: incidence rate ratio  
 PLUS-M: Prosthetic Limb Users Survey – Mobility; TUG: Timed Up and Go.



Supplementary Table 1. Performance-based test (PBT) model development (reduction)

| <b>Initial PBT Model: FSST, PLUS-M, TUG, NBWT</b>    |                                     |                                        |                            |                |        |
|------------------------------------------------------|-------------------------------------|----------------------------------------|----------------------------|----------------|--------|
| Variable                                             | Wald $X^2$ test of<br>model effects | Likelihood ratio $X^2$<br>omnibus test | Goodness-of-fit            |                |        |
|                                                      |                                     |                                        | $X^2$ goodness-of-fit      | Log Likelihood | AIC    |
| FSST                                                 | $X^2(1)=3.36$<br>$p=.067$           |                                        |                            |                |        |
| PLUS-M                                               | $X^2(1)=2.95$<br>$p=.086$           | $X^2(4)=7.59$<br>$p=.108$              | $X^2(40)=46.03$<br>$p=.20$ | -53.44         | 116.88 |
| TUG                                                  | $X^2(1)=1.46$<br>$p=.227$           |                                        |                            |                |        |
| NBWT                                                 | $X^2(1)=.503$<br>$p=.478$           |                                        |                            |                |        |
| <b>2nd iteration of PBT Model: FSST, PLUS-M, TUG</b> |                                     |                                        |                            |                |        |
| Variable                                             | Wald $X^2$ test of<br>model effects | Likelihood ratio $X^2$<br>omnibus test | Goodness-of-fit            |                |        |
|                                                      |                                     |                                        | $X^2$ goodness-of-fit      | Log Likelihood | AIC    |
| FSST                                                 | $X^2(1)=2.95$<br>$p=.086$           |                                        |                            |                |        |
| PLUS-M                                               | $X^2(1)=3.80$<br>$p=.051$           | $X^2(3)=7.08$<br>$p=.069$              | $X^2(41)=48.02$<br>$p=.20$ | -53.70         | 115.39 |
| TUG                                                  | $X^2(1)=1.64$<br>$p=.200$           |                                        |                            |                |        |
| <b>3rd iteration of PBT Model: FSST, PLUS-M</b>      |                                     |                                        |                            |                |        |
| Variable                                             | Wald $X^2$ test of<br>model effects | Likelihood ratio $X^2$<br>omnibus test | Goodness-of-fit            |                |        |
|                                                      |                                     |                                        | $X^2$ goodness-of-fit      | Log Likelihood | AIC    |
| PLUS-M                                               | $X^2(1)=4.75$<br>$p=.029$           | $X^2(2)=5.39$<br>$p=.068$              | $X^2(42)=43.49$<br>$p=.20$ | -54.54         | 115.08 |
| FSST                                                 | $X^2(1)=1.49$<br>$p=.223$           |                                        |                            |                |        |
| <b>Final PBT Model: PLUS-M</b>                       |                                     |                                        |                            |                |        |
| Variable                                             | Wald $X^2$ test of<br>model effects | Likelihood ratio $X^2$<br>omnibus test | Goodness-of-fit            |                |        |
|                                                      |                                     |                                        | $X^2$ goodness-of-fit      | Log Likelihood | AIC    |
| PLUS-M                                               | $X^2(1)=3.60$<br>$p=.058$           | $X^2(1)=3.83$<br>$p=.050$              | $X^2(43)=52.18$<br>$p=.10$ | -55.32         | 114.64 |

Supplementary Table 2. Performance-based test plus fall recall (PBT+FR) model development (reduction)

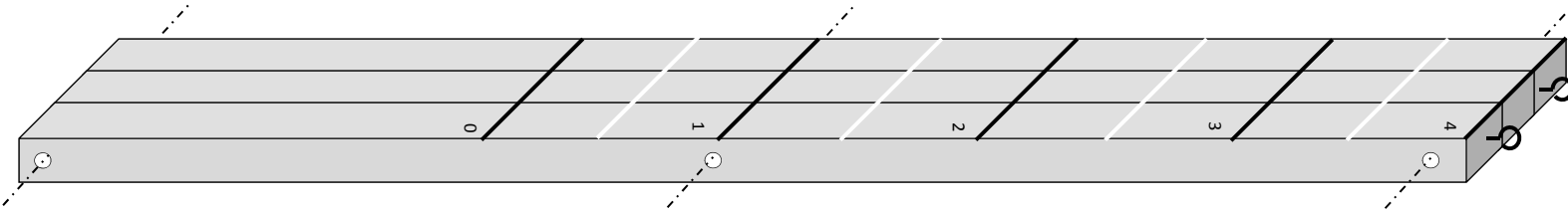
| <b>Initial PBT+FR Model: FSST, PLUS-M, TUG, NBWT, Fall recall</b>    |                                     |                                        |                          |                |        |
|----------------------------------------------------------------------|-------------------------------------|----------------------------------------|--------------------------|----------------|--------|
| Variable                                                             | Wald $X^2$ test of<br>model effects | Likelihood ratio $X^2$<br>omnibus test | Goodness-of-fit          |                |        |
|                                                                      |                                     |                                        | $X^2$ goodness-of-fit    | Log Likelihood | AIC    |
| FSST                                                                 | $X^2(1)=1.51$<br>p=.219             |                                        |                          |                |        |
| TUG                                                                  | $X^2(1)=3.26$<br>p=.071             |                                        |                          |                |        |
| Fall recall                                                          | $X^2(1)=4.16$<br>p=.041             | $X^2(5)=12.18$<br>p=.032               | $X^2(39)=33.40$<br>p=.20 | -51.15         | 114.29 |
| PLUS-M                                                               | $X^2(1)=.452$<br>p=.502             |                                        |                          |                |        |
| NBWT                                                                 | $X^2(1)=.136$<br>p=.713             |                                        |                          |                |        |
| <b>2nd iteration of PBT+FR Model: FSST, PLUS-M, TUG, Fall recall</b> |                                     |                                        |                          |                |        |
| Variable                                                             | Wald $X^2$ test of<br>model effects | Likelihood ratio $X^2$<br>omnibus test | Goodness-of-fit          |                |        |
|                                                                      |                                     |                                        | $X^2$ goodness-of-fit    | Log Likelihood | AIC    |
| FSST                                                                 | $X^2(1)=2.19$<br>p=.139             |                                        |                          |                |        |
| TUG                                                                  | $X^2(1)=3.16$<br>p=.076             | $X^2(4)=12.05$<br>p=.017               | $X^2(40)=34.64$<br>p=.20 | -51.21         | 112.43 |
| Fall recall                                                          | $X^2(1)=4.32$<br>p=.038             |                                        |                          |                |        |
| PLUS-M                                                               | $X^2(1)=.442$<br>p=.506             |                                        |                          |                |        |
| <b>Final PBT+FR Model: FSST, TUG, Fall recall</b>                    |                                     |                                        |                          |                |        |
| Variable                                                             | Wald $X^2$ test of<br>model effects | Likelihood ratio $X^2$<br>omnibus test | Goodness-of-fit          |                |        |
|                                                                      |                                     |                                        | $X^2$ goodness-of-fit    | Log Likelihood | AIC    |
| FSST                                                                 | $X^2(1)=1.98$<br>p=.150             |                                        |                          |                |        |
| TUG                                                                  | $X^2(1)=4.07$<br>p=.044             | $X^2(3)=5.39$<br>p=.009                | $X^2(41)=36.12$<br>p=.20 | -51.43         | 110.87 |
| Fall recall                                                          | $X^2(1)=6.92$<br>p=.009             |                                        |                          |                |        |

Supplementary Table 3. Fall recall (FR) model development (reduction)

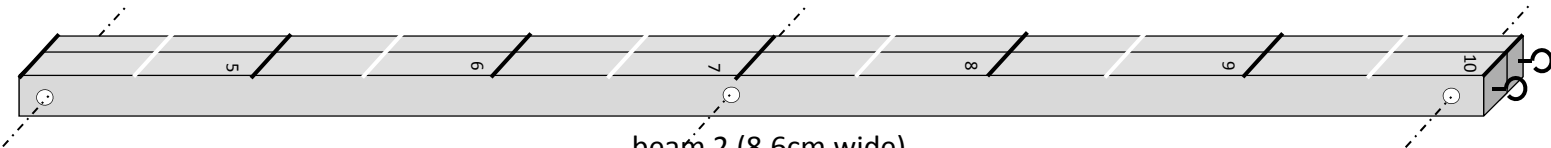
| Initial FR Model: Fall recall, PLUS-M |                                  |                                     |                          |                |        |
|---------------------------------------|----------------------------------|-------------------------------------|--------------------------|----------------|--------|
| Variable                              | Wald $X^2$ test of model effects | Likelihood ratio $X^2$ omnibus test | Goodness-of-fit          |                |        |
|                                       |                                  |                                     | $X^2$ goodness-of-fit    | Log Likelihood | AIC    |
| Fall recall                           | $X^2(1)=4.06$<br>p=.044          | $X^2(2)=8.50$<br>p=.014             | $X^2(42)=39.52$<br>p=.20 | -52.99         | 111.97 |
| PLUS-M                                | $X^2(1)=1.21$<br>p=.272          |                                     |                          |                |        |
|                                       |                                  |                                     |                          |                |        |
| Final FR Model: Fall recall           |                                  |                                     |                          |                |        |
| Variable                              | Wald $X^2$ test of model effects | Likelihood ratio $X^2$ omnibus test | Goodness-of-fit          |                |        |
|                                       |                                  |                                     | $X^2$ goodness-of-fit    | Log Likelihood | AIC    |
| Fall recall                           | $X^2(1)=6.02$<br>p=.014          | $X^2(1)=7.26$<br>p=.007             | $X^2(43)=45.54$<br>p=.20 | -53.61         | 111.21 |

## Appendix 3: NBWT Fabrication Guide

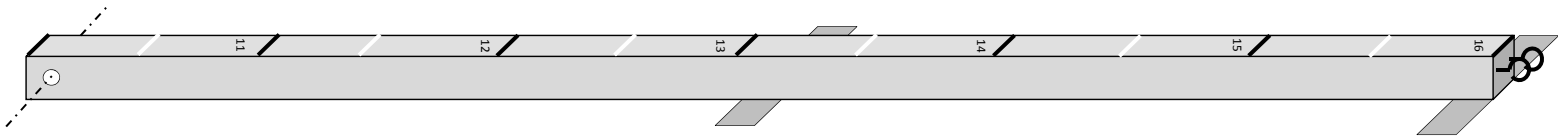
Narrowing Beam Walking Test (NBWT)



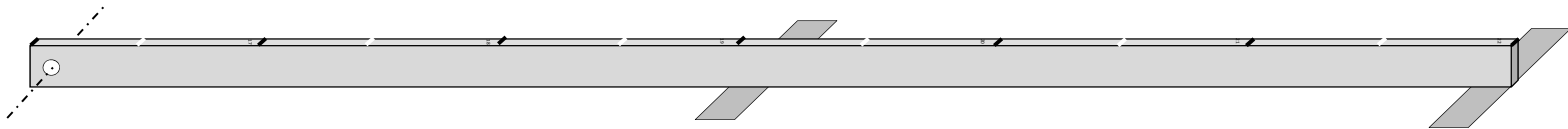
beam 1 (18.6cm wide)



beam 2 (8.6cm wide)



beam 3 (4.0cm wide)



beam 4 (2.0cm wide)

### **NBWT Parts List**

Lowes:

| Qty | Link                                                                                                                                                                                                                                                                                                                                                                                  |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2   | <a href="https://www.lowes.com/pd/ChoiceDek-Common-2-in-X-4-in-x-6-ft-Actual-1-47-in-x-3-41-in-x-6-ft-3-Pack-Beach-House-Gray-Composite-Deck-Handrail/50062619">https://www.lowes.com/pd/ChoiceDek-Common-2-in-X-4-in-x-6-ft-Actual-1-47-in-x-3-41-in-x-6-ft-3-Pack-Beach-House-Gray-Composite-Deck-Handrail/50062619</a><br>(three for beam 1, two for beam 2, two for beam 3 and 4) |
| 6   | <a href="https://www.lowes.com/pd/National-Hardware-Steel-Screw-Eye-Hook/1000246181">https://www.lowes.com/pd/National-Hardware-Steel-Screw-Eye-Hook/1000246181</a><br>(two for beam 1-2 connection, two for beam 2-3 connection, two for beam 3-4 connection)                                                                                                                        |

Home Depot:

| Qty | Link                                                                                                                                                                                                                                                                                                                                                           |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 4   | <a href="https://www.homedepot.com/p/Everbilt-5-in-x-5-in-Zinc-Plated-T-Plate-15068/202034097">https://www.homedepot.com/p/Everbilt-5-in-x-5-in-Zinc-Plated-T-Plate-15068/202034097</a><br>(two for Beam 1, two for Beam 2)                                                                                                                                    |
| 2   | <a href="https://www.homedepot.com/p/Everbilt-8-x-1-1-4-in-Phillips-Flat-Head-Wood-Screws-10-per-Pack-21041/202705535">https://www.homedepot.com/p/Everbilt-8-x-1-1-4-in-Phillips-Flat-Head-Wood-Screws-10-per-Pack-21041/202705535</a> (two packages for Beams 1 and 2)                                                                                       |
| 16  | <a href="https://www.homedepot.com/p/Everbilt-3-8-in-Zinc-Plated-Cut-Washer-807236/204633115">https://www.homedepot.com/p/Everbilt-3-8-in-Zinc-Plated-Cut-Washer-807236/204633115</a><br>(six for beam 1, six for beam 2 [inc. two for beam 1-2 connection], two for beam 3 [inc. two beam 2-3 connection], two for beam 4 [inc. two for beam 3-4 connection]) |
| 3   | <a href="https://www.homedepot.com/p/Everbilt-3-8-in-16-tpi-x-8-in-Zinc-Plated-Hex-Bolt-800936/204645580">https://www.homedepot.com/p/Everbilt-3-8-in-16-tpi-x-8-in-Zinc-Plated-Hex-Bolt-800936/204645580</a> (three for beam 1)                                                                                                                               |
| 1   | <a href="https://www.homedepot.com/p/Crown-Bolt-3-8-in-16-tpi-x-4-1-2-in-Zinc-Plated-Hex-Bolt-83156/204960793">https://www.homedepot.com/p/Crown-Bolt-3-8-in-16-tpi-x-4-1-2-in-Zinc-Plated-Hex-Bolt-83156/204960793</a> (one for beam 1-2 connection and first bolt in beam 2)                                                                                 |
| 2   | <a href="https://www.homedepot.com/p/Crown-Bolt-3-8-in-16-tpi-x-4-in-Zinc-Plated-Hex-Bolt-87266/204960792">https://www.homedepot.com/p/Crown-Bolt-3-8-in-16-tpi-x-4-in-Zinc-Plated-Hex-Bolt-87266/204960792</a> (two for beam 2)                                                                                                                               |
| 1   | <a href="https://www.homedepot.com/p/Crown-Bolt-3-8-in-16-tpi-x-3-in-Zinc-Plated-Hex-Bolt-83146/204960791">https://www.homedepot.com/p/Crown-Bolt-3-8-in-16-tpi-x-3-in-Zinc-Plated-Hex-Bolt-83146/204960791</a> (one for beam 2-3 connection)                                                                                                                  |
| 1   | <a href="https://www.homedepot.com/p/Crown-Bolt-3-8-in-16-tpi-x-1-1-2-in-Zinc-Plated-Hex-Bolt-87206/204960790">https://www.homedepot.com/p/Crown-Bolt-3-8-in-16-tpi-x-1-1-2-in-Zinc-Plated-Hex-Bolt-87206/204960790</a> (one for beam 3-4 connection)                                                                                                          |
| 1   | <a href="https://www.homedepot.com/p/Everbilt-3-8-in-x-16-in-Zinc-Plated-Wing-Nut-3-Piece-802391/204274204">https://www.homedepot.com/p/Everbilt-3-8-in-x-16-in-Zinc-Plated-Wing-Nut-3-Piece-802391/204274204</a> (one for beam 1-2 connection, one for beam 2-3 connection, one for beam 3-4 connection)                                                      |
| 5   | <a href="https://www.homedepot.com/p/Everbilt-3-8-in-Zinc-Grade-5-SAE-Hex-Nut-808686/204281818">https://www.homedepot.com/p/Everbilt-3-8-in-Zinc-Grade-5-SAE-Hex-Nut-808686/204281818</a> (three for beam 1, two for beam 2)                                                                                                                                   |
| 2   | <a href="https://www.homedepot.com/p/Roberts-2-1-2-in-x-25-ft-Roll-of-Rug-Traction-Anti-Slip-Rubber-Tape-50-545/100658264?keyword=075378505457">https://www.homedepot.com/p/Roberts-2-1-2-in-x-25-ft-Roll-of-Rug-Traction-Anti-Slip-Rubber-Tape-50-545/100658264?keyword=075378505457</a> (two rolls enough for all beams)                                     |

### **Fabrication Supplies**

Home Depot:

| Qty | Link                                                                                                                                                                                                                                                                                                         |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1   | <a href="https://www.homedepot.com/p/Diablo-10-in-x-72-Tooth-Trex-Composite-Material-Cutting-Saw-Blade-D1072CD/203754047?keyword=trex+blade">https://www.homedepot.com/p/Diablo-10-in-x-72-Tooth-Trex-Composite-Material-Cutting-Saw-Blade-D1072CD/203754047?keyword=trex+blade</a> (to cut composite beams) |
| 1   | <a href="https://www.homedepot.com/p/DEWALT-3-8-in-x-12-in-Black-Oxide-Drill-Bit-DW1610-G/202579647?keyword=028874116108">https://www.homedepot.com/p/DEWALT-3-8-in-x-12-in-Black-Oxide-Drill-Bit-DW1610-G/202579647?keyword=028874116108</a> (to drill thru composite beams)                                |
| 1   | <a href="https://www.homedepot.com/p/3M-17-6-oz-High-Strength-90-Spray-Adhesive-90-24/100151277">https://www.homedepot.com/p/3M-17-6-oz-High-Strength-90-Spray-Adhesive-90-24/100151277</a> (to secure rubber tape to beam)                                                                                  |

### **Fabrication Equipment**

Table saw, a miter saw (or circular saw), router, 3/8 drill, crescent wrench

### **Safety Equipment**

Respirators, safety goggles, and gloves

### **Notes**

To paint the lines and numbers on the beam, use either permanent paint markers or paint and a fine paint brush (I ended up using a combo of both and some spare black and white paint)

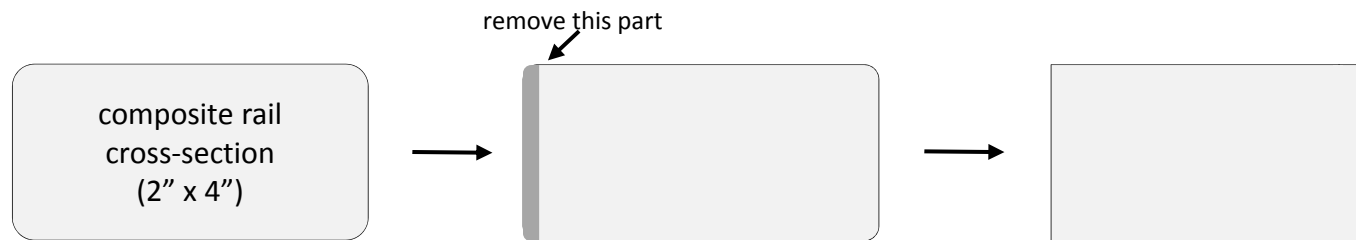
- <https://www.homedepot.com/p/Milwaukee-Inkzall-Fine-Point-Black-Marker-2-Pack-48-22-3105/205430341> (to mark 1' lines and numbers on all beams)
- <https://www.homedepot.com/p/Milwaukee-INKZALL-Fine-Point-Markers-Silver-2-Pack-48-22-3121/206973450> (to mark 6" lines on all beams)
- <https://www.homedepot.com/p/Angular-Artist-Paint-Brush-Set-4-Piece-AM-4044/205671666> (to paint the lines in the cross-cuts, if you have some paint you want to use)

I tried a stencil for the numbers, and I'm not sure it was ultimately much help. If you have a steady hand, you could probably free-hand it as easily.

- <https://www.amazon.com/CH-Hanson-10006-Single-Number/dp/B00065TU6E> (got this same one off ebay for less)

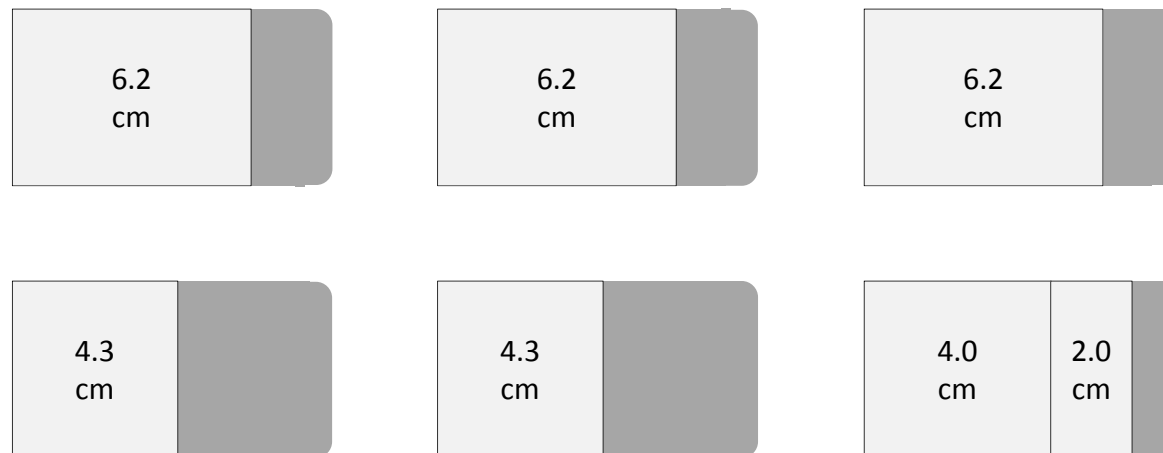
## Step 1: Rip the composite railings with a table saw and composite blade

- 1) Rip the curved edges off one edge of the composite railings to create a straight edge



the second edge will be made with the cuts noted below

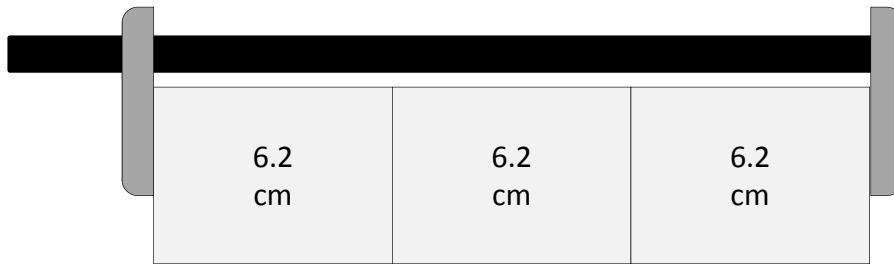
- 2) Rip three 6.2cm wide sections (beam 1), two 4.3cm wide sections (beam 2), one 4.0cm wide section (beam 3), and one 2.0cm wide section (beam 4)



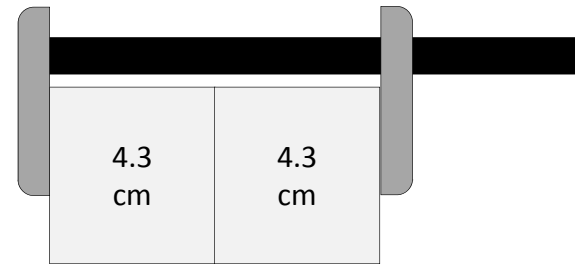
shown above are the six cross-sections of the composite rails and suggested cuts

## Step 2: Drill thru holes in beams with a 3/8" drill bit

- 1) Clamp together the 6.2cm segments and clamp together the 4.3cm segments

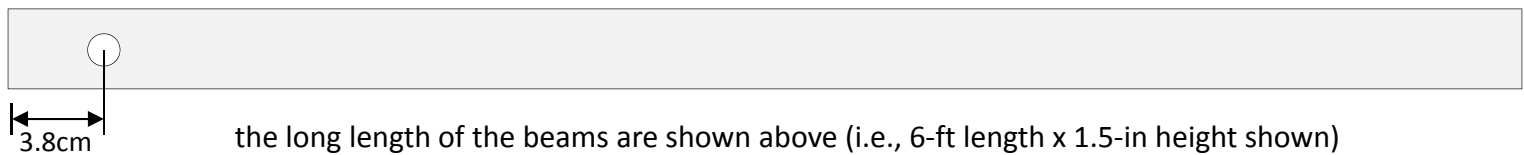
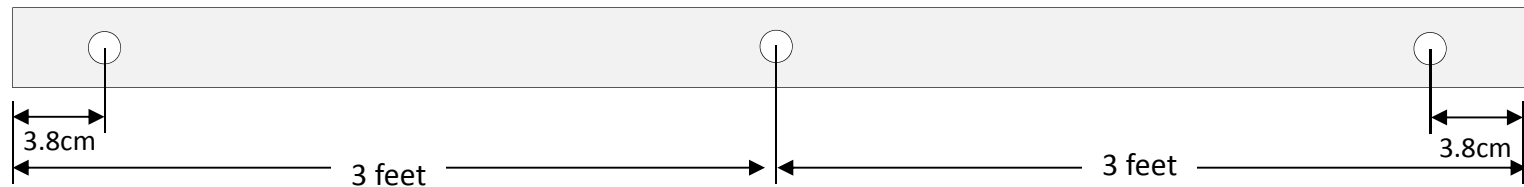


this will be beam 1 (18.6cm wide)

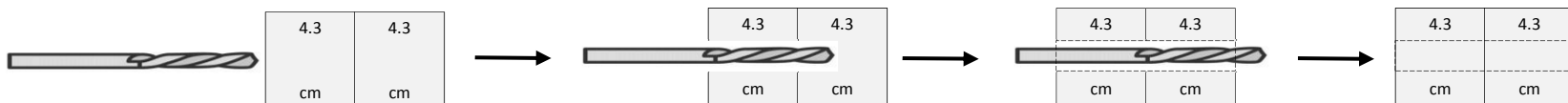


this will be beam 2 (8.6cm wide)

- 2) Drill a 3/8" thru-hole thru the beam segments where indicated

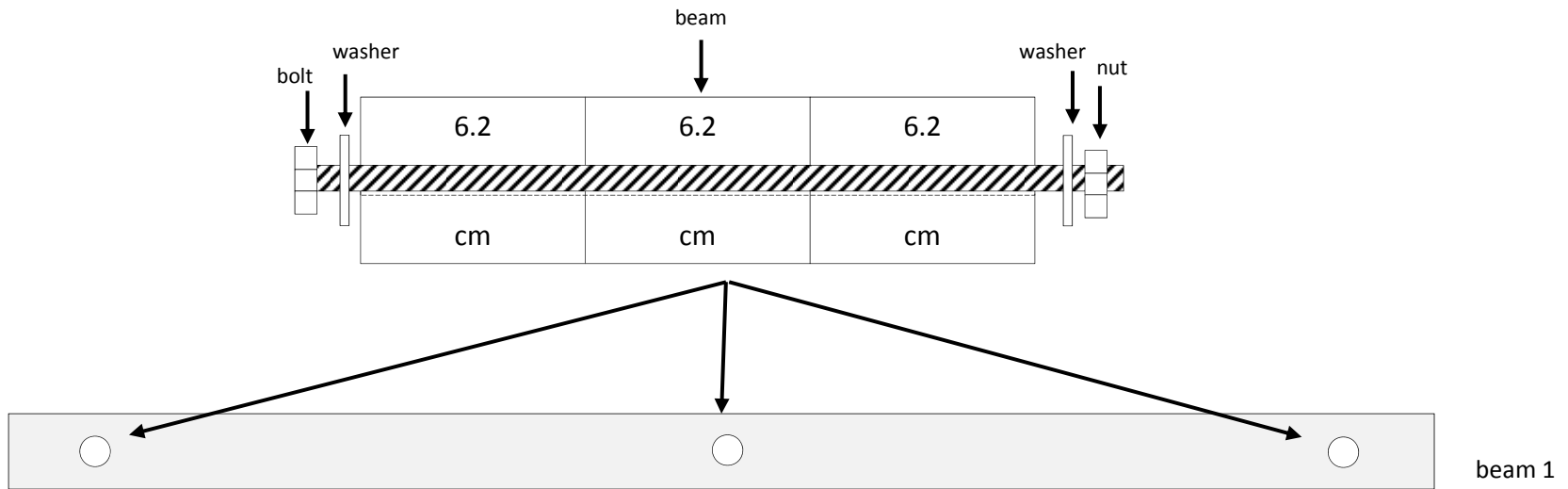


the long length of the beams are shown above (i.e., 6-ft length x 1.5-in height shown)



### Step 3: Assemble beam 1

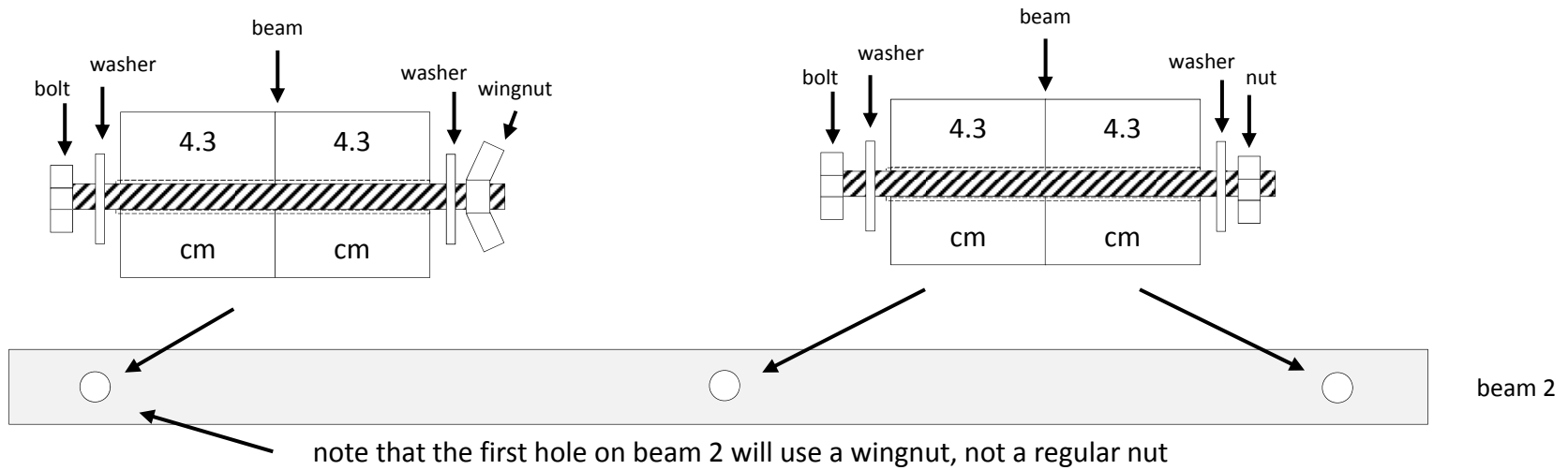
- 1) Loosely bolt together beam 1 (the 6.2cm segments) using 3/8" hardware



- 2) Lay beam 1 on the floor so that it aligns, and then tighten the hardware

## Step 4: Assemble beam 2

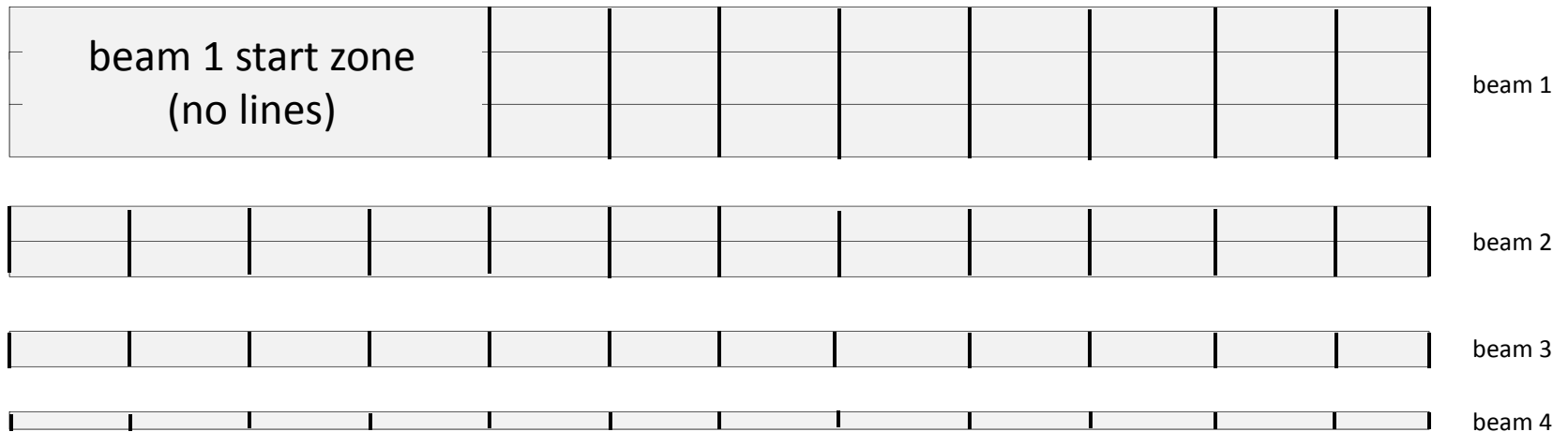
- 1) Loosely bolt together beam 2 (the 4.3cm segments) using 3/8" hardware



- 2) Lay beam 2 on the floor so that it aligns, and then tighten the hardware

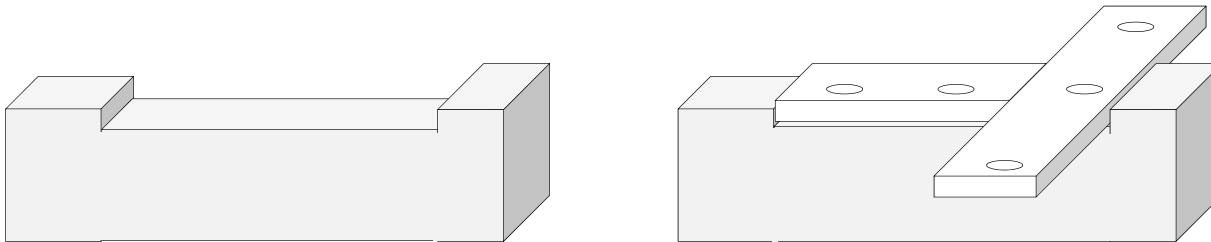
**Step 5:** Cut measurement lines on all beams using a miter or circular saw

- 1) Cut  $\frac{1}{4}$ " deep measurement lines every 6" across the walking surface of each of the beams as indicated. Note that the first 2 feet of beam 1 will not have lines.

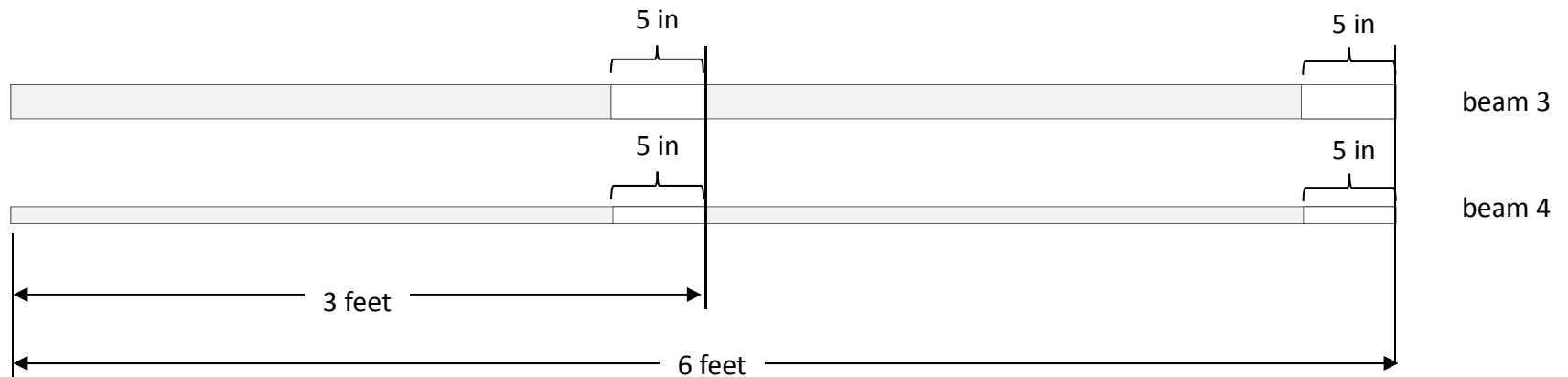


**Step 6:** Route out material on the bottom of the beam for the stabilizers

- 1) Route out a channel in the bottom of beams 3 and 4 so that the T-brackets will sit flush with the floor when the rubber tape is attached to the beam (the T-brackets should sit just above the beam surface).



- 2) Route one channel 5" distal to the center and far end of beams 3 and 4 (four total)



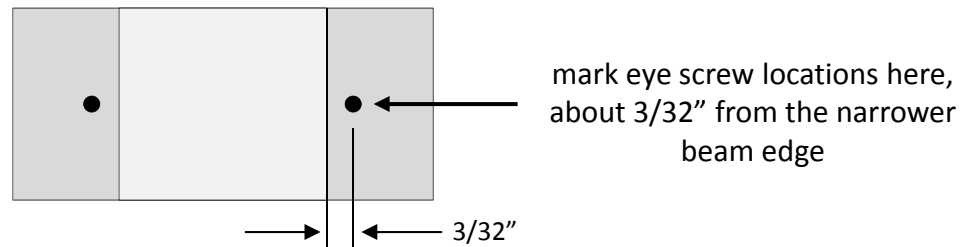
- 3) Attach T-brackets to the bottom of the beam with 1-1/4" wood screws (be sure to have countersunk side of the T-bracket facing the bottom of the beam).

## Step 7: Attach eye hooks to connect the beams end-to-end

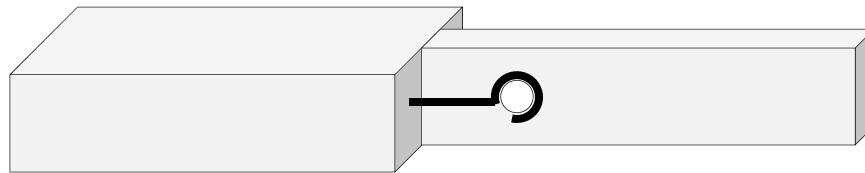
- 1) Lay two adjacent segments of the NBWT end-to-end.



- 2) Identify location to drill for the eye screws about  $\frac{3}{32}$ " to either side of the edges of the narrower beam segment (along the centerline).

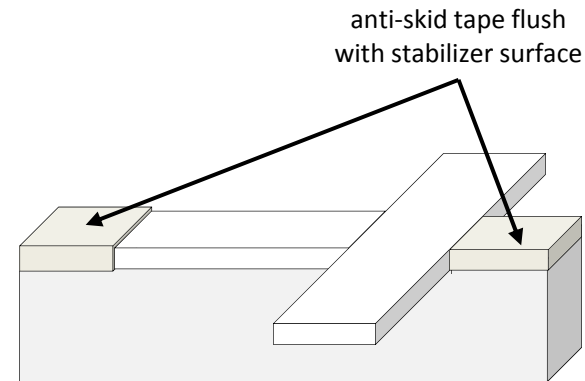
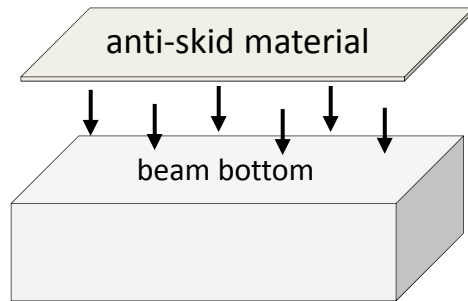


- 3) Drill pilot holes ( $\sim \frac{3}{32}$ ") for the  $\frac{3}{16}$ " eye screws. Insert eye screws so they line up with the holes in the adjacent beam.



## Step 8: Attach anti-skid material to bottom of the beam

- 1) Attach the anti-skid rubber tape to the bottom of each beam segment. Do not cover the stabilizers, as they will be flush to the ground. Use spray adhesive between the beam and anti-grip rubber tape for better adhesion.

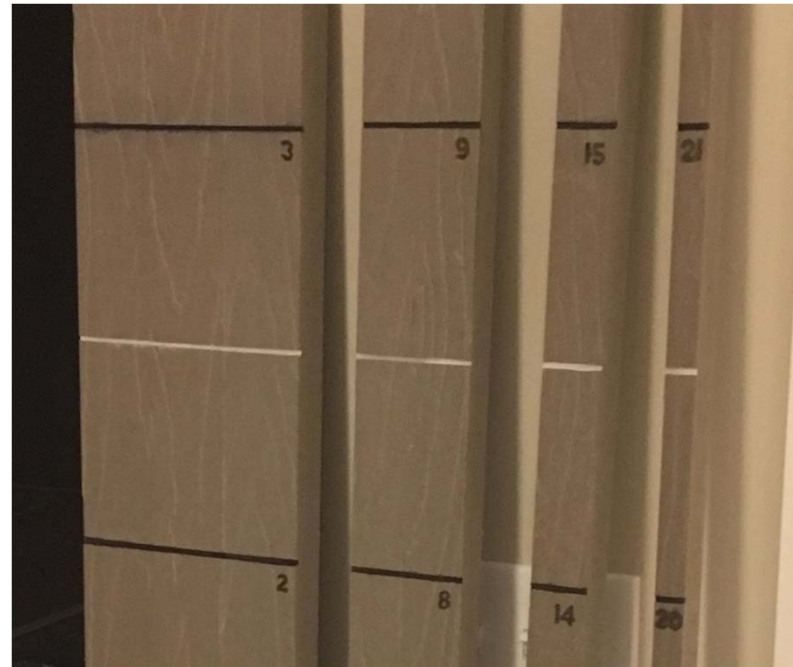


## Step 9: Paint lines and numbers to designate measurements

- 1) Paint the foot lines (0, 1, 2....20, 21, 22) with black paint; paint the half-foot (6") lines (0.5, 1.5, 2.5....19.5, 20.5, 21.5) with white paint
- 2) Paint numbers to designate the 1-foot lines (0, 1, 2...20, 21, 22). Make numbers as large as can fit on beam 4. Do not paint lines for the 6" marks.



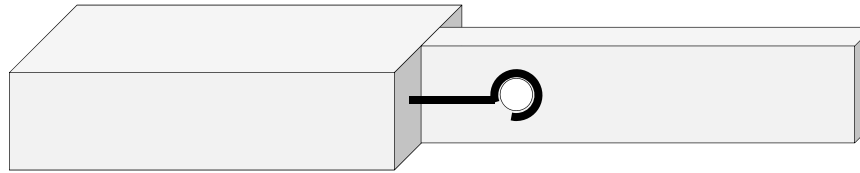
measurement lines for foot (black) and half-foot (white) distances



measurement numbers for each foot

**Step 10:** Insert hardware in beams 3 and 4, and assemble and administer NBWT

- 1) Lay beam segments on the floor, placing narrower beams between the eye screws.



- 2) Insert bolts through both eye screws and beam, secure with washer and wingnut



- 3) Secure the stabilizers (T-brackets) to the ground with high-quality duct tape.
- 4) Store flat on the ground (rubber tape up) to avoid warping.

## Finished NBWT



Note that the above version only has stabilizers on beam 1 – the instructions add a recommended set of stabilizers on beam 2