Abstract

Few established measures allow effective quantification of physical performance in severely injured service members. We sought to establish preliminary normative data in 180 healthy, active-duty service members for physical performance measures that can be readily implemented in a clinical setting. Interrater and test-retest reliability and minimal detectable change (MDC) values were also determined. Physical performance testing included self-selected walking velocity on level and uneven terrain, timed stair ascent, the sit-to-stand five times test, the four-square step test, and the 6-minute walk test. Data analysis included descriptive statistics, intraclass correlation coefficients, and MDC. Interrater and test-retest reliability were excellent for all measures (intraclass correlation coefficients >0.75). MDC values for timed measures were <0.3 seconds for interrater comparisons and <1.5 seconds for between-day comparisons. Physical performance measures had a narrow range of normal performance and were reliable and stable between days.

The prevalence of severe extremity trauma sustained during the conflicts in Iraq and Afghanistan as well as the potential long-term costs associated with these injuries have been well documented. Understanding of the time course of recovery associated with these combat injuries is limited, and limited evidence specifically quantifies the effect of surgical and rehabilitative interventions on physical function in patients with these injuries. Efforts are underway to quantify the efficacy of existing interventions to improve patient care. However, clinicians are confronted by the current paucity of established measures that allow for effective characterization of physical performance in both healthy and injured young adults.

Outcomes assessment following extremity trauma is typically performed using qualitative assessments and questionnaires that rely on patient self-report. Although these measures are useful for characterizing vital global outcomes such as quality of life, they typically lack sufficient resolution to fully describe disability, track gradual improvements over time, and link interventions to a change in a specific aspect of physical performance.

Quantitative assessment of physical performance in patients with severe extremity trauma typically includes basic functional activities such as rising from a chair and walking a short distance. Although appropriate for older adults (age >65 years) or more impaired patients,
# Physical Performance Assessment in Military Service Members

Few established measures allow effective quantification of physical performance in severely injured service members. We sought to establish preliminary normative data in 180 healthy, active-duty service members for physical performance measures that can be readily implemented in a clinical setting. Interrater and test-retest reliability and minimal detectable change (MDC) values were also determined. Physical performance testing included self-selected walking velocity on level and uneven terrain, timed stair ascent, the sit-to-stand five times test, the four-square step test, and the 6-minute walk test. Data analysis included descriptive statistics intraclass correlation coefficients, and MDC. Interrater and test-retest reliability were excellent for all measures (intraclass correlation coefficients >0.75). MDC values for timed measures were <0.3 seconds for interrater comparisons and <1.5 seconds for between-day comparisons. Physical performance measures had a narrow range of normal performance and were reliable and stable between days.
these measures typically demonstrate ceiling effects in more highly functioning persons (eg, athletes, military service members). In contrast, routine tests used to assess athletic populations often require significant areas of open space (eg, obstacle course, agility run) and can include activities that may place more severely involved patients at risk of further injury.\(^{15}\)

To address the need for reliable and relevant physical performance assessment measures, we sought to establish preliminary normative data for young, healthy service members and determine the reliability for measures that are readily implemented in a clinical setting. Selected measures were thought to be likely to reflect anticipated improvements resulting from commonly used interventions; be usable across a broad continuum of injury severity; demonstrate excellent reliability and sufficient resolution to detect relevant changes in performance; be implementable with minimal cost and training; and enable effective characterization of physical performance across a range of functional domains, including agility, mobility, balance, power, and exercise capacity.

## Methods

This descriptive and repeated-measures study included 180 healthy, active-duty service members training at Fort Sam Houston, Texas. Participants ranged in age from 18 to 43 years and had no current or recent history (within 6 months) of medical or neuromusculoskeletal disorders that limited participation in their military occupation specialty or physical training activities. Participants were excluded if they did not have full, pain-free motion of the spine and lower extremities or could not complete heel-toe walking, five deep squats, and five single-leg hops on each limb. The study was approved by the Brooke Army Medical Center Institutional Review Board, and all participants provided written consent.

Participants were randomly assigned to one of four physical performance testing stations, including self-selected walking velocity (SSWV) on level ground and on a loose-rock surface, timed stair ascent (TSA), the sit-to-stand five times (STS5) test, and the four-square step test (FSST). Testing order was counterbalanced and, after completing the exercise at each of the four physical performance stations, participants completed the 6-minute walk test (6MWT). Instruction, demonstration, and data collection for all measures were done by four physical therapists and five physical therapy students. To ensure uniformity of testing procedures, testers received 2 hours of training in the administration of the measures by two of the authors (S.W.S. and S.L.G.). To assess interrater reliability, two raters recorded performance for 25 participants as they completed all tests. To assess test-retest reliability, 20 participants returned for repeat testing 5 to 10 days after their initial assessment.

## Study Measures

Measurements included in the study reflect activities required for basic mobility and participation in civilian and military environments. For example, the SSWV test allows direct assessment of mobility, fall risk, and disability.\(^{16}\) SSWV on a loose-rock surface (SSWV\(_{RS}\)) was examinedsecondary to its importance as a critical task for optimal performance in military and athletic environments. In our study, participants were instructed to walk a distance of 20 meters at a normal, comfortable pace. Gait speed was calculated based on the time required to cross the middle 10 meters of the walking path. The rocky path consisted of stones measuring approximately 1 to 2 inches in diameter and ≥3 inches deep. The averages of three walking tests over ground (SSWV\(_{OOG}\)) and three SSWV\(_{RS}\) trials were used in the final analysis.

TSA is often used as an objective measure of mobility and power in older adults (age >65 years).\(^{17}\) However, limited research exists regarding normal TSA in younger adults. We define TSA as the ability to ascend 12 steps as quickly as possible, without using the hand railing, touching every step with at least one foot (ie, alternating steps). Timing began on the word “go” and stopped when both of the participant’s feet were on top of the 12th step. Participants performed three test trials and were allowed 1 minute of rest between trials.

The STS5 test is commonly used to assess lower extremity strength and mobility in older adults (age >65 years),\(^{18,19}\) but critical analysis of this measure in younger adults is lacking. Proper completion of the STS5 test requires participants to fully stand up and sit down five times as fast as possible with the arms crossed over the chest. Participants completed two trials and were allowed 1 minute of rest between trials.

Mobility and dynamic balance

---

None of the following authors or any immediate family member has received anything of value from or has stock or stock options held in a commercial company or institution related directly or indirectly to the subject of this article: Dr. Wilken, Dr. Darter, Mr. Goffar, Mr. Ellwein, Ms. Snell, Mr. Tomalis, and Dr. Shaffer.
were also assessed with the FSST. This measure has excellent interrater and test-retest reliability (intraclass correlation coefficient [ICC] = 0.99 and 0.98, respectively) and is a valid measure for identifying fall risk in older, community-dwelling adults, older adults who have sustained a transtibial amputation,20 and adults with vestibular dysfunction.21 In the FSST, the patient steps sequentially over four 1-in–diameter sticks or canes that are placed flat on the floor in the shape of a cross. In our study, participants began in the left rear square and were required to step over each cane as they moved as fast as possible in the following pattern: (1) forward, (2) sidestep to the right, (3) backward, and (4) sidestep to the left. They returned by sidestepping to the right, forward, left, and backward. Timing began when the participant’s foot was placed in the box in front of him or her and was stopped when both feet were placed in the final box. Participants completed one practice trial followed by four test trials.

The 6MWT is frequently used to assess aerobic fitness, endurance, and mobility by measuring the distance a person can walk on a level surface in 6 minutes.22,23 Normative times have been established across age spectrums,23-25 but test-retest reliability remains unclear.26,27 Participants were instructed to walk, not run, as far as they could for 6 minutes.22,23 Normative times for each measure. With the exception of the 6MWT, for which only one trial was conducted per participant, mean values were used for data analysis. Interrater and test-retest reliability was analyzed with ICC models 2,1 and 2,k (ie, two-way random, single measure and two-way random, average measure, respectively), and minimal detectable change (MDC) was determined at the 95% confidence level (MDC95). The standard error of the mean (SEM) represents the standard deviation (SD) of the measurement error and was calculated using the operation

\[
SD \times \sqrt{1 - ICC},
\]

with SD representing the pooled variance.28 The MDC95 is an extension of the

\[
\text{SEMMDC}_{95} = \text{SEM} \times z-score (95\% \text{ confidence interval}) \times \sqrt{2}
\]

and provides a boundary for the minimum amount of variation that is not due to chance.28 Assessment of ICC statistics was conducted using criteria described by Fleiss,29 with reliability coefficients ≥0.75 rated as excellent, 0.40 to 0.74 rated as fair to good, and <0.40 rated as poor.

**Results**

Participants’ demographic characteristics are reported in Table 1. Participants ranged in age from 18 to 43 years (mean ± SD = 24.5 ± 5.4), and 72.2% were men. Measures were reported for both men and women secondary to significant \(P < 0.05\) differences in height and weight, as well as in SSWVOG and TSA physical performance measures (Table 1). Participants had a high level of performance and limited variability in all measures except the 6MWT, in which the 5th percentile was 578.5 meters and the 95th percentile was 860.1 meters. Of 180 participants, 90% scored between 2.3 and 3.7 seconds on the TSA, 4.7 and 8.9 seconds on the STS5 test, and 3.9 and 7.8 seconds on the FSST. Participants also exhibited a limited range of performance on SSWVOG and SSWVRS, with 90% of scores ranging from 1.27 to 1.80 m/s and 0.95 to 1.56 m/s, respectively.

Interrater reliability was assessed in 25 of 180 participants and was found to be nearly perfect for all measures (ICC = 0.97 to 0.99, Table 2). In addition, MDC95 values reflected limited error between raters, with 0.05 m/s each for SSWVOG and SSWVRS, 0.19 seconds for the TSA, 0.27 seconds for the STS5 test, and 0.30 seconds for the FSST. Table 3 demonstrates differences between the mean and SD of the first and second sessions for the four performance measures as well as ICC and MDC95 values. All physical performance measures exhibited excellent test-retest reliability, with ICC ranging from 0.86 to 0.93. For each measure, the mean improved between sessions, with an MDC95 value of 81.25 meters for the 6MWT and MDC95 values of 0.37 seconds for the TSA, 1.12 seconds for the STS5 test, and 1.41 seconds for the FSST.

**Discussion**

We sought to establish preliminary normative data and determine the reliability of and MDC values for physical performance measures that can be readily implemented in a clinical setting. The young, healthy cohort that we examined reflected demographics similar to those of injured service members described in earlier reports.14 The selected measures were found to reliably allow for rapid assessment of a range of functional domains, including agility, mobility, balance, power, and exercise capacity.

In general, the mean and SD values...
reflect a consistent and high-level baseline physical ability in the tested cohort compared with assessments of physical ability in older or impaired cohorts described by other authors (FSST,21,30,31 STS5 test,18 6MWT,12,32,33 SSWV32,33). Our participants demonstrated limited variability and a narrow range of performance for all measures except the 6MWT.

Overall, our study provides initial evidence regarding expected range of performance in young, healthy adults for six physical performance measures: SSWVOG, SSWVRS, STS5, FSST, TSA, and 6MWT. These measures demonstrated excellent interrater reliability and reliability between days (ICC >0.75) that was equal to or greater than that reported in earlier studies.12,32,33 However, reliability of measures other than the 6MWT in young, active persons has not been published to date.

Interrater MDC values were <0.3 second for simple timed measures and <0.1 m/s for walking velocity measures. Between-day MDC values ranged from 0.37 to 1.41 seconds for simple timed measures. Standardized instructions with clearly defined start and stop points were used to minimize the MDC values for timed measures. Mean values for between-day comparisons demonstrated small but significant improvements between

### Table 1

**Baseline Characteristics of Participants**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean ± SD (range)</th>
<th>95% CI</th>
<th>Median (quartiles)</th>
<th>5th, 95th Percentiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yr)a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>24.5 ± 5.5 (18–43)</td>
<td>23.3–25.4</td>
<td>24 (20–27)</td>
<td>18, 36</td>
</tr>
<tr>
<td>Women</td>
<td>24.5 ± 5.1 (18–40)</td>
<td>22.9–26.0</td>
<td>24 (22–27)</td>
<td>18, 38</td>
</tr>
<tr>
<td>Height (cm)a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>175.9 ± 7.4 (152–193)</td>
<td>174.3–177.0</td>
<td>176 (171–182)</td>
<td>163, 188</td>
</tr>
<tr>
<td>Women</td>
<td>163.2 ± 6.3 (153–176)</td>
<td>161.3–165.2</td>
<td>162 (158–168)</td>
<td>153, 175.5</td>
</tr>
<tr>
<td>Weight (kg)a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>80.1 ± 12.4 (52.3–111.8)</td>
<td>78.0–82.6</td>
<td>79.3 (70.9–87.7)</td>
<td>60.9, 102.6</td>
</tr>
<tr>
<td>Women</td>
<td>64.7 ± 8.2 (50–88.2)</td>
<td>63.5–68.4</td>
<td>64.1 (59–70.6)</td>
<td>52.2, 82.4</td>
</tr>
<tr>
<td>SSWVOG (m/s)b</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>1.51 ± 0.17 (1.0–1.93)</td>
<td>1.48–1.54</td>
<td>1.52 (1.41–1.64)</td>
<td>1.27, 1.76</td>
</tr>
<tr>
<td>Women</td>
<td>1.48 ± 0.16 (1.1–1.8)</td>
<td>1.43–1.53</td>
<td>1.45 (1.37–1.60)</td>
<td>1.27, 1.80</td>
</tr>
<tr>
<td>SSWVRS (m/s)b</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>1.26 ± 0.18 (0.66–1.70)</td>
<td>1.23–1.29</td>
<td>1.25 (1.17–1.39)</td>
<td>0.98, 1.56</td>
</tr>
<tr>
<td>Women</td>
<td>1.19 ± 0.14 (0.86–1.54)</td>
<td>1.15–1.24</td>
<td>1.19 (1.11–1.24)</td>
<td>0.95, 1.47</td>
</tr>
<tr>
<td>Timed Stair Ascent (s)a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>2.82 ± 0.37 (2.27–4.5)</td>
<td>2.75–2.89</td>
<td>2.7 (2.6–3.0)</td>
<td>2.3, 3.5</td>
</tr>
<tr>
<td>Women</td>
<td>3.14 ± 0.36 (2.4–4.0)</td>
<td>3.03–3.25</td>
<td>3.2 (2.9–3.4)</td>
<td>2.5, 3.7</td>
</tr>
<tr>
<td>Sit-to-stand 5 Times Test (s)a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>6.0 ± 1.0 (4.4–11.0)</td>
<td>5.8–6.1</td>
<td>5.8 (5.3–6.4)</td>
<td>4.8, 8.9</td>
</tr>
<tr>
<td>Women</td>
<td>6.24 ± 1.1 (4.35–9.14)</td>
<td>5.9–6.6</td>
<td>6.1 (5.4–6.8)</td>
<td>4.7, 8.8</td>
</tr>
<tr>
<td>Four-square Step Test (s)a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>5.7 ± 1.0 (3.7–8.6)</td>
<td>5.5–5.9</td>
<td>5.6 (4.9–6.4)</td>
<td>3.9, 7.5</td>
</tr>
<tr>
<td>Women</td>
<td>6.0 ± 1.0 (4.0–8.1)</td>
<td>5.7–6.3</td>
<td>6.1 (5.4–6.7)</td>
<td>4.1, 7.8</td>
</tr>
<tr>
<td>Six-minute Walk Test (m)a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>724.9 ± 84.1 (509–1,007)</td>
<td>710.3–739.5</td>
<td>722.2 (670.3–778.5)</td>
<td>581.2, 860.1</td>
</tr>
<tr>
<td>Women</td>
<td>707.4 ± 69.8 (555–856)</td>
<td>687.5–727.2</td>
<td>707.8 (655.3–760.5)</td>
<td>578.5, 820.5</td>
</tr>
</tbody>
</table>

CI = confidence interval, SD = standard deviation, SSWVOG = self-selected walking velocity over ground, SSWVRS = self-selected walking velocity over loose-rock surface

a N = 180 (130 men, 50 women)
b N = 160 (117 men, 43 women)
sessions for all measures except TSA. Potentially more important is the relative magnitude of the MDC values compared with the types of changes expected in patients during the recovery process. Although it is necessary to determine population-specific MDC values for injured service members, these early results suggest that the measures tested likely provide sufficient resolution to detect meaningful change in performance.

Several factors should be taken into consideration when using our results to interpret physical performance data. Data were collected in a cohort of physically fit persons who were expected to perform similarly over time, thereby increasing the likelihood of obtaining high ICC and low MDC values.

The time between testing days was approximately 1 week; therefore, it remains unknown whether the same level of stability will be observed over much longer periods of time.

### Summary

To our knowledge, our study is the first to provide preliminary normative data, reliability, and MDC values for each of six physical performance measures (ie, SSWV_{OG}, SSWV_{RS}, TSA, STS5, FSST, 6MWT) in a cohort of young, healthy, active-duty service members. The measures are easy to understand, require little equipment, and demonstrate excellent reliability. Additionally, they can be used to assess a range of functional domains that are thought to be impaired in injured service members; these domains are amenable to intervention. Further research is required on the predictive validity, responsiveness, and influence that surgical interventions and rehabilitation have on each of these six physical performance measures to enhance the current level of care for younger civilians and military service members who have sustained polytraumatic extremity injuries.

### References

1. Cross JD, Ficke JR, Hsu JR, Masini BD, Wenke JC: Battlefield orthopaedic


