CHAPTER 2

Test-retest reliability of sensor-based sit-to-stand measures in young and older adults

G. R. H. (Ruben) Regterschot
Wei Zhang
Heribert Baldus
Martin Stevens
Wiebren Zijlstra

Chapter 2

ABSTRACT

This study investigated test-retest reliability of sensor-based sit-to-stand (STS) peak power and other STS measures in young and older adults. In addition, test-retest reliability of the sensor method was compared to test-retest reliability of the Timed Up and Go Test (TUGT) and Five-Times-Sit-to-Stand Test (FTSST) in older adults. Ten healthy young female adults (20-23 years) and 31 older adults (21 females; 73-94 years) participated in two assessment sessions separated by 3-8 days. Vertical peak power was assessed during three (young adults) and five (older adults) normal and fast STS trials with a hybrid motion sensor worn on the hip. Older adults also performed the FTSST and TUGT. The average sensor-based STS peak power of the normal STS trials and the average sensor-based STS peak power of the fast STS trials showed excellent test-retest reliability in young adults (intra-class correlation (ICC) ≥ 0.90; zero in 95% confidence interval of mean difference between test and retest (95% CI of D); standard error of measurement (SEM) ≤ 6.7% of mean peak power) and older adults (ICC ≥ 0.91; zero in 95% CI of D; SEM ≤ 9.9%). Test-retest reliability of sensor-based STS peak power and TUGT (ICC = 0.98; zero in 95% CI of D; SEM = 8.5%) was comparable in older adults, test-retest reliability of the FTSST was lower (ICC = 0.73; zero outside 95% CI of D; SEM = 14.4%). Sensor-based STS peak power demonstrated excellent test-retest reliability and may therefore be useful for clinical assessment of functional status and fall risk.
1. INTRODUCTION

Functional status and fall risk are often evaluated with simple field tests in older adults, such as the Timed Up and Go Test (TUGT) and Five-Times-Sit-to-Stand Test (FTSST) [1-4]. However, these field tests provide a crude outcome measure (time in seconds) with limited clinical relevance. For example, TUGT and FTSST scores do not provide a highly accurate prediction of future falls [5,6]. Therefore, methods that assess additional aspects of movement performance may contribute to the evaluation of functional status and fall risk in older people.

Studies show that leg strength and leg power decline during ageing [7,8], and that reduced strength and power of leg muscles is related to a lower mobility performance and an increased fall risk [8-10]. Therefore, assessment of leg strength and leg power may contribute to the evaluation of functional status and fall risk. The use of body-fixed motion sensors during sit-to-stand (STS) movements may provide clinicians with a practical method to estimate power of leg muscles, since a recent study showed that vertical peak power during the STS transfer can be estimated based on the vertical acceleration signal of hybrid motion sensors worn on the trunk in young and older adults [11].

However, test-retest reliability of STS peak power estimated with hybrid motion sensors is unknown. Therefore, this study investigated test-retest reliability of sensor-based STS peak power and related measures in young and older adults. Young adults were included to investigate whether reliability of sensor-based STS measures is consistent among young and older adults. Assuming consistent reliability in young and older adults, the reliability results in young adults were used to estimate the sample size needed for the older adult reliability study. This study also investigated the smallest number of STS trials necessary for excellent test-retest reliability of sensor-based STS measures. To determine whether the sensor-based STS measures can be of added value to clinicians, this study also compared the test-retest reliability of the sensor-based STS measures to the test-retest reliability of standard clinical measures of functional status (TUGT and FTSST) in older adults. Based on the excellent test-retest reliability of sensor-based STS peak power in young adults, we hypothesized that the test-retest reliability of sensor-based STS peak power would be excellent in older adults and comparable to the test-retest reliability of the TUGT and FTSST in older people.

2. METHODS

2.1 Participants

Ten healthy young female adults were recruited among students from the University of Groningen. They voluntarily decided to participate in this study. Age ranged from 20-23 years (mean±SD: 21.9±1.2 years), height ranged from 1.65-1.84 m (1.73±0.06 m) and body mass ranged from 55.5-77.0 kg (66.0±7.4 kg). All females were free of health issues...
that might influence STS performance. The required number of older adults was calculated using a statistical model [12]. An intra-class correlation (ICC) of 0.75 was defined as the minimally acceptable ICC, since ICC≥0.75 is considered to indicate excellent reliability [13]. The smallest ICC (ICC=0.90) of our most important outcome measure (sensor-based STS peak power) in young adults was chosen as the expected ICC. With two assessments (test and retest) per subject, the statistical model indicated that at least 27 subjects were needed to establish that the expected ICC (0.90) was significantly different from the minimally acceptable ICC (0.75) with a significance level of 0.05 and a power of 0.80.

Older adults were recruited in a residential care home, sheltered houses and a health care center. Inclusion criteria were: ≥70 years of age, being able to walk ≥10 meters (the use of a wheeled walker or cane was allowed), and being able to stand up from a chair. Exclusion criteria were: Cardiovascular/respiratory disorders, neurological disorders, severe comorbidity, cognitive disorders that affect comprehension or execution of the physical tests, simultaneous participation in an intervention or exercise program, orthopedic surgery in the previous six months, visual problems to a degree that makes it impossible to walk or stand up safely from a chair, a stroke within the last six months.

In total 31 healthy older adults (21 females, 10 males) voluntarily decided to participate. Age ranged from 73-94 years (mean±SD: 82.5±4.9 years), height ranged from 1.46-1.89 m (1.66±0.10 m) and body mass ranged from 48.0-104.4 kg (79.9±14.3 kg). Thirteen older adults reported they had fallen at least once in the year before this study. A fall was defined as ‘unintentionally coming to rest on the ground, floor or other lower level’ [14]. Self-reported number of falls in the last year ranged from 0-4.

All participants signed an informed consent before participating in the study. The study was approved by the Medical Ethical Committee of the University Medical Center Groningen, Groningen, the Netherlands.

2.2 Measurements

Subjects participated in two assessment sessions (test and retest), separated by five to seven days (young adults) and three to eight days (older adults). During both assessment sessions, young adults and older adults performed STS movements. Older adults also performed the FTSST and the TUGT during both assessment sessions.

2.2.1 STS movements

2.2.1.1 Test procedure
Young adults performed three normal as well as three fast STS transfers and older adults performed five normal STS transfers. When older adults had no difficulty with standing up from a chair, they were also requested to perform five fast STS transfers. Before rising from the chair, participants were instructed to sit against the back of the chair. Participants were
asked to perform the STS transfers with their arms crossed in front of the chest. After rising from the chair, participants stood still for five seconds before sitting down again. After sitting down, participants were sitting still for 10 seconds before standing up again. A chair with arm rests and a standard height of 0.47 m was used.

2.2.1.2 Data acquisition
During the normal and fast STS movements participants wore a hybrid motion sensor (π-Node, Philips) on the right side of the hip (a small distance above the trochanter major femoris). The hybrid motion sensor consisted of a 3D accelerometer (±2 g), a 3D gyroscope (±300 deg/s) and a 3D magnetometer (±2 Gauss) [15]. Data were sampled with 50 Hz and wirelessly transmitted to a PC by using a proprietary multipoint packetized radio protocol [16].

2.2.1.3 Data analysis
Matlab (The Mathworks, Inc.; version 7.12) was used for data analysis. The sensor orientations and accelerations in the global coordinate system were calculated by using quaternions and the data of the accelerometer, gyroscope and magnetometer in the sensor coordinate system [15]. Subsequently, the acceleration data in the global coordinate system were filtered with a low-pass Butterworth filter using 3 Hz as cut-off frequency.

The low-pass filtered vertical acceleration data were used to calculate the following STS measures:

1. **STS duration**: Time between the initiation of the forward trunk rotation before rising (identifiable as the first deflection of the acceleration compared to gravity) and the first intersection of the acceleration signal with the acceleration due to gravity, after the deceleration phase (for further explanation see Figure 3 in [17]).
2. **Maximal acceleration**: The maximal acceleration during the interval of STS duration.
3. **Maximal jerk**: The highest positive jerk during the acceleration phase of the rising movement. Jerk was computed as: \(Jerk = \frac{a_i - a_j}{(1/f_s)}\). The \(a\) represents acceleration, \(i\) the sample and \(f_s\) the sample frequency.
4. **Maximal velocity**: The maximal velocity during the interval of STS duration. Velocity was estimated by numerical integration of acceleration. The assumption was used that velocity was 0 m/s at the start of the STS duration interval.
5. **Peak power**: The peak power during the STS duration interval. Power was calculated for every sample in the STS duration interval [11]. Force and velocity were multiplied to estimate power: \(P_i = F_i \cdot v_i\). Force was calculated using the formula: \(F_i = m \cdot a_i\). The \(m\) stands for body mass.
6. **Scaled peak power**: Peak power scaled by body mass \((m)\), body height \((l)\) and gravity \((g)\). Scaled power was calculated as: \(P_{scaled} = P / (m \cdot g^{1.5} \cdot l^{0.5})\) [18].
2.2.2 FTSST
For the FTSST [2] older adults were asked to stand up from a chair (height 0.47 m) five times as fast as possible with the arms crossed in front of the chest. Start position was sitting against the back of the chair. Participants were instructed to reach a vertical trunk position when sitting down and not to touch the back of the chair. Time was measured until participant’s buttocks touched the chair after the fifth rising movement. Participants practiced the FTSST once before being timed.

2.2.3 TUGT
For the TUGT [1] older adults were instructed to rise from a chair (height 0.47 m), walk 3 m to a cone, turn around the cone, walk back to the chair, and sit down again. The test was performed at a self-preferred movement speed. The use of arm rests and walking aids (cane or wheeled walker) was allowed. Start and end position was sitting against the back of the chair. Time needed to complete the test was measured. Participants practiced the TUGT once before being timed.

2.3 Statistical analysis
Test-retest reliability of sensor-based STS measures was calculated separately for normal and fast STS movements. Reliability of sensor-based STS measures was calculated based on the first STS trial of the test and retest session as well as based on the average of two to three STS trials (young adults) and the average of two to five STS trials (older adults).

Test-retest reliability consists of relative and absolute test-retest reliability [19-20]. Relative test-retest reliability was examined by calculation of the ICC (two-way mixed model, type consistency) [19-23]. ICC’s were interpreted as follows: excellent reliability ICC≥0.75, fair to good reliability 0.40≤ICC<0.75 and poor reliability ICC<0.40 [13].

Absolute test-retest reliability was evaluated with the 95% confidence interval of the mean difference between test and retest (95%CI of D) as well as by the estimation of the standard error of measurement (SEM). The 95%CI of D was estimated to investigate whether a bias existed between test and retest and calculated as: 95%CI of D=D±t_{n-1}*SE [23]. In this formula n represents the number of subjects, t the value of the t-distribution corresponding to n-1 degrees of freedom, and SE=SD_{diff}/√n with SD_{diff} representing the standard deviation of differences. Zero in the 95%CI of D was considered to indicate excellent absolute reliability. The SEM is a measure of the precision of individual scores on a test [22]. The SEM was estimated as: SEM=SD*√(1-ICC), with SD being the SD of the scores from all subjects [22]. A SEM smaller than 10% of the mean test and retest score was considered to indicate excellent absolute reliability.

The smallest number of STS trials necessary for excellent test-retest reliability was defined as the smallest number of trials which resulted in ICC≥0.75, SEM<10% and zero in 95%CI of D. Statistical analyses were performed with Matlab (The Mathworks, Inc.; version 7.12).
3. RESULTS

In young adults two normal and two fast STS trials were excluded to prevent influence of missing samples on STS outcomes. The excluded normal STS trials concerned the first trial of a subject during the test session and the first trial of another subject during the retest session. The excluded fast STS trials concerned the second trial of a subject and the third trial of another subject during the test session.

In older adults three normal STS trials were excluded to avoid an effect of missing samples on the outcomes of this study. The exclusion concerned the third and fourth trial of a subject during the test session, and the third trial of another subject during the retest session. In total 18 older adults performed the five fast STS trials during test and retest session. Three older adults were not able to perform the FTSST.

3.1 Reliability sensor-based STS measures in young adults

Highest reliability outcomes were obtained by averaging over three STS trials. Based on the average of three normal STS trials, all STS measures demonstrated excellent absolute and relative reliability, except duration and maximal jerk (Table 1). Maximal velocity, peak power and scaled peak power also revealed excellent reliability based on one normal STS trial (Table 1). Based on the average of three fast STS trials, only duration, maximal velocity, peak power and scaled peak power showed excellent absolute and relative reliability (Table 1). Peak power and scaled peak power also demonstrated excellent reliability based on the average of two fast STS trials (Table 1).

3.2 Reliability sensor-based STS measures in older adults

The average score of five STS trials resulted in the highest test-retest reliability results. Based on the average of five normal STS trials, only maximal velocity, peak power and scaled peak power demonstrated excellent relative and absolute reliability (Table 2). Maximal acceleration was the only measure that showed highest absolute reliability when averaged over less than five normal STS trials. Absolute and relative reliability of maximal acceleration were excellent based on the average of three and four normal STS trials. The STS measures that showed excellent reliability based on the average of five normal STS trials also showed excellent reliability based on the average of two normal STS trials (Table 2). Based on the average of five fast STS movements, all measures showed excellent relative and absolute reliability, except maximal jerk (Table 2). Duration, maximal acceleration, maximal velocity and peak power also demonstrated excellent reliability based on the average of less than five fast STS trials (Table 2).
Chapter 2

3.3 Reliability standard clinical measures

The FTSST showed good relative reliability, but poor absolute reliability (Table 2). The TUGT demonstrated excellent absolute and relative reliability (Table 2).

4. DISCUSSION

Sensor-based STS measures showed consistent test-retest reliability results in young and older adults. Based on the average of three (young adults) and five (older adults) normal and fast STS trials, maximal velocity, peak power and scaled peak power showed excellent reliability in young as well as older adults, while STS duration, maximal acceleration and maximal jerk showed mixed reliability results in both groups. Reliability of maximal velocity, peak power and scaled peak power was higher during normal STS than during fast STS in young and older adults. Of these three STS measures, peak power based on the average of five normal STS trials demonstrated highest reliability in older adults. At least two normal STS trials are necessary for a reliable assessment of sensor-based STS peak power in older persons. The test-retest reliability of sensor-based STS peak power based on the average of five normal STS trials was comparable to the test-retest reliability of the TUGT. The FTSST showed lower test-retest reliability.

Several reasons may explain the mixed reliability of STS duration, maximal acceleration and maximal jerk. The relatively low ICC values of normal STS duration may have been caused by relatively low variation in duration between individuals. For example, during the test session the SD of normal STS duration (0.27s) was 12.7% of the mean normal STS duration (2.12s) in older adults, while the SD of normal STS peak power, scaled peak power and maximal velocity ranged between 24.5% and 33.2% of mean values. Maximal acceleration showed consistently the lowest SEM, but almost always a lower ICC than peak power, scaled peak power and maximal velocity. This may also be explained by small variation between individuals. For example, during the test session the SD of maximal acceleration of normal STS (0.29m/s²) was only 2.6% of the mean maximal acceleration during normal STS (11.09m/s²) in older adults. Maximal jerk showed consistently the largest SEM, indicating large individual variation. Jerk was calculated between two subsequent samples and therefore very sensitive to individual variation.

Other studies show comparable TUGT reliability (ICC=0.96-0.99; SEM=1.7s or 11.8%) [1,24], but higher FTSST reliability (ICC=0.89-0.95; SEM=0.9s or 6.3%) [25,26]. The lower FTSST reliability in the present study may be explained by fatigue due to the performance of repeated STS movements, leading to less consistent FTSST performance. This hypothesis is supported by the SEM (2.53s or 14.4%) of the FTSST, which indicated large individual variation in performance. The TUGT may have suffered less from fatigue by repeated STS movements, since the STS transfer is only a small part of the TUGT.
Table 1 | Results on sensor-based STS measures (mean of three STS trials) in young adults. Absolute and relative test-retest reliability results are presented as well as the smallest number of STS trials required for excellent test-retest reliability (ICC≥0.75, SEM<10% and zero in 95%CI of D).

<table>
<thead>
<tr>
<th>Outcome measure</th>
<th>Nr of subjects</th>
<th>Test mean±SD</th>
<th>Retest mean±SD</th>
<th>D±SD4#</th>
<th>95%CI of D</th>
<th>SEM (%)</th>
<th>ICCb</th>
<th>Trials required</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Normal STS</strong></td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration (s)</td>
<td></td>
<td>1.66±0.17</td>
<td>1.64±0.16</td>
<td>-0.01±0.22</td>
<td>-0.20--&gt;0.17</td>
<td>0.14 (8.5%)</td>
<td>0.25</td>
<td>Unknown</td>
</tr>
<tr>
<td>Maximal acceleration (m/s²)</td>
<td></td>
<td>11.65±0.36</td>
<td>11.83±0.39</td>
<td>0.18±0.33</td>
<td>-0.10--&gt;0.46</td>
<td>0.18 (1.6%)</td>
<td>0.75</td>
<td>2</td>
</tr>
<tr>
<td>Maximal jerk (m/s³)</td>
<td></td>
<td>10.32±2.39</td>
<td>12.39±2.91</td>
<td>2.07±2.39</td>
<td>0.07--&gt;4.06</td>
<td>1.40 (12.3%)</td>
<td>0.75</td>
<td>Unknown</td>
</tr>
<tr>
<td>Maximal velocity (m/s)</td>
<td></td>
<td>0.61±0.07</td>
<td>0.63±0.07</td>
<td>0.01±0.04</td>
<td>-0.02--&gt;0.05</td>
<td>0.02 (3.2%)</td>
<td>0.92</td>
<td>1</td>
</tr>
<tr>
<td>Peak power (W)</td>
<td></td>
<td>410.5±69.1</td>
<td>423.3±81.4</td>
<td>-12.8±30.4</td>
<td>-12.6--&gt;38.2</td>
<td>15.0 (3.6%)</td>
<td>0.96</td>
<td>1</td>
</tr>
<tr>
<td>Peak power scaledc</td>
<td></td>
<td>0.15±0.02</td>
<td>0.16±0.02</td>
<td>0.00±0.01</td>
<td>-0.01--&gt;0.01</td>
<td>0.01 (3.6%)</td>
<td>0.92</td>
<td>1</td>
</tr>
<tr>
<td><strong>Fast STS</strong></td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration (s)</td>
<td></td>
<td>1.29±0.19</td>
<td>1.24±0.16</td>
<td>-0.04±0.13</td>
<td>-0.15--&gt;0.07</td>
<td>0.07 (5.5%)</td>
<td>0.84</td>
<td>3</td>
</tr>
<tr>
<td>Maximal acceleration (m/s²)</td>
<td></td>
<td>13.51±0.49</td>
<td>13.40±0.77</td>
<td>-0.10±0.71</td>
<td>-0.70--&gt;0.49</td>
<td>0.41 (3.1%)</td>
<td>0.56</td>
<td>Unknown</td>
</tr>
<tr>
<td>Maximal jerk (m/s³)</td>
<td></td>
<td>24.70±2.83</td>
<td>21.18±4.69</td>
<td>-3.52±5.84</td>
<td>-8.41--&gt;1.36</td>
<td>4.77 (20.8%)</td>
<td>-0.31</td>
<td>Unknown</td>
</tr>
<tr>
<td>Maximal velocity (m/s)</td>
<td></td>
<td>0.92±0.11</td>
<td>0.94±0.14</td>
<td>0.02±0.09</td>
<td>-0.06--&gt;0.10</td>
<td>0.05 (5.2%)</td>
<td>0.84</td>
<td>3</td>
</tr>
<tr>
<td>Peak power (W)</td>
<td></td>
<td>669.9±118.1</td>
<td>698.2±175.1</td>
<td>28.3±90.5</td>
<td>-47.4--&gt;103.9</td>
<td>46.1 (6.7%)</td>
<td>0.90</td>
<td>2</td>
</tr>
<tr>
<td>Peak power scaledc</td>
<td></td>
<td>0.25±0.04</td>
<td>0.26±0.05</td>
<td>0.01±0.03</td>
<td>-0.02--&gt;0.04</td>
<td>0.02 (6.4%)</td>
<td>0.88</td>
<td>2</td>
</tr>
</tbody>
</table>

a SEM expressed as percentage of the mean test and retest score.
b ICC two-way mixed model, type consistency, average measures.
c Dimensionless numbers.
Table 2: Results on sensor-based STS measures (mean of five STS trials) and standard clinical tests in older adults. Absolute and relative test-retest reliability results are presented as well as the smallest number of STS trials required for excellent test-retest reliability (ICC ≥ 0.75, SEM < 10% and zero in 95% CI of D).

<table>
<thead>
<tr>
<th>Outcome measure</th>
<th>Nr of subjects</th>
<th>Test mean±SD</th>
<th>Retest mean±SD</th>
<th>D±SD_{diff}</th>
<th>95% CI of D</th>
<th>SEM (%)</th>
<th>ICCb</th>
<th>Trials required</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal STS</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration (s)</td>
<td>2.12±0.27</td>
<td>2.07±0.33</td>
<td>-0.05±0.30</td>
<td>-0.16→0.07</td>
<td>0.17 (8.3%)</td>
<td>0.66</td>
<td></td>
<td>Unknown</td>
</tr>
<tr>
<td>Maximal acceleration (m/s²)</td>
<td>11.09±0.29</td>
<td>11.21±0.37</td>
<td>0.12±0.25</td>
<td>0.02→0.21</td>
<td>0.14 (1.2%)</td>
<td>0.83</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Maximal jerk (m/s³)</td>
<td>7.85±2.71</td>
<td>8.34±2.55</td>
<td>0.49±2.15</td>
<td>-0.33→1.31</td>
<td>1.17 (14.5%)</td>
<td>0.80</td>
<td></td>
<td>Unknown</td>
</tr>
<tr>
<td>Maximal velocity (m/s)</td>
<td>0.49±0.12</td>
<td>0.52±0.14</td>
<td>0.02±0.07</td>
<td>0.00→0.05</td>
<td>0.04 (7.5%)</td>
<td>0.91</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Peak power (W)</td>
<td>389.1±129.2</td>
<td>409.0±137.5</td>
<td>19.9±54.5</td>
<td>-0.8→40.7</td>
<td>27.7 (6.9%)</td>
<td>0.96</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Peak power scaledc</td>
<td>0.12±0.03</td>
<td>0.13±0.03</td>
<td>0.01±0.02</td>
<td>0.00→0.01</td>
<td>0.01 (7.9%)</td>
<td>0.90</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Fast STS</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration (s)</td>
<td>1.62±0.31</td>
<td>1.61±0.29</td>
<td>-0.01±0.20</td>
<td>-0.12→0.09</td>
<td>0.11 (6.7%)</td>
<td>0.87</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Maximal acceleration (m/s²)</td>
<td>11.86±0.51</td>
<td>11.87±0.48</td>
<td>0.01±0.36</td>
<td>-0.17→0.19</td>
<td>0.19 (1.6%)</td>
<td>0.85</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Maximal jerk (m/s³)</td>
<td>11.78±2.99</td>
<td>12.79±3.23</td>
<td>1.02±3.70</td>
<td>-0.82→2.85</td>
<td>2.29 (18.7%)</td>
<td>0.46</td>
<td></td>
<td>Unknown</td>
</tr>
<tr>
<td>Maximal velocity (m/s)</td>
<td>0.71±0.18</td>
<td>0.72±0.15</td>
<td>0.01±0.12</td>
<td>-0.05→0.07</td>
<td>0.06 (8.7%)</td>
<td>0.86</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Peak power (W)</td>
<td>588.7±204.5</td>
<td>597.1±187.6</td>
<td>8.4±113.7</td>
<td>-48.1→64.9</td>
<td>58.5 (9.9%)</td>
<td>0.91</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Peak power scaledc</td>
<td>0.18±0.05</td>
<td>0.19±0.04</td>
<td>0.00±0.03</td>
<td>-0.02→0.02</td>
<td>0.02 (10.0%)</td>
<td>0.84</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Standard clinical tests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Five Times Sit to Stand test (s)</td>
<td>28</td>
<td>18.47±4.78</td>
<td>16.64±4.88</td>
<td>-1.82±3.58</td>
<td>-3.21→0.43</td>
<td>2.53 (14.4%)</td>
<td>0.73</td>
<td>-</td>
</tr>
<tr>
<td>Timed Up and Go test (s)</td>
<td>31</td>
<td>17.06±10.14</td>
<td>16.50±10.28</td>
<td>-0.56±2.09</td>
<td>-1.33→0.20</td>
<td>1.43 (8.5%)</td>
<td>0.98</td>
<td>-</td>
</tr>
</tbody>
</table>

*a* SEM expressed as percentage of the mean test and retest score.

*b* ICC two-way mixed model, type consistency, average measures.

*c* Dimensionless numbers.
This study showed that sensor-based STS peak power and the TUGT have a comparable test-retest reliability in older adults. In addition, the test-retest reliability of sensor-based STS peak power is comparable to the test-retest reliability of the FTSST as reported by other studies. A recent study demonstrated that sensor-based STS peak power is more sensitive to changes in functioning than standard clinical measures (e.g. TUGT) [17]. Together these findings indicate that sensor-based STS peak power may have a higher clinical validity than standard clinical measures of functional status.

Reliability of sensor-based normal STS measures was determined in an older adult sample that seems a representative group of the target population of older adults. This suggests that reliability of sensor-based normal STS measures can be generalized to the target population. Several factors may limit the generalizability of the reliability of sensor-based fast STS measures. First, the sample size used to determine reliability of sensor-based fast STS measures (n=18) was smaller than the necessary sample size calculated using the statistical model (n=27). Second, reliability of sensor-based fast STS measures was determined in older adults who had no difficulty performing STS movements. Hence, it is not possible to generalize the reliability of sensor-based fast STS measures to mobility-limited older adults. Third, fatigue due to repeated STS movements may have influenced reliability of sensor-based fast STS measures. This may explain the lower reliability of maximal velocity, peak power and scaled peak power during fast STS compared to normal STS.

In conclusion, sensor-based STS peak power, scaled peak power and maximal velocity showed excellent test-retest reliability in young and older adults. Of these three STS measures, STS peak power based on the average of five normal STS trials showed highest reliability in older adults. Since leg power is related to functional status and fall risk [8-10], the excellent reliability indicates that sensor-based STS peak power may be useful for clinical assessment of functional status and fall risk. Studies investigating discriminative ability of sensor-based STS measures are in progress.

ACKNOWLEDGMENT
This study was supported by a grant from ZonMw (program ‘Diseasemanagement chronische ziekten’; project number 40-00812-98-09014). The sponsor was not involved in the research, the writing of the manuscript or in the decision to submit the manuscript for publication.
REFERENCES


