TEST–RE-TEST RELIABILITY OF QUADRICEPS MUSCLE STRENGTH MEASURES IN PEOPLE WITH MORE SEVERE CHRONIC OBSTRUCTIVE PULMONARY DISEASE

André NYBERG, PhD, Didier SAEY, PhD, Mickaël MARTIN, MSc and François MALTAIS, MD
From the Heart and Lung Institute, University of Laval, Quebec, Canada

Objective: To investigate the interday test–retest reliability of volitional and non-volitional measurements of isometric quadriceps strength using a strain-gauge in people with severe to very severe chronic obstructive pulmonary disease.

Design: Cross-sectional study. Volitional quadriceps measurements consisted of isometric maximal voluntary contractions. Non-volitional measurements were obtained during magnetic potentiated twitch stimulations of the femoral nerve.

Setting: Research centre laboratory.

Participants: Twenty-four individuals with severe to very severe chronic obstructive pulmonary disease (percentage of predicted forced expiratory volume in 1 s, 37% predicted).

Results: Maximal voluntary contractions and potentiated twitch stimulations measures demonstrated excellent interday test–retest relative reliability (ICC 0.97 and 0.80, respectively), while absolute reliability measures were different between techniques (SEM 1.4 kg, CV 3.2%, MDC 3.9 kg vs SEM 1.5 kg, CV 12.2%, MDC 4.2 kg for maximal voluntary contractions and potentiated twitch stimulations, respectively).

Conclusion: The results supports that maximal voluntary contraction and potentiated twitch stimulation measurements of isometric quadriceps strength are reliable in people with severe to very severe chronic obstructive pulmonary disease as evident excellent relative reliability using both techniques, although the former technique appears to have better absolute reliability.

Key words: pulmonary disease; chronic obstructive; isometric contraction; magnetic stimulation; muscle strength; quadriceps muscle; reproducibility of results.

Accepted Apr 16, 2018; Epub ahead of print xx, 2018
J Rehabil Med 2018; 50: 00–00

Correspondence address: André Nyberg, Institut universitaire de cardiologie et de pneumologie de Québec, (IUCPQ), G1V 4G5 Quebec, Canada. E-mail: andre.nyberg@criucpq.ulaval.ca

Quadriceps muscle dysfunction has been recognized as one of the most important systemic manifestations of chronic obstructive pulmonary disease (COPD), intimately linked with everyday physical activity (1), exercise intolerance (2), health-related quality of life (3), healthcare service use (4), and mortality (5, 6). A corollary of the prognostic and clinical importance of quadriceps muscle dysfunction in COPD is that it is recommended to assess quadriceps muscle strength in this group of people (6). Such an assessment could be performed using either volitional, effort-dependent manoeuvres or non-volitional techniques, such as electrical or magnetic stimulations of the femoral nerve (7, 8). The recent American Thoracic Society (ATS)/European Respiratory Society (ERS) statement on limb muscle dysfunction in COPD (6), recommended that isometric measurements of quadriceps muscle strength using strain-gauges should be favoured over other alternative strategies, due to its simplicity, availability and the quality of information provided (6, 9). However, while volitional, effort-dependent manoeuvres are the most clinically feasible way of measuring quadriceps strength (8), it is perhaps not suitable for all persons, as it relies on the motivation and full cooperation of the individual performing the test (7). Thus, the use of non-volitional strategies, such as magnetic stimulation of the femoral nerve, to evaluate quadriceps strength have received increasing attention during the last decade and has been used to document the effects of different exercise modalities on muscle strength as well as on quadriceps fatigability in people with COPD (10–14).

Irrespective of whether volitional or non-volitional measurement techniques are used to assess quadriceps muscle strength, the reliability of these techniques...
needs to be established before they can be implemented and used in clinical practice. In people with moderate COPD (forced expiratory volume in 1 s (FEV1) 65±21%), isometric (maximal voluntary contraction; MVC) measurement using a strain gauge has demonstrated excellent interday test–retest reliability (15), while the reliability of non-volitional techniques are yet to be fully determined among people with COPD. Other than a reported within-subject test–retest coefficient of variation (CV) of 5±6%, reported in a previous study by Saey et al., to the best of our knowledge, no test–retest reliability assessment of magnetic stimulation of the femoral nerve has been reported previously. Furthermore, the potential reliability of volitional and non-volitional measurements of quadriceps strength in people with more severe disease has not been well documented, and is yet to be determined. Establishing the reliability of these measurements in people with more severe disease is of importance, considering that the prevalence of quadriceps weakness increases with increasing disease severity. For example, the prevalence rate of muscle weakness is approximately 20% in people with mild and moderate COPD and over 40% in more severe disease (16). Thus, the primary objective of this study was to determine absolute and relative interday test–retest reliability of isometric volitional (MVC) and non-volitional (TWp) assessment of isometric quadriceps muscle force in people with severe to very severe COPD. We hypothesized that both measurement strategies would demonstrate excellent relative reliability, as evidenced by an intraclass correlation coefficient (ICC) >0.75.

METHODS

Participants and study design

A total of 24 individuals with stable COPD, 40 years of age or older, who had a cumulative (current or past) smoking history of >10 pack years and a post-bronchodilator FEV1 ≤50% predicted (17) were included. Exclusion criteria were: a recent exacerbation (<6 weeks prior to study participation), neuromuscular and/or orthopaedic disorders that compromised participation in testing, recent cancer, unstable cardiac disease/cardiac stimulator, current asthma, low body-weight or obesity (body mass index <20 or >30 kg/m²), resting O2-pulse saturation <85% or a daily dose >10 mg of systemic prednisone. For the purpose of interday test and retest comparisons, participants attended our research laboratory on 2 occasions during 1 week (minimum 48 h apart). All test procedures were conducted independently by the same assessor, who had approximately 5 years’ experience with quadriceps muscle testing in patients with COPD. In addition, test conditions (time of day, setting, setup of equipment, instructions, etc.) were identical between test occasions. All participants provided written informed consent, and the research was approved by the local ethics committee (CER: 21111). Recruitment and data collection were performed from November 2014 to June 2016. For purposes of characterization, pulmonary function assessment was performed in accordance with recommended guidelines (18, 19), and assessment of body composition was performed by bioelectrical impedance (InBody520, Body Composition Analyzer, Biospace, Seoul, Korea). Also, the COPD assessment test (20) was used to evaluate the impact of COPD in daily life and the level of physical activity was assessed with the Voorrips physical activity questionnaire (21), previously used in COPD (11, 22) (higher score = more active).

Volitional and non-volitional assessment of quadriceps strength

Isometric MVC was assessed using a fixed strain gauge dynamometer with a signal analogue force transducer and amplifier (Biopac MP150; Biopac Systems, Goleta, CA, USA). The fixed strain gauge was positioned to be perpendicular to the leg and parallel to the floor. Magnetic stimulation of the femoral nerve of the dominant leg was performed using a 70-mm figure-of-eight coil powered by 2 coupled magnetic stimulators (Magstim 200 monopulse Bistim; Magstim Co. Ltd, Whitland, Dyfed, UK) (8, 23). Participants were seated in a semi-reclined chair (N-K 330 Exercise table; N-K Products, Elsinore, CA, USA) that provided a 90° knee joint angle and approximately 120° of hip extension to facilitate stimulations of the femoral nerve, as used previously (15, 24). Arms were held crossed in front of the chest throughout testing. At first, to locate the optimal stimulation position, switches were administered at various locations in the vicinity of the femoral triangle with the stimulator intensity set at 50–70% of maximum output. To ensure that maximal twitches were delivered with the magnetic stimulator, a ramp stimulation protocol was performed at the beginning of every session, consisting of 1 unpotentiates twitch stimulation (TW) at increasing intensities (30, 50, 60, 70, 80, 85, 90, 95 and 100% of stimulator output). Supramaximality of the stimulation (i.e. a plateau in twitch force) was assumed if the change in the force of the contractions from the last 2 stimulator intensities (95–100%) were smaller than the mean CV%, the latter calculated, as described by Shechtman (25, 26). Thereafter, participants performed 3 MVCs, each lasting approximately 5 s, with each MVC separated by 30 s of rest. A single twitch was manually administered by the assessor at a stimulator intensity of 100% at the end of the 5 s contraction and another with the leg relaxed 5 s after the MVC manoeuvre (26, 27) in order to determine the superimposed (TWu) and the potentiated twitch (TWp) stimulations force. The percent activation during the MVC manoeuvre was quantified by comparing the superimposed twitches with the potentiated twitches, taking into account the force level at the time for stimulation: % voluntary activation = (1–TWu/TWp) * (quadriceps strength at stimulation/MVC force) *100 (26). The mean of the highest (peak) value of 2 reproducible manoeuvres from 3 attempts (allowing no more than 10% variance) was used for analysis for both MVC and TWp measurement.

Sample size

In order to detect an ICC greater than 0.75 with a confidence interval width of 0.3, based on a study by Tofari et al. (26) on the interday test–retest reliability of volitional and non-volitional assessment of quadriceps muscle force in healthy younger adults, we estimated that 24 participants would be necessary (28). All measurements were planned to be performed by a single assessor, during 2 occasions (test–retest). A convenient sampling method was used for recruitment of participants, i.e. people with COPD at our regional hospital, who, after screening of their medical records appeared to fulfil inclusion criteria, were contacted by a research nurse and invited to participate.
Muscle strength measures in COPD

Analyses

Interday test–retest relative reliability (i.e. day 1 vs day 2) of MVC and TWp force was determined via calculation of ICCs using a 2-way random single measure ICC procedure (ICC 2.1). ICC values were interpreted as <0.4 = poor, 0.4–0.75 = fair to good and >0.75 = excellent, as previously used in COPD (15). The limits of agreement between measures taken on 2 days was established using Bland–Altman plots (29) for all measurements. For each measure, the proportion of data points within 1.96 standard deviations (SD) of the mean difference between test–retest values was used to describe agreement. Pearson correlation coefficients between absolute test–retest differences and individual test–retest mean values were also calculated in order to detect any heteroscedasticity in the data. This was performed considering that the amount of error of measurement should be uncorrelated to the size of the measured value in order to properly estimate ICCs (30). Furthermore, in addition to measuring ICCs (relative reliability), measures of absolute reliability were included in the analysis, including the standard error of measurement (SEM), and the (CV%) (25). In order to facilitate the interpretation of reliability results, the minimal detectable change (MDC) was also calculated as described by Beckerman et al. (31). Pearson correlation coefficients were calculated to assess the relationship between volitional and non-volitional quadriceps strength. R values in the range 0.0–0.1 were considered trivial, 0.1–0.3, small, 0.3–0.5, moderate, 0.5–0.7, large, 0.7–0.9, very large and 0.9–1.0, extremely large (15, 32). The mean TWp/MVC ratio will be used to determine the validity of the twitch measurements (8). A level of 0.05 was considered valid for statistical significance. For data management and statistical analysis, the IBM Statistical package for Social Science (SPSS) version 23.0 was used.

RESULTS

Baseline characteristics of included participants are presented in Table I. One assessor performed all measurements; all 24 participants had complete data for test and retest occasions, giving a total of 48 observations (24 pre–24 post) for MVC and TWp measurements, respectively. The majority were male, had severe airway obstruction, a medium total body fat free mass just below 70%, a low level of physical activity, and a medium impact of the disease in their daily life (mean COPD Assessment Test (CAT) score <20). No evidence of heteroscedasticity was found, as evident by insignificant Pearson correlation coefficients between the magnitude of SEM and the size of the measured values (MVC: r = 0.03, p = 0.89, TWp: r = 0.13, p = 0.56, respectively). Thus, the requirements for applying ICCs analyses were adequate. All participants reached a plateau in torque during the ramp stimulation protocol, as evidenced by the fact that the change from the last 2 stimulator intensities (95–100%) of the ramp protocol (range –4 to 11%) was smaller than the mean CV% for the resting twitch (13%) (p > 0.05) and percentage activation was 94% (4.4%). The mean TWp/MVC ratio for fully potentiated twitches was 0.42 (0.27 for un-potentiated twitches). These results, which are consistent with previous publication on the topic (8), support the validity of the twitch measurements.

Mean values for test and retest days, including relative and absolute interday reliability estimates (ICC, SEM, CV%, MDC), are shown in Table II. ICC values for both measurement techniques were above 0.75, thus indicating excellent relative reliability. No differences existed between test and retest values for MVC (0.5 kg, 95% CI –0.4 to 1.2 kg, p > 0.05) while, for TWp, the retest value was lower than the test value (–1.0 kg, 95% CI –0.1 to –1.9 kg, p > 0.05). This test–retest differences in TWp was mainly explained by 2 individuals with a mean test–retest difference of –5.7 kg (0.8), i.e. beyond the 95% CI limits of agreement of the mean difference between test–retest values (Fig. 3). If removing the data from these subjects absolute reliability data for TWp became similar to those of MVC with no significant differences between test–retest (–0.6 kg, 95% CI –1.3 to 0.1 kg, p > 0.05). These individuals also had a significant difference during the ramp simulation protocol between test–retest, at 90, 95 and 100% of stimulator intensity (mean difference 2.1, 1.4 and 1.6 kg, respectively, all p < 0.05), which was not seen for the whole group (mean difference 0.2

Table I. Participants’ characteristics

<table>
<thead>
<tr>
<th>Measure</th>
<th>COPD (n = 24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years, mean (SD)</td>
<td>66 (6)</td>
</tr>
<tr>
<td>Sex, % males, n (%)</td>
<td>19 (79)</td>
</tr>
<tr>
<td>Pack years, mean (SD)</td>
<td>53 (28)</td>
</tr>
<tr>
<td>Body mass index, kg/m²</td>
<td>24 (4)</td>
</tr>
<tr>
<td>Fat free mass, kg, mean (SD)</td>
<td>46 (9)</td>
</tr>
<tr>
<td>Dominant leg muscle mass, kg, mean (SD)</td>
<td>6.9 (1.6)</td>
</tr>
<tr>
<td>FEV1/FVC, mean (SD)</td>
<td>3.52 (8.3)</td>
</tr>
<tr>
<td>FEV1, %, mean (SD)</td>
<td>36 (9)</td>
</tr>
<tr>
<td>FEV1, I, mean (SD)</td>
<td>0.97 (0.31)</td>
</tr>
<tr>
<td>FVC, %, mean (SD)</td>
<td>81.9 (24.3)</td>
</tr>
<tr>
<td>FVC, I, mean (SD)</td>
<td>2.82 (0.79)</td>
</tr>
<tr>
<td>RV, %, mean (SD)</td>
<td>142 (23)</td>
</tr>
<tr>
<td>DLCO, %, mean (SD)</td>
<td>45 (11)</td>
</tr>
<tr>
<td>MVV, I, mean (SD)</td>
<td>38.7 (12.3)</td>
</tr>
<tr>
<td>Physical activity, score, mean (SD)</td>
<td>4.7 (3.2)</td>
</tr>
<tr>
<td>CAT, score, mean (SD)</td>
<td>19.5 (7.2)</td>
</tr>
</tbody>
</table>

Table II. Mean values for volitional and non-volitional measures of quadriceps muscle strength (kg) including measures of relative and absolute reliability

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean (SD)</th>
<th>Retest Mean (SD)</th>
<th>Relative reliability</th>
<th>Absolute reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>MVC, kg</td>
<td>26.9 (7.7)</td>
<td>27.3 (7.6)</td>
<td>0.97</td>
<td>0.92–0.99</td>
</tr>
<tr>
<td>TWp, kg</td>
<td>11.1 (3.2)</td>
<td>10.0 (3.5)</td>
<td>0.80</td>
<td>0.59–0.91</td>
</tr>
</tbody>
</table>

*Significantly different from test, p < 0.05.

SD: standard deviation; FVC: forced vital capacity, FEV1: forced expiratory volume in 1 second; RV: functional residual capacity; MV; residual volume; DLCO: diffusion capacity for carbon monoxide; MVV: maximal voluntary ventilation; CAT: COPD Assessment Test.
kg at all intensities, all \( p > 0.05 \). Fig. 1 demonstrates the correlation between volitional and non-volitional quadriceps force measurements taken from the first test, while Figs 2 and 3 demonstrate test–retest Bland-Altman analyses for MVC and \( TW_p \), respectively.

**DISCUSSION**

The main finding of the present study was that the interday test–retest reliability of volitional assessment of isometric quadriceps muscle maximal strength in people with COPD with severe to very severe airway obstructions was highly satisfactory, as evident by excellent relative as well as low absolute measures of reliability. Non-volitional assessment of quadriceps strength also demonstrated excellent relative reliability, while absolute measures of reliability were high.

Volitional assessments of quadriceps muscular strength in people with COPD have recently been found to be highly reliable in people with moderate airflow obstructions (15). Our findings indicate that this measurement strategy is also reliable in people with more severe airway obstructions. Establishing the reliability in people with more severe disease is of importance, considering that the prevalence of muscle weakness is higher in this population in comparison with people with mild or moderate airway obstruction (40% vs 20%, respectively) (16). The relative (ICC = 0.97) and absolute reliabilities (MDC 3.9 kg or 14.4% of MVC) found in our study were consistent with those reported in people with less severe COPD (ICC = 0.96, MDC 18.0%) (15). Our findings were also consistent with those made in healthy adults (26), with regard both to measures of relative (ICC 0.97 vs 0.98) and absolute reliability (CV% 3.2% vs 4.5%). To confirm if a maximal activation has occurred, \( TW \) on an ongoing MVC manoeuvre was used with the underlying premise that, if the muscle is not maximally activated, the superimposed twitch will result in an increase in force above that of the volitional force (33). In our sample, we calculated that the voluntary activation was 94%, thus indicating that the observed MVC was close to a true MVC. The observed % voluntary activation in the present study was similar (94% vs 93%) to that previously observed among healthy young adults (26).

The strong relationship between \( TW \) and MVC observed in our data \((r = 0.79)\) was in line, albeit slightly lower, with that previously reported in people with COPD in studies by Polkey et al. (8) \((r = 0.85)\) and Rodrigues et al. \((r = 0.86)\) (15), while larger than repor-
tested by Saey & Trosters (14) \((r=0.61)\) and Ju & Chen (34) \((r=0.40)\), in patients with COPD or Hammegård et al. (35) in healthy subjects \((r=0.40)\). The test–retest difference in MVC \((0.46 \text{ kg})\) in the present trial was lower than that reported by previous investigators \((1.0–1.1 \text{ kg})\) (15, 34).

In line with previous reports (15, 34), our data supports the recommendations of the ATS/ERS on using isometric measurements for assessing quadriceps muscle strength (6), demonstrating high relative and absolute reliability across disease severities (15, 34). Thus, indicating that this technique could be a reliable method, which, in due course, could be used in the clinical setting. However, before implementing MVC measurements of isometric strength becomes a reality in clinical practice, there is a need to establish and validate predictive normal values for isometric quadriceps strength (15). Furthermore, even though the quadriceps muscle is important, the reliability of MVC measurements also needs to be established for other muscle groups.

Non-volitional techniques are alternative and complementary strategies to assess muscle strength. Factors such as the site of the stimulation, the positioning of the person \(\text{e.g. supine/reclined}\), the experience of the tester and the body characteristics may all compromise the quality of these measurements (8, 34, 35). For example, submaximal stimulations have been reported to be more likely in overweight patients and/or those with a thicker skinfold (34, 35), while the positioning of the person affects the length of the tested muscle and hence the results (36). With regard to the experience of the tester, the assessor in the present trial was experienced in both MVC and TW\(_p\) measurements. Similar to volitional assessments, the interday relative reliability was also excellent for TW\(_p\) \((\text{ICC } 0.80)\), comparable to that found recently among healthy young men \((\text{mean age } 23.5 \text{ years}, \text{ICC } 0.82)\) (26). The interday mean difference in TW\(_p\) \((1.0 \text{ kg})\) observed in the present trial is similar to the reported intraday difference in non-volitional isometric quadriceps strength measurements in COPD \((\text{TW\(_p\) } 0.9 \text{ kg})\) (34). Despite these encouraging reliability data, the coefficient of variation was slightly larger in our sample \((\text{CV} \% 12.2\%)\) than previously reported in healthy young men \((\text{CV} \% 6.8–11.8\%)\) (26, 37). As the ICC was fairly similar \((0.80 \text{ vs } 0.82)\), the difference in CV\% across studies can be explained mainly by the larger variation in our sample in relation to mean TW\(_p\) forces compared with the study on healthy adults (26). In addition, the test–retest variation \(\text{(approximately } 10\%)\), as well as the MDC, was large, which probably reduces the ability to detect a real change. Furthermore, as TW\(_p\) measurements could also be used to determine the presence of contractile fatigue after a training session \(\text{(i.e. a reduction in quadriceps force \(\geq 15\%\) after training)}\) (24), further research is also warranted to determine the reliability of TW\(_p\) measurements in these situations. Even though the use of magnetic stimulation is unlikely to offer significant advantages for people who are able to perform a MVC, this technique could be used in patients who could be prohibited from performing a true MVC; e.g. patients in the intensive care unit, although this would require un-potentiated stimulations, not potentiated stimulations as assessed in this study (8).

**Study limitations**

The reduction in TW\(_p\) force on retest days was unexpected. This phenomenon could be explained by 2 individuals, in whom the mean reduction in TW\(_p\) forces at retest exceeded the 95% limits of agreement as illustrated in the Bland-Altman plot \(\text{(Fig. 3)}\). A likely explanation for this is that the positioning of the stimulation probe was different between occasions on these individuals, despite all the care that was taken to ensure a stable positioning from one measurement to the other. This statement is supported by the large differences in force elicited during the ramp stimulation protocol between occasions. Nevertheless, as these individuals negatively affected both measures of relative and absolute reliability for TW\(_p\) measurements, these findings are a limitation of the study and emphasize the importance of standardized testing procedures to achieve reliable measurements. Another potential limitation was that the sample consisted mainly of men, and that a convenient sampling method was used, which could reduce the external validity of our findings.

**Conclusion**

In summary, the results obtained from this study indicate that volitional and non-volitional assessment strategies of isometric quadriceps force using a strain-gauge seem to be reliable in people with severe to very severe COPD, but that MVC measurements seem to have better absolute reliability in this population.

**REFERENCES**


