Journal: PTC; Volume ; Issue:

DOI: 10.3138/PTC.2016-42

Article

The Unsupported Upper Limb Exercise Test in People Without Disabilities: Assessing the Within-Day Test–Retest Reliability and the Effects of Age and Gender

Ana Oliveira, MSc, PT;*† Joana Cruz, MSc, PhD, PT;*†
Cristina Jácome, MSc, PhD, PT;* Alda Marques, MSc, PhD, PT*†

*Respiratory Research and Rehabilitation Laboratory (Lab3R), School of Health Sciences, University of Aveiro (ESSUA), Aveiro, Portugal;

†Institute for Research in Biomedicine, University of Aveiro (ESSUA), Aveiro, Portugal.

Correspondence to: Alda Marques, Respiratory Research and Rehabilitation Laboratory (Lab3R), School of Health Sciences, University of Aveiro (ESSUA), Agras do Crasto – Campus Universitário de Santiago, Edifício 30, 3810–193 Aveiro, Portugal; amarques@ua.pt.

Contributors: All authors designed the study; or collected, analyzed, or interpreted the data; and drafted or critically revised the article; and approved the final draft.

Competing Interests: None declared. This work was partially supported by Fundo Europeu de Desenvolvimento Regional through the Programa Operacional Competitividade e Internacionalização and the Fundação para Ciência e a Tecnologia (FCT) under project UID/BIM/04501/2013 and by National Funds through FCT, in the context of projects UID/CEC/00127/2013 and Incentivo/EEI/UI0127/2014.

Acknowledgements: The authors acknowledge all participants for their enrolment in this research.

Journal: PTC; Volume ; Issue:

DOI: 10.3138/PTC.2016-42

Abstract

Purpose: To estimate the within-day test–retest reliability and standard error of measurement (SEM) of the unsupported upper limb exercise test (UULEX) in adults without disabilities and to determine the effects of age and gender on performance of the UULEX. ***Method:*** A cross-sectional study was conducted with 100 adults without disabilities (44 men, mean age 44.2 [SD 26] y; 56 women, mean age 38.1 [SD 24.1] y). Participants performed three UULEX tests to establish within-day reliability, measured using an intra-class correlation coefficient (ICC) model 2 (two-way random effects) with a single rater (ICC[2,1]) and SEM. The effects of age and gender were examined using two-factor mixed-design analysis of variance (ANOVA) and one-way repeated-measures ANOVA. For analysis purposes, four sub-groups were created: younger adults, older adults, men, and women. ***Results:*** Excellent within-day reliability and a small SEM were found in the four sub-groups (younger adults: ICC[2,1] = 0.88; 95% CI: 0.82, 0.92; SEM ~40 s; older adults: ICC[2,1] = 0.82; 95% CI: 0.72, 0.90; SEM ~50 s; men: ICC[2,1] = 0.93; 95% CI: 0.88, 0.96; SEM ~30 s; women: ICC[2,1] = 0.85; 95% CI: 0.78, 0.91; SEM ~45 s). Younger adults took, on average, 308.24 seconds longer than older adults to perform the test; older adults performed significantly better on the third test ($p < 0.0001$; $\eta^2 = 0.096$). Gender effects were not found ($p > 0.05$). ***Conclusion:*** The within-day test–retest reliability and SEM values of the UULEX may be used to define the magnitude of the error obtained with repeated measures. One UULEX test seems to be adequate for younger adults to achieve reliable results, whereas three tests seem to be needed for older adults.

Key Words: outcome assessment; standard error of measurement; UULEX; within-day reliability.

Résumé

Objectif : évaluer la fiabilité d’un test-retest en une même journée et l’erreur type de mesure (ETM) du test d’exercice des membres supérieurs sans appui (UULEX) chez des adultes sans incapacités et déterminer les effets de l’âge et du sexe sur leur exécution. ***Méthode :*** les chercheurs ont réalisé une étude transversale auprès de 100 adultes sans incapacités (44 hommes, d’un âge moyen de 44,2 ans [ÉT 26], et 56 femmes, d’un âge moyen de 38,1 ans [ÉT 24,1]). Les participants ont effectué trois UULEX pour établir la fiabilité du test-retest en une même journée, mesurés à l’aide du modèle 2 de coefficient de corrélation intraclasse (ICC, effets aléatoires bilatéraux) comportant un ICC(2,1) et une ETM à un seul évaluateur. Les chercheurs

DOI: 10.3138/PTC.2016-42

ont examiné les effets de l'âge et du sexe à l'aide d'une analyse de variance bifactorielle à mesures mixtes et d'une analyse de variance unifactorielle à mesures répétées. Pour les besoins de l'analyse, les chercheurs ont créé quatre sous-groupes: jeunes adultes, adultes plus âgés, hommes et femmes. ***Résultats :*** les quatre sous-groupes affichaient une excellente fiabilité en une même journée et une petite ETM (jeunes adultes: ICC[2,1] = 0,88 [IC 95% : 0,82, 0,92] et ETM~40 secondes; adultes plus âgés : ICC[2,1] = 0,82 [IC 95 % : 0,72, 0,90] et ETM~50 secondes; hommes: ICC[2,1] = 0,93 [IC 95% : 0,88, 0,96] et ETM~30 secondes; femmes: ICC[2,1] = 0,85 [IC 95 % : 0,78, 0,91] et ETM~45 secondes). En moyenne, les jeunes adultes ont effectué le test pendant 308,24 secondes de plus que les adultes plus âgés; ceux-ci ont obtenu un résultat nettement meilleur au troisième test ($p < 0,0001$; $\eta^2 = 0,096$). Il n'y avait pas d'effets selon le sexe ($p > 0,05$). ***Conclusion :*** il est possible d'utiliser la fiabilité de test-retest en une même journée et les valeurs d'ETM de l'UULEX pour définir l'importance de l'erreur obtenue lors de mesures répétées. Un UULEX semble suffire pour que les jeunes adultes obtiennent des résultats fiables, tandis que trois tests semblent nécessaires chez les adultes plus âgés.

Mots-clés : évaluation des résultats cliniques; erreur type de mesure; fiabilité en une même journée; UULEX

Research has defined functional physical fitness as the physiological capacity to safely perform daily physical activities without extreme fatigue.[1] A great number of daily physical activities include upper limb movements,[2] which are particularly affected by the aging process.[3] Aging results in significant declines in cardiorespiratory and muscle fitness,[3] and it may also lead to chronic disease, decreases in the performance of activities of daily living, and, consequently, loss of independence and poor quality of life.[3,4] Increased physical activity levels from a young age have been associated with good indicators of physical fitness and a decreased risk of chronic conditions in later life.[5,6] Thus, promoting physical activity and physical fitness throughout an individual's life is currently an international priority.[7–9]

Non-governmental and governmental organizations' strategies to increase physical activity in populations without disabilities include community exercise programmes.[9] Such

DOI: 10.3138/PTC.2016-42

programmes require the use of simple and economical instruments to evaluate the programmes' impact on participants' health. Simple, easy-to-use, and validated physical tests already exist to assess lower limb exercise capacity in populations without disabilities,[10–12] but few are available to assess upper limb exercise capacity. To our knowledge, the arm crank ergometer is the only exercise-tolerance test of the upper limb that has been studied for reliability in people without disabilities,[13] and good to excellent test–retest (intra-class correlation coefficient [ICC] = 0.76), inter-observer (ICC = 0.82), and inter-ergometer (ICC = 0.63) reliability have been reported.[13] Nevertheless, arm ergometers are expensive and require trained health professionals to conduct and interpret the tests; thus, their use is somewhat limited in community exercise programmes.

The unsupported upper limb exercise test (UULEX) is a symptom-limited, incremental test first described by Takahashi and colleagues[14] to assess upper limb peak exercise capacity in patients with chronic obstructive pulmonary disease (COPD). UULEX test–retest reliability has been studied in patients with COPD, and it has shown excellent ICCs for UULEX total exercise time (ICC = 0.98), with no significant differences among tests conducted within 2–4 days.[14] The UULEX's excellent reliability results, along with its higher portability, ease of implementation, and low cost, make it a promising measure for use in community exercise programmes with people without disabilities.

Nevertheless, evidence shows that the measurement properties of a given test are specific to a particular population.[15] Thus, before recommending use of the UULEX with people without disabilities, its reliability – namely, its within-day test–retest and standard error of measurement (SEM) – should be established for younger and older people without disabilities, as well as for men and women, because significantly different performances on endurance tests have been

found with these populations.[16,17] SEM values will inform health and sports professionals about the minimum number of tests needed to achieve a reliable baseline measure. This information will enhance the assessment of upper limb peak exercise capacity in people without disabilities and the development of exercise programmes for this population.

As a result, this study aimed to estimate the within-day test–retest reliability and SEM in adults without disabilities. The secondary aim was to estimate the role of age and gender in the performance of the UULEX.

Method

Study design

We conducted a cross-sectional study with people without disabilities between September 2012 and September 2015. The reliability sections of this study were described following the guidelines for reporting reliability and agreement studies.[18] Before data collection, the Ethics Committee of the School of Health Sciences, University of Aveiro, Aveiro, Portugal, approved the study, and all participants signed an informed consent sheet.

Participants

Participants without disabilities and aged older than 18 years were recruited from the university campus and surrounding community. Exclusion criteria were the presence of one or more of the following conditions: acute (within the past month) or chronic respiratory disease,

cardiac disease, cognitive or neurological impairment, or significant musculoskeletal disorder (e.g., kyphoscoliosis).

Sample size

The sample size for test–retest reliability was determined according to the study by Bonett,[19] which established that a sample size of 36 individuals was sufficient to estimate an ICC model 2 (two-way random effects) with a single rater ICC(2,1) = 0.8; 95% CI: width of 0.2; $\alpha = 0.05$; $k = 3$.

Measures and procedures

First, we collected socio-demographic (age and gender), anthropometric, and clinical data from the participants. Anthropometric data (weight and height measurements) were used to calculate BMI. Clinical data included smoking status, evaluated through a two-question survey on current and previous smoking habits – that is, number of years as a smoker and the usual quantity of cigarettes smoked over a 24-hour period to calculate the number of pack-years. A brief cardiorespiratory examination, testing for dyspnoea, fatigue, and heart and respiratory rates, was conducted before carrying out the UULEX, and it provided the baseline physiological values. The modified Borg scale (MBS) was used to assess dyspnoea and fatigue[20] because it is the most widely used and recommended scale for monitoring patients during exercise.[21,22]

Finally, participants performed three UULEX tests with a 30-minute rest period between each one.[14] Participants were instructed to sit on a chair with their back resting against the chair back and with their knees, hips, and ankles at approximately 90°. Participants held a plastic bar

(0.2 kg) and faced an A0-size chart (0.84 cm × 120 cm), marked with eight different coloured bands (or levels), each 84 centimetres wide, 8 centimetres high, and spaced 15 centimetres apart (from the middle of one band to the middle of the next) (see Figure 1). The lowest band, which was the starting level, was placed in line with the participants' knees; this position was maintained across the three tests. The highest level of arm elevation was determined by measuring the distance to the highest band on the chart that the participants could reach while seated with their arms fully extended.[14]

(Insert Figure 1 about here.)

Participants were instructed to lift the bar to the first band and then return it to the neutral position (i.e., resting the bar on their thighs); they were to do this for 2 minutes at a constant rate of 30 lifts per minute, marked by a metronome. Afterward, they were asked to progress to the next band, repeating the same task for 1 minute per band. When participants reached the highest band, they received a heavier bar (0.5 kg) and performed the same task for another minute at that band. Thereafter, the weight of the bar was increased by 0.5 kilogram every minute, with the participants always performing the task at the highest band.

Each UULEX test was either continued until participants were exhausted or stopped because they exhibited abnormal physiological responses – that is, they reached 90% of their maximum age-predicted heart rate (206.9 – [0.67 × age])[23] or peripheral oxygen saturation less than 85%, they experienced pain, or they were unable to continue performing the test correctly.[14] Participants were encouraged with standardized sentences each minute, and there was no practice test.[24] As soon as a participant finished each test, we recorded the band reached, total exercise time, self-reported dyspnoea and local muscle fatigue (using the MBS), and peak heart and

respiratory rates. Trained physiotherapists, with experience in applying these tests, collected the data.

Data analysis

We used descriptive statistics to describe the sample. Participants' characteristics were compared between younger (aged 18–50 y) and older (aged ≥ 50 years) adults and between men and women using independent *t*-tests for normally distributed data, Mann–Whitney *U* tests for non-normally distributed and ordinal data, and χ^2 tests for categorical data.

We determined relative and absolute within-day test–retest reliability for the total sample and for each age and gender group; this has been recommended in reliability studies.[25] To assess relative reliability between the tests (i.e., tests 1 and 2, tests 2 and 3, and tests 1 and 3), we used ICC(2,1).[26] ICC values were assigned as follows: more than 0.75 = excellent, 0.40–0.75 = fair to good, and less than 0.40 = poor.[27]

We calculated SEM because it is a measure of absolute reliability (i.e., it indicates the extent to which a score varies on repeated measurements) and because it provides a value for measurement error in the same units as the measurement itself. For these reasons, it can also be used in everyday clinical practice.[28] We also calculated 95% CI.[29]

The differences between the tests were assessed using two-way mixed-design analysis of variance (ANOVA) to determine whether there were significant differences in total exercise time, dyspnoea, fatigue, peak heart rate, and peak respiratory rate (dependent variables) between the tests and whether there was any interaction between these differences and age (comparison 1 independent variable) or gender (comparison 2 independent variable).

Differences among the three tests between the age (independent variable) and gender (independent variable) groups were assessed using general linear models for repeated measures (one-way ANOVA) for total exercise time, dyspnoea, fatigue, peak heart rate, and peak respiratory rate (dependent variables). If we found a statistically significant difference among the tests, post hoc comparisons between tests (i.e., tests 1 and 2, tests 2 and 3, and tests 1 and 3) were performed and corrected for multiple comparisons using the Bonferroni method.[30]

We performed all statistical analyses using IBM SPSS Statistics, version 20.0 (IBM Corp., Armonk, NY), and plots were created using GraphPad Prism, version 5.01 (GraphPad Software, San Diego, CA). All tests were two-tailed, and an effect was considered statistically significant at $p < 0.05$.

Results

Participant characteristics

A total of 142 adults without disabilities were screened for participation in the study. Of these, 7 declined to participate, and 35 did not complete the three UULEX tests. One hundred participants completed the study (44 men, mean age 44.2 [SD 26] y, and 56 women, mean age 38.1 [SD 24.1] y). At baseline, older adults and men presented with a significantly higher BMI than younger adults (mean BMI 28.9 [SD 5.3] vs. 22.8 [SD 4.0], $p < 0.001$) and women (mean BMI 26.0 [SD 4.8] vs. 24.3 [SD 5.7], $p = 0.038$), respectively, and were classified as overweight according to World Health Organization criteria.[7] Older adults also smoked significantly more pack-years than younger adults (median number of pack-years 2.7 [first–third quartile 0.0–18.3]

vs. 0.2 [first–third quartile 0.0–0.5], $p = 0.044$). The remaining variables did not present significant differences between groups. (Participant characteristics are presented in Table 1.)

(Insert Table 1 about here.)

Within-day test–retest reliability and standard error of measurement

Table 2 presents the ICC and SEM values obtained between tests 1 and 2, tests 2 and 3, and tests 1 and 3. We found excellent within-day test–retest relative reliability among the three tests for the total exercise time, both in the total sample (ICC[2,1] = 0.91, 95% CI: 0.87, 0.93) and in all sub-groups (younger adults: ICC[2,1] = 0.88, 95% CI: 0.82, 0.92; older adults: ICC[2,1] = 0.82, 95% CI: 0.72, 0.90; men: ICC[2,1] = 0.93, 95% CI: 0.88, 0.96; women: ICC[2,1] = 0.85, 95% CI: 0.78, 0.91). Small SEM values were identified for these four sub-groups (approximately 40 seconds, 50 seconds, 30 seconds, and 45 seconds for younger adults, older adults, men, and women, respectively).

(Insert Table 2 about here.)

Independent of age and gender, tests 1 and 2 seemed to present higher within-day reliability values and lower SEM values than tests 2 and 3 and tests 1 and 3 (see Table 2).

Effect of age

We found a significant interaction between the time spent performing the UULEX and participants' age ($p = 0.024$, $\eta^2 = 0.038$), suggesting that younger adults were able to perform the tests longer (on average, 308.24 s longer) than older adults. When comparing the tests within

DOI: 10.3138/PTC.2016-42

each age group, no differences were found among the three tests in total test time in the younger group ($p = 0.67$, $\eta^2 = 0.006$). However, older adults performed significantly better on the third test than on the first and second tests ($p = 0.028$, $\eta^2 = 0.096$).

Regarding participants' performance by age group, 36 (56%) younger adults achieved their best result (i.e., longest total exercise time) on one of the first two tests, and 28 (50%) older adults achieved their best result on one of the last two tests, with the majority ($n = 17$, 47%) performing the best on the last one. Eight participants in the younger group (13%) presented exactly the same result among the three tests (5%) or between two of the three tests (8%). Mean differences in the younger group were 7.5 seconds (1%) between the first and second tests, –12.4 seconds (–0.9%) between the second and third tests, and –4.9 (–0.3%) between the first and third tests. Mean differences in the older adult group were –15.1 seconds (1%) between the first and second tests, 52.8 seconds (21.4%) between the second and third tests, and 37.6 seconds (22.4%) between the first and third tests.

We did not observe any interaction between the clinical parameters at the end of the UULEX and age (dyspnoea, $p = 0.38$; fatigue, $p = 0.18$; heart rate, $p = 0.18$; respiratory rate, $p = 0.16$). However, we found significant increases from baseline and after each test among all clinical parameters in both groups ($p < 0.05$), except for dyspnoea in the older group (see Table 3 and Figure 2). Both younger and older participants ended the tests mainly because of fatigue (younger group, 92%; older group, 81%). Other reasons for test interruption were performing a test incorrectly (younger group, 6%; older group, 6%), low back pain (younger group, 2%; older group, 11%), and dyspnoea (younger group, 0%; older group, 3%).

(Insert Table 3 and Figure 2 about here.)

DOI: 10.3138/PTC.2016-42

Effect of gender

We did not find significant interactions between time spent performing the UULEX and participants' gender ($p = 0.43$, $\eta^2 = 0.008$). In addition, when comparing the tests within each group, we did not find differences among the three tests in total exercise time, either in men ($p = 0.34$, $\eta^2 = 0.025$) or in women ($p = 0.75$, $\eta^2 = 0.005$).

Both men ($n = 18$, 41%) and women ($n = 19$, 34%) achieved their best result on test 3. Mean differences in men were 12.4 seconds (2.2%) between tests 1 and 2, 13.8 seconds (9.6%) between tests 2 and 3, and 26.2 seconds (13.0%) between tests 1 and 3. Mean differences in women were –10.9 seconds (0.4%) between tests 1 and 2, 8.9 seconds (5.2%) between tests 2 and 3, and –2.0 seconds (3.8%) between tests 1 and 3.

We did not observe any interaction between the clinical parameters at the end of the UULEX and gender (dyspnoea, $p = 0.40$; fatigue, $p = 0.80$; heart rate, $p = 0.64$; respiratory rate, $p = 0.97$). When analyzing the groups independently, we found significant increases from baseline and after each test among all clinical parameters ($p < 0.05$). We also observed significant increases in heart rate in both men and women between test 1 and tests 2 and 3 (see Table 4 and Figure 3). Both men and women ended the tests mainly because of fatigue (men, 86.4%; women, 83.9%). Other reasons for test interruption were performing a test incorrectly (men, 2.3%; women, 5.4%), low back pain (men, 11.4%; women, 7.1%), and dyspnoea (men, 0%; women, 1.8%).

(Insert Table 4 and Figure 3 about here.)

Discussion

DOI: 10.3138/PTC.2016-42

Excellent within-day test–retest reliability and small SEM values were found for UULEX total test time in participants without disabilities. Younger adults were able to perform the test longer (on average, 308.24 s longer) than older adults, and significant improvements were observed from the first to the third test in older individuals. Gender did not influence the time spent performing the UULEX or the number of tests needed to achieve the best performance.

Our study found excellent values for within-day test–retest relative reliability (ICCs 0.79–0.95); these values were similar to those reported for the arm crank ergometer (ICC = 0.76), the gold standard for assessing upper limb physical fitness.[13] Nevertheless, slightly higher reliability values have been found in studies that included patients with COPD (ICC = 0.98).[14] These differences may be due to the higher variability of the sample recruited in the present study. Patients with COPD recruited from hospitals are often more homogeneous in their characteristics than people from the community. Important confounders in people recruited in the community can include substantially different levels of physical activity, fitness status, and motivation. Such factors should be taken into account when interpreting these results and in further investigations.

Other studies that have assessed the reliability of physical fitness tests, such as the 6-minute walk test (ICC 0.88–0.94) and the 5-times-sit-to-stand test (ICC = 0.72), conducted with community-dwelling older adults, have presented results similar to ours.[31] Nevertheless, our study conducted different types of reliability tests than previous studies (i.e., within- vs. between-day reliability); thus, caution should be taken when establishing comparisons.

According to our best knowledge, this was the first study to assess the SEM of the UULEX total time; the results were approximately 40 seconds, 50 seconds, 30 seconds, and 45

seconds for younger adults, older adults, men, and women, respectively. The SEM values are an estimate of repeated measures;[32] this means that if an individual exceeds this value in a repeated measure, a further test might be necessary. In addition, the SEM is considered a fixed characteristic of a measure, regardless of the population under investigation.[28] Thus, the SEM values of this study may apply not only to people without disabilities but also to other populations, in which repeated measures might be needed to reach an individual's real performance on the UULEX.

The analysis of group performance showed that older adults significantly increased their test time on the third test (22.4%), whereas no differences were found between tests in younger adults, men, or women. This substantial increase from the first and second tests to the third test may be due to familiarization with the test, motivation, or learning effects. Although simple and easy to perform, the UULEX requires participants to learn a movement pattern and synchronize it with a sound signal. Thus, it can be argued that, by the third test, older adults felt more familiar with these conditions and their anxiety levels decreased, allowing them to achieve better performance. Also, the UULEX is physically demanding[14] and, knowing that the third test was the last one to be performed, older people may have been motivated to perform it at their maximum level.

Finally, and similar to the suggestion made by Takahashi and colleagues,[14] a learning effect may have occurred. Other studies that have assessed the repeatability of fitness tests performed in community-dwelling older adults without disabilities have reported the need for a practice run to overcome a learning effect.[14,33,34] Compared with young adults, older adults may present with some level of cognitive impairment, reduced perceived ability, or diminished self-

efficacy, which may influence their test performance.[33] Therefore, a larger number of repetitions may be needed for older adults to familiarize themselves with the UULEX and achieve a better performance.

A drawback of using the UULEX in clinical practice, especially considering the results obtained for older adults, may be the number of tests needed to achieve a person's real performance. Considering that at least three tests have to be performed, approximately 90 minutes must be set aside to complete the evaluation, and this may not be feasible in a community-based program. Thus, our conclusions should be read with caution, considering that we analyzed only within-day test–retest reliability. For example, a previous study conducted with older patients with COPD found low variability in UULEX time in three repetitions conducted 2–4 days apart.[14] In our study, participants were requested to stay in the data collection facilities for approximately 90 minutes, until the end of the third test, whereas in the study by Takahashi and colleagues,[14] participants could go back to their daily activities and return another day to repeat the test. Although our within-day assessment could have resulted in participants quickly becoming familiar with the UULEX because the tests were carried out so close to each other, it may also have reduced participants' motivation because they spent large amounts of time without performing any activity, apart from the UULEX.

Finally, participants' signs and symptoms after each test increased significantly when compared with baseline values, but they were no different among the three tests. This was an expected and desirable result, showing that, at the end of the UULEX, participants were close to their maximum level of performance. Fatigue was the main cause for ending the tests, a result that agrees with those of previous studies.[14,35] Younger participants without disabilities took

approximately 5 minutes longer to achieve exhaustion than older adults without disabilities and patients with COPD.[14,35] This substantial increase in the time needed to complete the UULEX may discourage health professionals from using it in their clinical practice. Thus, as with other tests of exercise capacity,[12,36] it might be valuable to modify the UULEX protocol for people without disabilities, especially young adults. One practical suggestion is to further increase the weight of the bar or increase the rhythm of the metronome.

Our study has some limitations that need to be acknowledged. First, it was set up only for ICC tests because the main aim was to present UULEX reliability in people without disabilities. Nevertheless, we acknowledge that the lack of power calculations for other statistical tests may have affected the study's external validity and therefore limit the conclusions that can be drawn about the effects of age and gender on performing the UULEX. That is, the small number of older participants enrolled may have contributed to the absence of a plateau across the three tests of UULEX in this group. Nevertheless, we verified the homogeneity of the sample in both groups, and other reliability studies that have used walk tests with patients with COPD have also concluded that patients' performance increases, even after taking nine measurements.[37]

In addition, we estimated differences in age and gender groups independently and without considering any interactions between them. It would be interesting to assess whether there were gender differences in the UULEX outcomes of the younger and older adults; however, our samples were too small to draw such conclusions. Future studies could use our results to conduct power analysis and perform these comparisons because age and gender may have a concurrent effect on performance of the UULEX.

Another limitation of our study is the absence of between-day reliability data. It is well known that studies in which repeated tests are performed at short time intervals can yield better reliability values than those in which repeated tests are performed at longer time intervals (i.e., days or weeks).[15,38–40] Our study aimed to assess UULEX reliability to establish the upper limb peak exercise capacity of people without disabilities at a given moment; if it is to be used to assess the effects of an intervention, between-day reliability tests should also be conducted.

Conclusions

The UULEX showed adequate within-day reliability after two and three tests, and small SEM values were found for younger and older adults and for men and women. Our findings also suggest that a single UULEX test may be sufficient to achieve individual real performance in younger adults, men, and women, and at least three tests are needed for older adults.

These results may be useful in clinical practice to define what can be expected and what represents a real change in repeated measures in people without disabilities.

Key Messages

What is already known on this topic

The unsupported upper limb exercise test (UULEX) is a simple and cost-effective upper limb test widely used in patients with chronic obstructive pulmonary disorder. It also has great potential for guiding the development and evaluation of community-based exercise programmes

in younger and older populations without disabilities. Nevertheless, its measurement properties have not been well studied.

What this study adds

Our study found excellent within-day test–retest reliability and small SEM values for UULEX test time in adults without disabilities. It also found that at least three UULEX tests are needed for older adults to achieve real performance, whereas only one test is needed for younger adults. No significant differences were found in gender in the performance of the UULEX. These results can be directly applied in clinical practice to define what can be expected and what represents a real change in repeated measures.

References

<jrn>1. Toraman F, Sahin G. Age responses to multicomponent training programme in older adults. Disabil Rehabil. 2004;26(8):448–54. https://doi.org/10.1080/09638280310000166312. Medline:15204466</jrn>

<jrn>2. Magermans DJ, Chadwick EK, Veeger HE, et al. Requirements for upper extremity motions during activities of daily living. Clin Biomech (Bristol, Avon). 2005;20(6):591–9. https://doi.org/10.1016/j.clinbiomech.2005.02.006. Medline:15890439</jrn>

<jrn>3. Milanović Z, Pantelić S, Trajković N, et al. Age-related decrease in physical activity and functional fitness among elderly men and women. Clin Interv Aging. 2013;8:549–56. https://doi.org/10.2147/CIA.S44112. Medline:23723694</jrn>

<jrn>4. Paterson DH, Jones GR, Rice CL. Ageing and physical activity: evidence to develop exercise recommendations for older adults. Can J Public Health. 2007;98(Suppl 2):S69–108. Medline:18213941</jrn>

<jrn>5. Kujala UM. Physical activity, genes, and lifetime predisposition to chronic disease. Eur Rev Aging Phys Act. 2011;8(1):31–6. https://doi.org/10.1007/s11556-010-0077-2.</jrn>

<jrn>6. Durstine JL, Gordon B, Wang Z, et al. Chronic disease and the link to physical activity. J Sport Health Sci. 2013;2(1):3–11. https://doi.org/10.1016/j.jshs.2012.07.009.</jrn>

<bok>7. World Health Organization. Diet, physical activity and health – Report by the Secretariat Fifty-Fifth World Health Assembly. Geneva: World Health Organization; 2002.</bok>

<jrn>8. Kohl HW III, Craig CL, Lambert EV, et al.; Lancet Physical Activity Series Working Group. The pandemic of physical inactivity: global action for public health. Lancet. 2012;380(9838):294–305. https://doi.org/10.1016/S0140-6736(12)60898-8. Medline:22818941</jrn>

<jrn>9. Bauman A, Merom D, Bull FC, et al. Updating the evidence for physical activity: summative reviews of the epidemiological evidence, prevalence, and interventions to promote “active aging.” Gerontologist. 2016;56(Suppl 2):S268–80. https://doi.org/10.1093/geront/gnw031. Medline:26994266</jrn>

<jrn>10. Kervio G, Carre F, Ville NS. Reliability and intensity of the six-minute walk test in healthy elderly subjects. Med Sci Sports Exerc. 2003;35(1):169–74. https://doi.org/10.1097/00005768-200301000-00025. Medline:12544651</jrn>

<jrn>11. Gabriel KKP, Rankin RL, Lee C, et al. Test-retest reliability and validity of the 400-meter walk test in healthy, middle-aged women. J Phys Act Health. 2010;7(5):649–57. https://doi.org/10.1123/jpah.7.5.649. Medline:20864761</jrn>

<jrn>12. Dourado VZ, Guerra RLF, Tanni SE, et al. Reference values for the incremental shuttle walk test in healthy subjects: from the walk distance to physiological responses. J Bras Pneumol. 2013;39(2):190–7. https://doi.org/10.1590/S1806-37132013000200010. Medline:23670504</jrn>

<jrn>13. Bulthuis Y, Drossaers-Bakker W, Oosterveld F, et al. Arm crank ergometer is reliable and valid for measuring aerobic capacity during submaximal exercise. J Strength Cond Res. 2010;24(10):2809–15. https://doi.org/10.1519/JSC.0b013e3181e31242. Medline:20885199</jrn>

<jrn>14. Takahashi T, Jenkins SC, Strauss GR, et al. A new unsupported upper limb exercise test for patients with chronic obstructive pulmonary disease. J Cardiopulm Rehabil. 2003;23(6):430–7. https://doi.org/10.1097/00008483-200311000-00007. Medline:14646791</jrn>

<jrn>15. Mesquita R, Janssen DJ, Wouters EF, et al. Within-day test-retest reliability of the Timed Up & Go test in patients with advanced chronic organ failure. Arch Phys Med Rehabil. 2013;94(11):2131–8. https://doi.org/10.1016/j.apmr.2013.03.024. Medline:23583345</jrn>

<jrn>16. Shvartz E, Reibold RC. Aerobic fitness norms for males and females aged 6 to 75 years: a review. Aviat Space Environ Med. 1990;61(1):3–11. Medline:2405832</jrn>

<jrn>17. Probst VS, Hernandes NA, Teixeira DC, et al. Reference values for the incremental shuttle walking test. Respir Med. 2012;106(2):243–8. https://doi.org/10.1016/j.rmed.2011.07.023. Medline:21865021</jrn>

<jrn>18. Kottner J, Audigé L, Brorson S, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. J Clin Epidemiol. 2011;64(1):96–106. https://doi.org/10.1016/j.jclinepi.2010.03.002. Medline:21130355</jrn>

<jrn>19. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. Stat Med. 2002;21(9):1331–5. https://doi.org/10.1002/sim.1108. Medline:12111881</jrn>

<bok>20. Borg G. Borg's Perceived Exertion and Pain Scales. Champaign (IL): Human Kinetics; 1998.</bok>

<jrn>21. Spruit MA, Troosters T, Trappenburg JC, et al. Exercise training during rehabilitation of patients with COPD: a current perspective. Patient Educ Couns. 2004;52(3):243–8. https://doi.org/10.1016/S0738-3991(03)00098-3. Medline:14998593</jrn>

<jrn>22. Crisafulli E, Clini EM. Measures of dyspnea in pulmonary rehabilitation. Multidiscip Respir Med. 2010;5(3):202–10. https://doi.org/10.1186/2049-6958-5-3-202. Medline:22958431</jrn>

<jrn>23. Gellish RL, Goslin BR, Olson RE, et al. Longitudinal modeling of the relationship between age and maximal heart rate. Med Sci Sports Exerc. 2007;39(5):822–9. https://doi.org/10.1097/mss.0b013e31803349c6. Medline:17468581</jrn>

<jrn>24. Janaudis-Ferreira T, Hill K, Goldstein RS, et al. Relationship and responsiveness of three upper-limb tests in patients with chronic obstructive pulmonary disease. Physiother Can. 2013;65(1):40–3. https://doi.org/10.3138/ptc.2011-49. Medline:24381380</jrn>

<jrn>25. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. Clin Rehabil. 1998;12(3):187–99. https://doi.org/10.1191/026921598672178340. Medline:9688034</jrn>

<jrn>26. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;86(2):420–8. https://doi.org/10.1037/0033-2909.86.2.420. Medline:18839484</jrn>

<bok>27. Fleiss J. The design and analysis of clinical experiments. New York: Wiley-Interscience; 1986.</bok>

<jrn>28. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005;19(1):231–40. Medline:15705040</jrn><jrn>29. 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(7):745–50. https://doi.org/10.1093/ptj/77.7.745. Medline:9225846</jrn>

<jrn>30. Dunn OJ. Multiple comparisons using rank sums. Technometrics. 1964;6(3):241–52. https://doi.org/10.1080/00401706.1964.10490181.</jrn>

<jrn>31. Mijnarends DM, Meijers JM, Halfens RJ, et al. Validity and reliability of tools to measure muscle mass, strength, and physical performance in community-dwelling older people: a systematic review. J Am Med Dir Assoc. 2013;14(3):170–8. https://doi.org/10.1016/j.jamda.2012.10.009. Medline:23276432</jrn>

<bok>32. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. 3rd ed. Upper Saddle River (NJ): Prentice Hall; 2009.</bok>

<jrn>33. Ferrini R, Edelstein S, Barrett-Connor E. The association between health beliefs and health behavior change in older adults. Prev Med. 1994;23(1):1–5. https://doi.org/10.1006/pmed.1994.1001. Medline:8016025</jrn>

<jrn>34. Troosters T, Gosselink R, Decramer M. Six minute walking distance in healthy elderly subjects. Eur Respir J. 1999;14(2):270–4. https://doi.org/10.1034/j.1399-3003.1999.14b06.x. Medline:10515400</jrn>

<jrn>35. Janaudis-Ferreira T, Hill K, Goldstein RS, et al. Resistance arm training in patients with COPD: a randomized controlled trial. Chest. 2011;139(1):151–8. https://doi.org/10.1378/chest.10-1292. Medline:20724740</jrn>

<jrn>36. Probst VS, Hernandes NA, Teixeira DC, et al. Reference values for the incremental shuttle walking test. Respir Med. 2012;106(2):243–8. https://doi.org/10.1016/j.rmed.2011.07.023. Medline:21865021</jrn>

<jrn>37. Knox AJ, Morrison JF, Muers MF. Reproducibility of walking test results in chronic obstructive airways disease. Thorax. 1988;43(5):388–92. https://doi.org/10.1136/thx.43.5.388. Medline:3194867</jrn>

<jrn>38. Shin SH, Ro H, Lee OS, et al. Within-day reliability of shoulder range of motion measurement with a smartphone. Man Ther. 2012;17(4):298–304. https://doi.org/10.1016/j.math.2012.02.010. Medline:22421186</jrn>

<jrn>39. Wong AYL, Kawchuk G, Parent E, et al. Within- and between-day reliability of spinal stiffness measurements obtained using a computer controlled mechanical indenter in individuals with and without low back pain. Man Ther. 2013;18(5):395–402. https://doi.org/10.1016/j.math.2013.02.003. Medline:23465962</jrn>

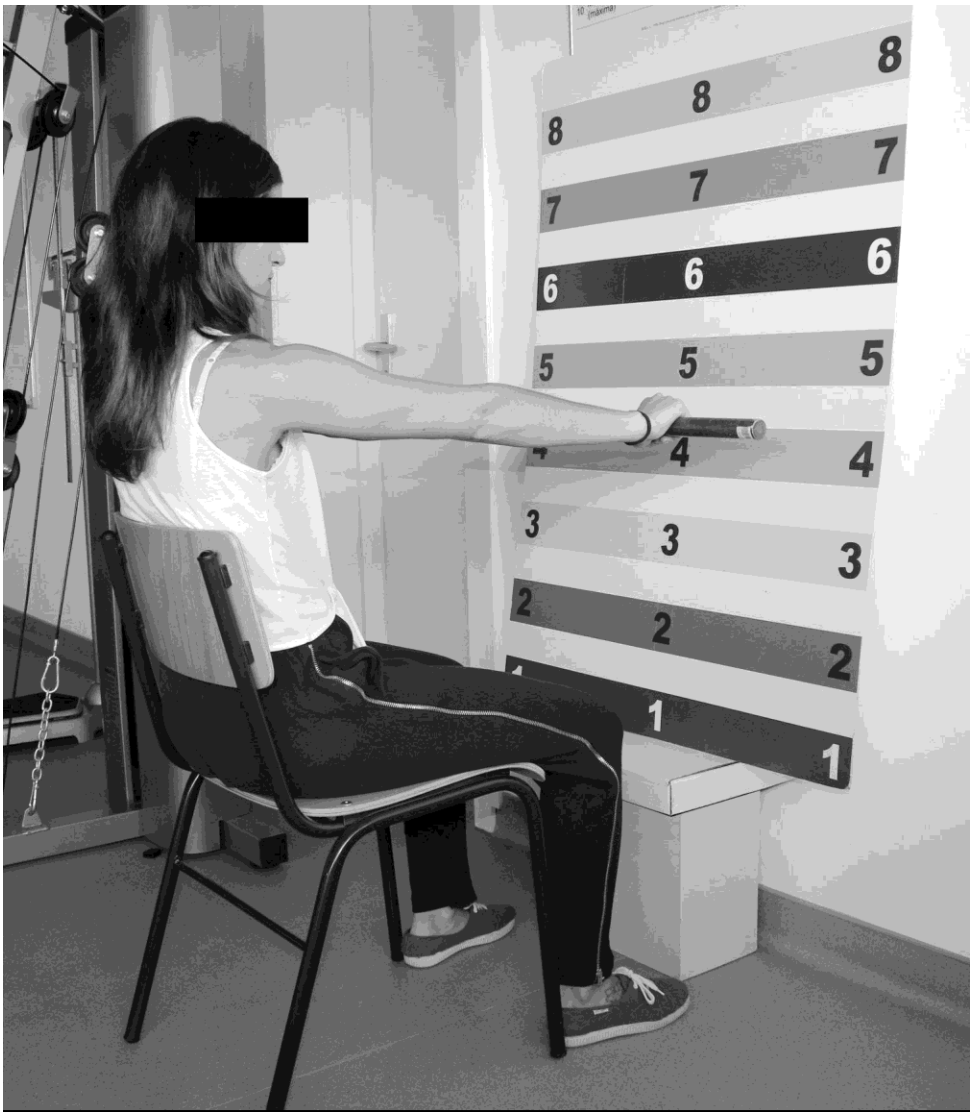
<jrn>40. Ferber R, McClay Davis I, Williams DS III, et al. A comparison of within- and between-day reliability of discrete 3D lower extremity variables in runners. J Orthop Res. 2002;20(6):1139–45. https://doi.org/10.1016/S0736-0266(02)00077-3. Medline:12472220</jrn>

Journal: PTC; Volume ; Issue:

DOI: 10.3138/PTC.2016-42

Figure captions

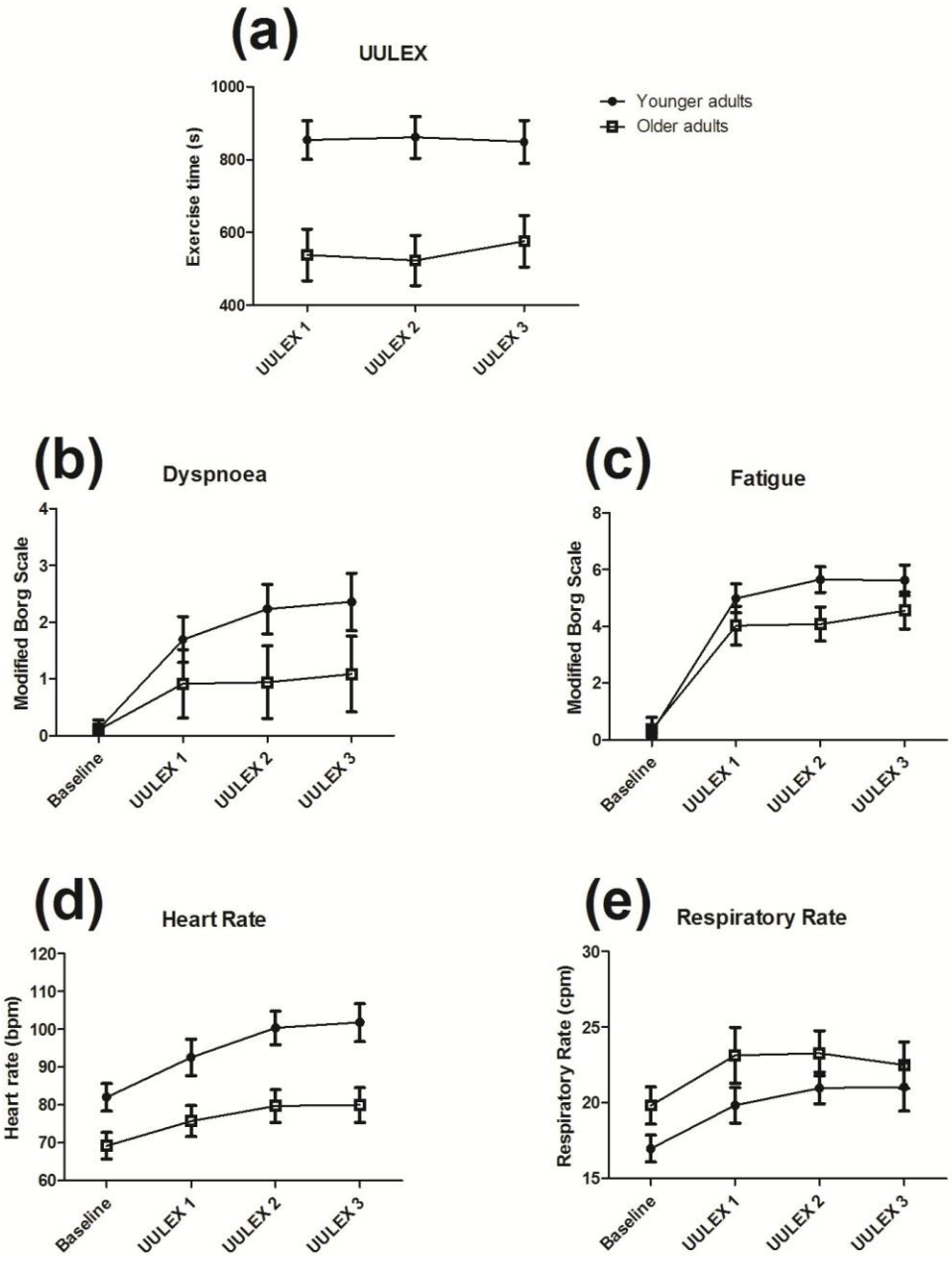
Figure 1. Set-up for the unsupported upper limb exercise test.



Journal: PTC; Volume ; Issue:

DOI: 10.3138/PTC.2016-42

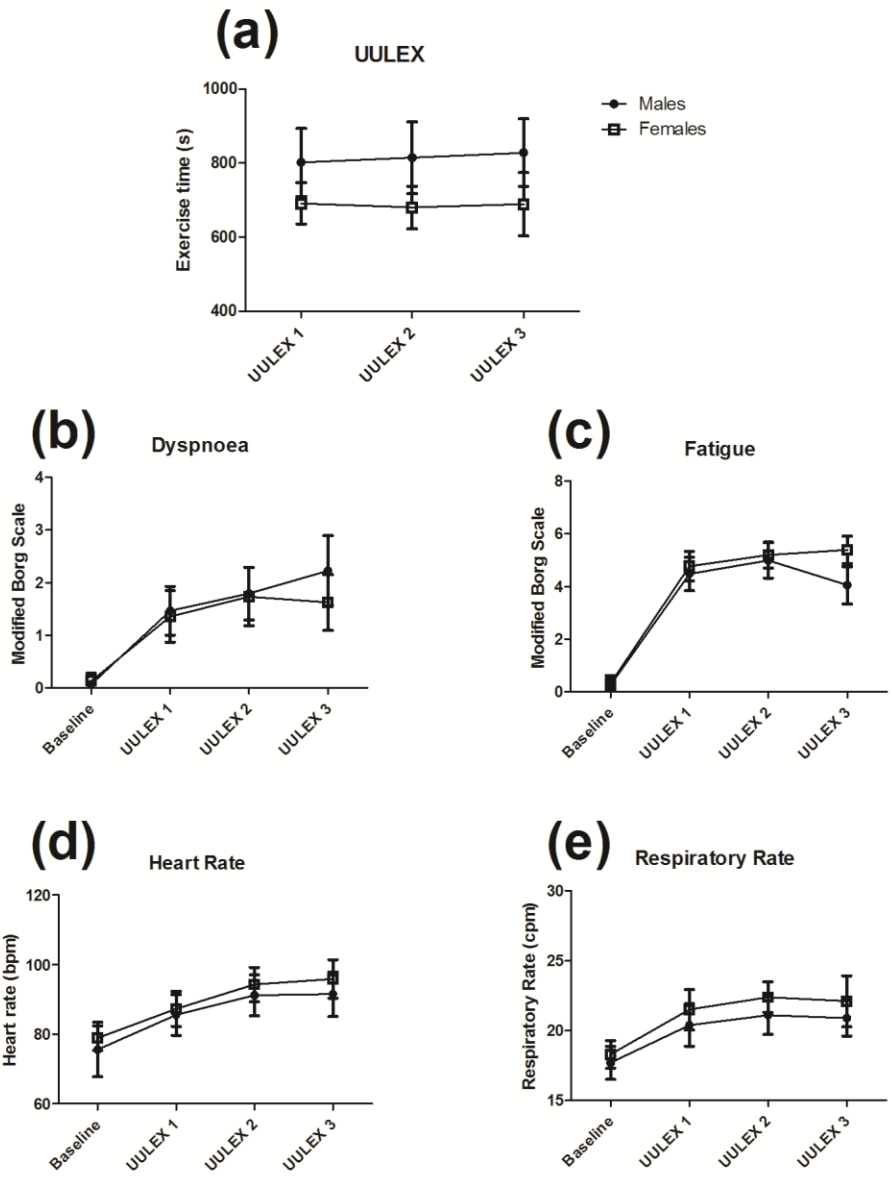
Figure 2. Comparisons between age groups at the different moments of evaluation for (a) total exercise time, (b) dyspnoea, (c) fatigue, (d) heart rate, and (e) respiratory rate.



Journal: PTC; Volume ; Issue:

DOI: 10.3138/PTC.2016-42

Figure 3. Comparisons between male and female participants at the different evaluation times for (a) total exercise time, (b) dyspnoea, (c) fatigue, (d) heart rate, and (e) respiratory rate.



Journal: PTC; Volume ; Issue:

DOI: 10.3138/PTC.2016-42

Table 1. Characteristics of Participants (*n* = 100)

Characteristic	Total sample (*n* = 100)	Younger adults (*n* = 64)	Older adults (*n* = 36)	*p*-value	Men (*n* = 44)	Women (*n* = 56)	*p*-value
Men, no. (%)	44.0 (44)	27.0 (42)	17.0 (47)	0.050	–	–	–
Age (y), mean (SD)	40.8 (25.0)	23.5 (8.2)	71.4 (11.5)	< 0.001*	44.2 (26.0)	38.1 (24.1)	0.88
BMI (kg/m^2), mean (SD)	24.8 (5.8)	22.8 (4.0)	28.9 (5.3)	< 0.001*	26.0 (4.8)	24.3 (5.7)	0.038*
Smoking status							
Never/former/current, %	84/1/15	84/14/2	83/0/17	0.715	33/1/10	51/0/5	0.076
No. years as a smoker, median (1st–3rd quartile)†	3.0 (0.6–5.0)	3.0 (0.2–4.0)	4.0 (0.5–40.8)	0.265	3.0 (0.5–5)	3.0 (1.5–5.5)	0.839
No. pack-years, median (1st–3rd quartile)†	0.1 (0.0–3.2)	0.2 (0.0–0.5)	2.7 (0.0–18.3)	0.044*	0.5 (0.2–5.0)	0.1 (0.0–0.6)	0.227

* $p < 0.05$.

†Data presented for smokers and former smokers only.

Journal: PTC; Volume ; Issue:

DOI: 10.3138/PTC.2016-42

Table 2. ICC and SEM Values of the UULEX between Tests for Total Sample, Age Groups, and Gender Groups

Parameter	Total sample (*n* = 100)	Younger adults (*n* = 64)	Older adults (*n* = 36)	Men (*n* = 44)	Women (*n* = 56)
Tests 1 and 2					
ICC(2,1) (95% CI)	0.93 (0.90, 0.95)	0.91 (0.85, 0.94)	0.87 (0.76, 0.93)	0.95 (0.92, 0.98)	0.88 (0.81, 0.93)
SEM (95% CI), s	26.36 (25.15, 30.63)	28.87 (25.35, 33.54)	37.96 (33.33, 44.10)	21.04 (18.47, 24.44)	35.86 (31.48, 41.66)
Tests 2 and 3					
ICC(2,1) (95% CI)	0.90 (0.85, 0.93)	0.89 (0.82, 0.93)	0.79 (0.61, 0.89)	0.93 (0.87, 0.96)	0.84 (0.77, 0.90)
SEM (95% CI), s	39.89 (30.03, 46.34)	41.54 (36.47, 48.26)	57.21 (50.23, 66.45)	35.64 (31.29, 41.40)	47.50 (41.71, 55.18)
Tests 1 and 3					
ICC(2,1) (95% CI)	0.88 (0.83, 0.92)	0.85 (0.76, 0.90)	0.81 (0.66, 0.90)	0.90 (0.82, 0.94)	0.84 (0.75, 0.90)
SEM (95% CI), s	41.69 (36.60, 48.43)	44.74 (39.28, 51.98)	54.74 (48.06, 63.59)	37.43 (32.86, 43.48)	49.16 (43.16, 57.11)

ICC(2,1) = intra-class correlation coefficient, model 2 (two-way random effects) with a single rater; SEM = standard error of measurement; UULEX = unsupported upper limb exercise.

Journal: PTC; Volume ; Issue:

DOI: 10.3138/PTC.2016-42

Table 3. Results of Total Exercise Time and Clinical Parameters for Each UULEX Test in Younger (*n* = 64) and Older (*n* = 36) Adults

Measure	Baseline	Test 1	Test 2	Test 3	*p*-value
Younger adults					
UULEX total time, mean (SD); 95% CI	–	853.5 (210.7); 800.9, 906.1	861.0 (228.3); 804.0, 918.0	848.5 (238.1); 789.1, 908.0	0.669
Dyspnoea (MBS), median (1st–3rd quartile)	0.0	1.0 (0.0–3.0)*	2.0 (0.5–3.0)*	2.0 (1.0–3.0)*	< 0.0001
Fatigue (MBS), median (1st–3rd quartile)	0.0	5.0 (3.0–6.8)*	6.0 (4.0–7.0)*†	5.0 (4.0–7.0)*†	< 0.0001
Heart rate (bpm), mean (SD); 95% CI	82.1 (14.5); 76.5, 85.7	92.6 (19.4);* 87.8, 97.4	100.4 (17.7);* 96.0, 104.8	101.8 (20.2);* (96.8, 106.8)	< 0.0001
Respiratory rate (cpm), mean (SD); 95% CI	17.0 (3.6); 16.1, 17.9	19.8 (4.7);* 18.7, 21.0	21.0 (4.2);* 19.9, 22.0	21.0 (6.4);* 19.5, 22.7	< 0.0001
Older adults					
UULEX total time, mean (SD); 95% CI	–	538.6 (208.8); 467.9, 609.2	523.4 (202.5); 454.9, 592.0	576.2 (208.9);† 505.5, 656.9	0.028
Dyspnoea (MBS), median (1st and 3rd quartile)	0.0	0.0 (0.0–2.0)	0.0 (0.0–1.0)	0.0 (0.0–2.0)	0.001
Fatigue (MBS), median (1st and 3rd quartile)	0.0	4.0 (3.0–5.0)*	4.0 (3.0–5.0)*	5.0 (3.0–6.0)*	< 0.0001
Heart rate (bpm), mean (SD); 95% CI	69.2 (10.3); 65.7, 72.7	75.8 (12.0)*; 71.7, 19.8	79.7 (12.8);* 75.4, 84.0	80.0 (13.6);* 75.4, 84.6	< 0.0001
Respiratory rate (cpm), mean (SD); 95% CI	19.8 (3.6); 18.6, 21.0	23.1 (5.5);* 21.3, 25.0	23.3 (4.4);* 21.8, 24.8	22.5 (4.5);* 21.0, 24.0	0.0003

*$p < 0.05$ from baseline.

†$p < 0.05$ from test 1.

UULEX = unsupported upper-limb exercise; MBS = modified Borg scale; bpm = beats per minute; cpm = cycles per minute.

Journal: PTC; Volume ; Issue:

DOI: 10.3138/PTC.2016-42

Table 4. Results of Total Exercise Time and Clinical Parameters for Each UULEX Test in Men (*n* = 44) and Women (*n* = 56)

Measure	Baseline	Test 1	Test 2	Test 3	*p*-value
Men					
UULEX total time, mean (SD); 95% CI	–	802.1 (302.1); 710.3, 894.0	814.5 (317.8); 717.9, 911.2	828.3 (300.1); 737.1, 919.5	0.336
Dyspnoea (MBS), median (1st–3rd quartile)	0.0	1.0 (0.0–3.0)*	2.0 (0.0-–3.0)*	2.0 (0.0–3.0)*	< 0.0001
Fatigue (MBS), median (1st–3rd quartile)	0.0	4.0 (3.0–5.0)*	5.0 (3.0–7.0)*	5.0 (4.0–7.0)*	< 0.0001
Heart rate (bpm), mean (SD); 95% CI	75.6 (25.8); 76.5, 85.7	85.5 (19.2);* 87.8, 97.4	91.2 (19.5);*† 96.0, 104.8	91.5 (21.1);*† 96.8, 106.8	< 0.0001
Respiratory rate (cpm), mean (SD); 95% CI	17.7 (3.9); 16.1, 17.9	20.4 (5.0);* 18.7, 21.0	21.1 (4.5);* 19.9, 22.0	20.9 (4.2);* 19.5, 22.7	< 0.0001
Women					
UULEX total time, mean (SD); 95% CI	–	691.4 (208.1); 635.7, 747.1	680.5 (215.6); 622.7, 738.2	689.4 (317.8); 632.7, 746.0	0.753
Dyspnoea (MBS), median (1st–3rd quartile)	0.0	0.5 (0.0–2.0)*	1.0 (0.0–3.0)*	1.0 (0.0–2.0)*	< 0.0001
Fatigue (MBS), median (1st–3rd quartile)	0.0	5.0 (3.0–6.0)*	5.0 (4.0–7.0)*	5.0 (4.0–7.0)*	< 0.0001
Heart rate (bpm), mean (SD); 95%	78.9 (13.3); 76.5, 85.7	87.3 (18.8);* 87.8, 97.4	94.3 (18.4);*† 96.0, 104.8	95.9 (20.6);*†	< 0.0001

Journal: PTC; Volume ; Issue:

DOI: 10.3138/PTC.2016-42

CI				96.8, 106.8	
Respiratory rate (cpm), mean (SD); 95% CI	18.3 (3.7); 18.6, 21.0	21.5 (5.4);* 21.3, 25.0	22.4 (4.1);* 21.8, 24.8	22.1 (6.8);* 21.0, 24.0	< 0.0001

*$p < 0.05$ from baseline.

†$p < 0.05$ from test 1.

UULEX = unsupported upper-limb exercise; MBS = modified Borg scale; bpm = beats per minute; cpm = cycles per minute.