

Reliability, Agreement and Minimal Detectable Change of the Timed Up & Go and the 10-Meter Walk tests in Older Patients with COPD

Alda Marques^{1,2} PhD, Joana Cruz^{1,2} MSc, Sara Quina³ PT, Maria Regêncio³ PT,

Cristina Jácome^{1,2} MSc

¹Lab 3R – Respiratory Research and Rehabilitation Laboratory, School of Health Sciences, University of Aveiro (ESSUA), Aveiro, Portugal

²Center for Health Technology and Services Research (CINTESIS), School of Health Sciences, University of Aveiro (ESSUA), Aveiro, Portugal

³School of Health Sciences, University of Aveiro (ESSUA), Aveiro, Portugal

Corresponding author:

Alda Marques

Lab 3R – Respiratory Research and Rehabilitation Laboratory, School of Health Sciences of the University of Aveiro (ESSUA); Campus Universitário de Santiago;

Agras do Crasto, Edifício 30, 3810-193 Aveiro, Portugal

Telephone number: 00351 234 372 462

Email address: amarques@ua.pt

Running head: Reliability and MDC of TUG and 10MWT in COPD

KEY WORDS: CHRONIC lung disease, functional balance, gait speed, measurement properties; walking.

Abstract

This study aimed to determine the interrater and intrarater reliability and agreement and the minimal detectable change (MDC) of the Timed Up & Go (TUG) test and the 10-Meter Walk Test (10MWT) in older patients with Chronic Obstructive Pulmonary Disease (COPD). Patients (≥ 60 years old) living in the community were asked to attend 2 sessions with 48-72h interval. In session 1, participants completed the TUG and 10MWT twice (2 trials) and were assessed by 2 raters. In session 2, they repeated the tests twice and were assessed by 1 rater. Interrater and intrarater reliability were calculated for the exact scores (using data from trial 1) and mean scores (mean of 2 trials) using Intraclass Correlation Coefficients (ICC_{2,1} and ICC_{2,2}, respectively).

Interrater and intrarater agreement were explored with the Bland & Altman method. The MDC₉₅ was calculated from the standard error of measurement.

Sixty participants (72.43 ± 6.90 years old) completed session 1 and 41 participants session 2. Excellent ICC values were found for the TUG test (interrater: ICC_{2,1}=0.997 ICC_{2,2}=0.999; intrarater: ICC_{2,1}=0.921 ICC_{2,2}=0.964) and 10MWT (interrater: ICC_{2,1}=0.992 ICC_{2,2}=0.997; intrarater: ICC_{2,1}=0.903 ICC_{2,2}=0.946). Good interrater and intrarater agreement was also found for both tests. The MDC₉₅ was 2.68s and 1.84s for the TUG and 0.40m/s and 0.30m/s for the 10MWT considering the exact and mean scores, respectively. Findings suggest that the TUG test and the 10MWT are reliable and have acceptable measurement error. Therefore, these measures may be used to assess functional balance (TUG) and gait (10MWT) deficits in older patients with COPD.

Introduction

Chronic Obstructive Pulmonary Disease (COPD) is one of the most prevalent chronic diseases among adults aged 60 and older [1] with systemic manifestations that can contribute to functional performance decline [2]. Impairments such as peripheral skeletal muscle dysfunction and reduced exercise capacity are well-documented systemic manifestations in patients with COPD [3]. Poor balance and slow gait speeds have also been described in these patients [4-9]. As a result, patients with COPD may experience difficulties in performing daily life activities that require functional mobility and balance control (e.g., moving around, dressing) [6] and be at higher risk of falling [4].

To manage and monitor balance and gait speed deficits in research and clinical settings, clinicians should routinely measure these parameters in patients with COPD. For this purpose, they need simple and reliable measures that can easily be applied in the clinical context and reflect true changes over time or in response to interventions. Several clinical measures have been developed to assess balance and gait deficits in older adults, such as the Timed Up & Go (TUG) test [10] and the 10-Meter Walk Test (10MWT) [11-13]. These are quick and easy-to-administer measures that do not require special equipment or space, being suitable to be applied in most clinical settings such as home care, (sub)acute rehabilitation facilities or long-term care facilities. The TUG test [10] is frequently used to assess functional balance, mobility skills and fall risk of older people [14, 15] and has been increasingly used in patients with COPD [4-7, 16, 17]. It incorporates a series of tasks necessary for independent living, such as walking, turning and sitting [18], thus providing a comprehensive representation of balance and functional mobility capacities of an individual. Another practical measure to assess functional mobility is the 10MWT [11-13]. This test assesses the short-duration gait

Reliability and MDC of TUG and 10MWT in COPD

speed [19, 20] and has been widely used for measuring walking ability of patients with neurologic impairment [13, 21]. Recent studies have demonstrated that gait speed is related to a number of clinical outcomes in COPD [8, 9]. However, studies employing the 10MWT in patients with COPD are limited [22]. As this is a simple and quick measure that can be used in different settings (e.g., home-based, community-based), where space and time are not easily available, the 10MWT may provide relevant information in routine assessments of patients with COPD.

Measurement properties of these clinical measures, such as interrater and intrarater reliability, have been studied in specific populations [23-27] but they have been scarcely examined in patients with COPD [22, 28]. Specifically, Mesquita et al. [28] assessed the within-day reliability of the TUG test in older patients with advanced grades of COPD, and Rozenberg et al. [22] assessed the within-day and between-days reliability and the construct validity of the usual and fast walking speeds using one version of the 10MWT in the same population. Therefore, further research is needed to assess the intrarater but also the interrater reliability in patients at all COPD grades. This is fundamental to determine the proportion of variance obtained in measurements conducted by different raters and/or under similar assessment conditions and to define to what extent the changes observed in a patient's balance and gait scores are significant or associated with measurement error [29]. Therefore, the purpose of this study was to determine the interrater and intrarater reliability and agreement and the minimal detectable change (MDC) of the TUG test and the 10MWT in older patients with COPD. Knowing the measurement properties and associated measurement error of these simple tests is crucial before their widespread use in the clinical management of these patients.

Methods

Study design

A [reliability and agreement](#) study was conducted. Ethical approval was obtained from the Central Regional Health Administration and the Hospital Center of Baixo Vouga. This study was described following the guidelines for reporting reliability and agreement studies (GRRAS) [29].

Participants

Patients with COPD were recruited from 2 primary care centers and 1 district hospital, from November 2012 to February 2013. Patients were identified by the physicians of the institutions, who ensured the fulfillment of eligibility criteria. Patients were included if they were: i) ≥ 60 years old; ii) diagnosed with COPD according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria [30]; iii) living in the community; iv) clinically stable (no hospital admissions or exacerbations) for 1 month prior to the study; v) able to ambulate with or without a walking aid (but without the assistance of another person); and vi) able to follow simple instructions. Patients were excluded if they had severe musculoskeletal (e.g., [severe osteoarthritis](#)), neurological (e.g., [Parkinson's disease](#)), cardiovascular (e.g., [stroke](#)) or psychiatric (e.g., [dementia conditions/disorders](#)) and/or severe auditory/visual impairments [that could interfere with the measurements](#).

Patients were informed by their physicians about the study and asked about their willingness to participate. Patients who agreed to participate were contacted by the researchers to schedule an appointment. The written informed consent was obtained before data collection.

Reliability and MDC of TUG and 10MWT in COPD

Data collection

Participants were asked to attend 2 testing sessions with a 48–72h interval.

In the first session, patients completed a questionnaire with sociodemographic (age, gender) and health-related (smoking status, incidence of ≥ 1 falls in the preceding year, use of walking aids) information. [Level of dyspnea](#) was collected using the modified Medical Research Council (mMRC) dyspnea scale [31]. Height and weight were recorded to calculate the body mass index (BMI). Lung function was assessed with a portable spirometer (MicroLab 3500, CareFusion, Kent, UK) according to the guidelines [32]. [This device uses the reference equations for lung function as defined in Quanjer et al. \[33\]](#). COPD severity was determined using the GOLD criteria [30].

Participants performed the TUG test and 10MWT twice in each testing session. Two previously trained clinicians with experience in applying these tests collected the data. Both raters evaluated the 2 trials of the TUG and 10MWT conducted in the first session to enable assessing interrater reliability. In the second session, only 1 of the raters evaluated 2 trials for further assessment of intrarater reliability. Prior to testing, participants received specific instructions on how to perform the tests and took a practice trial. Participants were encouraged to rest between trials, as needed. Effort was made to keep all factors associated with the testing sessions consistent, specifically the time of day, area in which the tests were performed, order of testing and use of a walking aid (if needed). No verbal encouragement was given during the tests. [A random number generator was used to randomly assign the sequence of data collection \(simple randomization\)](#).

Reliability and MDC of TUG and 10MWT in COPD

Measures

The TUG test [10] was performed in a corridor with a 3-meter mark on the floor (identified with adhesive tape) and a chair without armrests properly placed and fixed at the beginning mark. Instructions given to participants were as follow: when given the initiation word “Go”, they should rise from a chair, walk 3 meters as quickly and as safely as possible, turn around, return to the chair, and sit down with their back leaning against the backrest [34]. The fast TUG has been used in COPD [6, 7, 16, 17] and has been related to a higher risk of falls in community-dwelling older adults [15]. The time (seconds [s]) needed to perform the entire sequence was recorded using a stopwatch.

The 10MWT [11-13] was used to assess gait speed. Although this test is known as the 10MWT, gait speed is actually measured over 6 meters. The test was conducted in a corridor with 2 marks on the floor (identified with adhesive tape) denoting the 2 ends of a 10-meter walkway and another 2 marks identifying the 2 and 8 meters. Specific verbal instructions were given prior to the test: “*I will say: ready, set, go. When I say ‘Go’, walk as fast as you safely can until I say stop*”. Participants walked for 10 meters (m) at their fastest speed and the time for the intermediate 6 meters was measured to allow for acceleration and deceleration, starting when toes of the leading foot crossed the 2-meter mark and stopping when toes of the leading foot crossed the 8-meter mark. Maximum gait speed was used since a previous study has found that maximum speed declines more steeply than comfortable speed with increasing age and lower extremity muscle strength [11]. Gait speed (m/s) was calculated by dividing the distance walked on each trial (6m) by the time (s) required to complete the trial.

In both tests, participants who used a walking aid on a daily basis were asked to use it during the tests.

Reliability and MDC of TUG and 10MWT in COPD

Data analysis

Statistical analyzes were conducted using SPSS version 22 (IBM Corp., Armonk, NY). Descriptive statistics were used to characterize the sample and to provide the scores of each test. Sociodemographic and health-related characteristics were compared between patients who completed the 2 testing sessions and those who failed to attend the second session, using independent t-tests for normally distributed data, Mann-Whitney U-tests for non-normally distributed and ordinal data and Chi-square tests for categorical data. Level of significance was set at 0.05.

Relative reliability was assessed for the TUG test and 10MWT using the Intraclass Correlation Coefficient (ICC). Interrater reliability was computed using the scores obtained from the 2 raters in session 1 and intrarater reliability using the scores from 1 rater in sessions 1 and 2. The ICC_{2,1} was computed for the exact scores using the scores from trial 1 and ICC_{2,2} for the mean scores of 2 trials [35]. Mean scores were used in addition to the exact scores since they are considered better estimates of true scores and can provide better reliability estimates [36]. This methodology has been used in other reliability studies [26, 37]. The ICC ranges from 0 to 1 and ‘good’ to ‘excellent’ reliability can be considered for ICC>0.75 [35]. However, to consider a measurement suitable for clinical use, ICC values should exceed 0.9 [29]. The 95% confidence intervals (95%CI) were also computed.

The 95% limits of agreement (LoA) using the Bland & Altman plots [38] were performed to assess agreement between raters (interrater reliability) and sessions (intrarater reliability). The presence/absence of systematic error (e.g., due to learning effects or fatigue) between within-day trials was also assessed through the paired student t-test and the LoA, using the trials 1 and 2 of session 1 measured by one rater.

Reliability and MDC of TUG and 10MWT in COPD

To assess absolute reliability, the standard error of measurement (SEM) was calculated for the TUG and 10MWT using scores obtained by 1 rater in sessions 1 and 2. The SEM indicates the extent to which a score varies on repeated measurements [39] and was calculated using the equation 1:

$$SEM = SD \sqrt{(1 - ICC)} \quad (1)$$

where SD is the standard deviation of the scores obtained from all individuals and ICC is the intrarater reliability coefficient. SEM was also expressed as a percentage of the mean (SEM%), which is independent of the units of measurement, defined as (equation 2):

$$SEM\% = (SEM/mean) \times 100 \quad (2)$$

where mean is the mean of all scores obtained in sessions 1 and 2. The SEM (absolute value) was also used to determine the MDC, which is the smallest change in a patient's score which indicates a true change in patient's performance rather than the result of random measurement error. The MDC at the 95% level of confidence (MDC₉₅) was calculated as follows (equation 3):

$$MDC_{95} = SEM \times 1.96 \times \sqrt{2} \quad (3)$$

The MDC can also be expressed as a percentage (MDC%), defined as (equation 4):

$$MDC\% = (MDC_{95}/mean) \times 100 \quad (4)$$

where mean is the mean of all scores obtained in the testing sessions. A MDC% below 30% is considered acceptable [40].

Sample size

Sample size was determined as described by Gwet [41] for interrater reliability studies, using the results from a previous pilot study (not published). Based on 2 raters, it was

Reliability and MDC of TUG and 10MWT in COPD

determined that 60 patients would be necessary for an ICC of 0.9 and an expected 95%CI length of 0.1.

Results

Participants

Sixty-six patients were contacted and invited to participate in the study. However, 3 refused to participate as they did not perceive the study as relevant and 3 were unable to attend the health center. Therefore, 60 patients (35 males) with a mean(\pm SD) age of 72 ± 7 years participated in the study. They had a BMI of 29 ± 6 kg/m² and a FEV₁ of $65\pm 23\%$ of the predicted. Twenty five (42%) participants had fallen in the preceding year. None of them used a walking aid. Participants' characteristics are presented in Table 1.

(table 1)

From the 60 participants, 19 were lost to follow up due to unavailability to attend the health center (n=16) and respiratory exacerbations (n=3). Thus, 41 participants attended session 2. There were no significant differences between baseline characteristics of patients who completed the 2 testing sessions and the dropouts ($p>0.05$).

Interrater reliability and agreement

Scores of the TUG test and the 10MWT measured by 2 raters in session 1 and interrater reliability results are presented in Table 2. Results are shown for the exact scores and the mean scores. The ICCs for the interrater reliability were excellent on both tests, with the lower limit of the 95%CI being above 0.9 in both measures, either considering the exact scores (TUG: ICC_{2,1}=0.997 95%CI 0.995–0.998; 10MWT: ICC_{2,1}=0.992 95%CI

Reliability and MDC of TUG and 10MWT in COPD

0.987–0.995) or the mean scores (TUG: $ICC_{2,2}=0.999$ 95%CI 0.999–1; 10MWT: $ICC_{2,2}=0.997$ 95%CI 0.995–0.998).

(table 2)

The Bland & Altman plots for the measures collected by the 2 raters are shown in Figure 1. Excellent agreement (i.e., the mean of the differences close to 0 and tight LoA) was found between raters for the TUG test and the 10MWT. No evidence of systematic bias was observed.

(figure 1)

Intrarater reliability, agreement and minimal detectable change (MDC_{95})

Table 3 presents the scores of the TUG and 10MWT measured by the same rater in sessions 1 and 2 and intrarater reliability results. Intrarater reliability of TUG was excellent, with ICCs achieving the recommended value of 0.9. This was found for the exact scores ($ICC_{2,1}=0.921$ 95%CI 0.855–0.957) and mean scores ($ICC_{2,2}=0.964$ 95%CI 0.931–0.981). Similar results were obtained for the 10MWT (exact scores: $ICC_{2,1}=0.903$ 95%CI 0.826–0.947, $p<0.001$; mean scores: $ICC_{2,2}=0.946$ 95%CI 0.898–0.971, $p<0.001$).

Good intrarater agreement results were found for both tests (Figure 2), with no evidence of systematic bias.

(figure 2)

No significant differences were found between trials of session 1 for the TUG test ($8.94\pm 3.23s$ vs. $8.79\pm 3.21s$, $p=0.108$) and the 10MWT ($1.41\pm 0.46m/s$ vs. $1.43\pm 0.44m/s$, $p=0.107$). The Bland & Altman plots showed no evidence of systematic error between within-day trials.

(figure 3)

Reliability and MDC of TUG and 10MWT in COPD

The SEM of the TUG test was 0.97s (SEM%=12.1) using exact scores and 0.66s (SEM%=8.4) using the mean scores, indicating a small variation in the measurement error. This test had a MDC₉₅ of 2.68s (MDC%=33.5) and 1.84s (MDC%=23.2) for the exact and mean scores, respectively.

Regarding the 10MWT, the SEM was 0.14m/s (SEM%=9.6) for the scores obtained in trial 1 and 0.11m/s (SEM%=7.1) for the mean scores of trials 1 and 2. The 10MWT also showed an acceptable measurement error, with a MDC₉₅ of 0.40m/s (MDC%=26.5) for the exact scores and 0.30m/s (MDC%=19.6) for the mean scores.

(table 3)

Discussion

This study evaluated the interrater and intrarater reliability and agreement of the TUG test and the 10MWT in older patients with COPD. These clinical measures showed excellent interrater and intrarater reliability, with ICC values exceeding the recommended 0.9 [29], and good agreement results. Estimates of measurement error (SEM, SEM%) and the minimal amount of change necessary to assume a meaningful change in patient's performance (MDC₉₅, MDC%) were relatively low for both measures. Findings suggest that the TUG test and the 10MWT are reliable measures to assess functional balance and gait deficits (respectively) in older patients with COPD. Excellent results were found for interrater and intrarater reliability of the TUG test (ICC>0.9). Similar results have been found in reliability studies conducted in other populations, such as Parkinson's disease (interrater ICC=0.99;[42] intrarater ICC 0.80-0.98) [40, 42], Alzheimer's Disease (intrarater ICC=0.99) [26] and stroke (intrarater ICC=0.96) [43]. Regarding agreement and absolute reliability (SEM and MDC₉₅) of the TUG, results were also encouraging since a small variation of measurement error was

Reliability and MDC of TUG and 10MWT in COPD

found. A previous study conducted by Mesquita et al. [28] has also assessed the intrarater reliability and agreement parameters of the TUG test in patients with COPD. The authors found lower reliability and agreement results than those of the present study (first and second trials: ICC 0.85 95%CI 0.76–0.91; SEM=1.76; MDC₉₅= 4.88, MDC%=43; for the Bland & Altman plots the reader is directed to the original article) [28]. However, comparison between the previous and the present studies should be done with caution since patients' disease severity and the methodologies employed were different: disease severity (severe-to-very-severe COPD grades vs. all COPD grades, respectively), TUG pace (usual speed vs. fast speed), existence of a practice trial (no reference to a practice trial vs. existence of a practice trial), and time period between trials (within-day vs. between-day trials). As previously mentioned, the MDC₉₅ indicates the smallest real change of an outcome measure for a 95% confidence level. It should be noted that the MDC% of the TUG exact scores was slightly higher (33.5%) than the value considered acceptable (i.e., <30%) [40]. When the mean scores from the 2 trials were considered, the MDC% decreased to 23.2% though. These findings suggest that, when using TUG scores in COPD, clinicians should calculate the average of 2 measurements in order to reduce the magnitude of measurement error and thus be more confident that a real statistical change has occurred [36]. Nevertheless, further research is warranted before firm conclusions can be drawn.

The MDC values of the TUG test obtained in the present study may be used by clinicians to identify a true statistical change (beyond measurement error) in a patient's TUG score over time or in response to interventions. Previous studies using the TUG test to evaluate the effects of an intervention in patients with COPD have found pre-posttest mean differences of -1.5s [44] and -1.7s [17], which are lower than the MDC values found in the present study (2.68s and 1.84s for the exact and mean scores,

Reliability and MDC of TUG and 10MWT in COPD

respectively). This suggests that, on average, the interventions did not produce a true statistical change in patients' TUG scores. Nevertheless, it is not known the proportion of patients achieving a degree of improvement that was beyond the MDC, which is a more informative method for describing the effects of the intervention than the overall mean change [45]. Further research is needed to clarify this issue.

No evidence of systematic error was found for the TUG test between trials 1 and 2 during session 1. Mesquita et al. [28] have found that trial 2 of the TUG test was significantly better than trial 1 (possibly due to learning effects) but there were no differences between trials 2 and 3. In the present study, a practice trial was allowed before trials 1 and 2, which may have corresponded to trial 1 of the study from Mesquita et al. [28], thereby supporting their results.

The 10MWT also showed excellent (relative and absolute) reliability and good agreement in patients with COPD. It has been recognized that gait speed is an important functional measure which has been linked to fall risk [46], fear of falling [47] and survival [48] in older adults. It also predicts future functional decline [49, 50]. Hence, gait speed has been considered a “*functional vital sign*” [51]. Gait speed has not been extensively studied in COPD; nevertheless, recent studies have suggested that it can also be an indicator of functional decline in this population since it has been associated to patients' symptoms (e.g., dyspnea, fatigue), exercise tolerance and quality of life [8, 9]. As the 10MWT was found to be reliable in patients with COPD and may be applied in different settings, future studies should explore the value of gait speed measured by the 10MWT in assessing functional capacity of patients with COPD and its relationship with other clinical measures.

In the present study, gait speed was calculated by considering the time needed to walk the intermediate 6 meters of the 10-meter course, to account for acceleration and

Reliability and MDC of TUG and 10MWT in COPD

deceleration. However, a great variation exists in the literature concerning the distance [52, 53], testing protocol (e.g., presence/absence of specific distances for acceleration/deceleration) and pace (self-selected comfortable pace vs. fast pace) used [22, 52, 53]. Often, these testing parameters are chosen based more on tester preference and convenience, especially in some clinical settings where space is limited [54]. Thus, clinicians and researchers should clearly indicate which measurement method they have used in order to enable comparison among studies. Furthermore, the ICC, SEM and MDC_{95} of the present study should only be considered for studies using the same methodology and population. Unlike the TUG test, the $MDC\%$ of the the exact scores of the 10MWT was fairly below the value considered acceptable ($MDC\%=26.5$). This suggests that a single measure may be sufficient to assess changes in patients' gait speed without substantially increasing the measurement error. The non-significant results found for the differences between within-day trials further support these results, by suggesting that there was no evidence of systematic error. Further research is needed to support these findings.

Limitations

There are limitations in the present study that need to be acknowledged. Patients were 60 years old or over and most of them had mild ($n=20$) and moderate ($n=25$) COPD, which may limit the generalizability of the findings to the all COPD population. Future studies should assess younger patients and use a more balanced sample of COPD grades. The number of participants lost to follow up was another limitation of the present study. Nevertheless, the characteristics of dropouts were similar to that of completers, thereby they might not have considerable influence in the reported results. Physiological measures, such as heart rate and peripheral oxygen saturation, were not

Reliability and MDC of TUG and 10MWT in COPD

measured before and after each assessment, thus, it is currently unknown if the resting period was sufficient for patients to recover between trials. Another limitation concerns the use of the ICC. It is acknowledged that this reliability coefficient depends on the characteristics of the individuals included in the analysis: a large ICC can be found when the between-subjects variability is high, even if there is poor between-raters/trial-to-trial variability; a low ICC can be found if the between-subjects variability is low, even when the between-raters/trial-to-trial variability is low [39]. Therefore, studies conducted in populations with different characteristics are likely to report ICC values different from the ones found in this study. For example, the lower ICC values found for the TUG test in the study of Mesquita et al. [28] may be attributable to the exclusive inclusion of patients at advanced grades of the disease (i.e., patients with similar characteristics). In the present study, all COPD grades were considered which may have lead to a great between-subjects variability and, thus, higher ICC results. Finally, measures used to calculate intrarater reliability were collected with a 48–72h interval. This shorter reassessment interval was chosen since, in clinical practice (e.g., physical therapy), the period between visits is often smaller than 1 week. However, it is recognized that random measurement error is time-dependent and, thus, longer periods could lead to lower reliability values and higher measurement error [55]. Future studies could assess intrarater reliability of the TUG test and the 10MWT in various time intervals in order to study the effect of time (i.e., shorter vs. longer intervals) on measurement error.

Conclusions

This study showed that the TUG test and the 10MWT are highly reliable and have acceptable measurement error, thus they **may** be used to assess functional balance and

Reliability and MDC of TUG and 10MWT in COPD

gait deficits (respectively) in older patients with COPD. Results also indicated that small improvements in test scores may be sufficient to detect true changes in a patient's performance. These findings may help clinicians to interpret when a change corresponds to a true change (i.e., beyond measurement error) in a patient's performance.

Acknowledgements

The authors would like to acknowledge all institutions and patients involved for their participation in this research. A special thanks to Catarina Furtado and Susana Martinho for their contribution in data collection.

Declaration of Interest statement

The authors have no conflict of interest.

References

1. Gershon AS, Warner L, Cascagnette P, Victor JC, To T. Lifetime risk of developing chronic obstructive pulmonary disease: a longitudinal population study. *Lancet*. 2011;378(9795):991-6.
2. Barnes PJ, Celli BR. Systemic manifestations and comorbidities of COPD. *Eur Respir J*. 2009;33(5):1165-85.
3. Decramer M, De Benedetto F, Del Ponte A, Marinari S. Systemic effects of COPD. *Respir Med*. 2005;99(Suppl 2):S3-S10.
4. Beauchamp MK, Hill K, Goldstein RS, Janaudis-Ferreira T, Brooks D. Impairments in balance discriminate fallers from non-fallers in COPD. *Respir Med*. 2009;103(12):1885-91.
5. Butcher SJ, Meshke JM, Sheppard MS. Reductions in functional balance, coordination, and mobility measures among patients with stable chronic obstructive pulmonary disease. *J Cardiopulm Rehabil Prev*. 2004;24(4):274-80.
6. Cruz J, Marques A, Jácome C, Gabriel R, Figueiredo D. Global Functioning of COPD Patients With and Without Functional Balance Impairment: An Exploratory Analysis Based on the ICF Framework. *Journal of COPD*. 2015;12(2):207-16.
7. Jácome C, Cruz J, Gabriel R, Figueiredo D, Marques A. Functional balance in older adults with chronic obstructive pulmonary disease. *J Aging Phys Act*. 2014;22(3):357-63.
8. Ilgin D, Ozalevli S, Kilinc O, Sevinc C, Cimrin AH, Ucan ES. Gait speed as a functional capacity indicator in patients with chronic obstructive pulmonary disease. *Ann Thorac Med*. 2011;6(3):141-6.
9. Karpman C, DePew ZS, LeBrasseur NK, Novotny PJ, Benzo RP. Determinants of gait speed in COPD. *Chest*. 2014;146(1):104-10.

Reliability and MDC of TUG and 10MWT in COPD

10. Podsiadlo D, Richardson S. The timed 'Up and Go' Test: a Test of Basic Functional Mobility for Frail Elderly Persons. *J Am Geriatr Soc.* 1991;39(2):142-8.
11. Bohannon RW. Comfortable and maximum walking speed of adults aged 20-79 years: reference values and determinants. *Age Ageing.* 1997;26(1):15-9.
12. Bohannon RW, Andrews AW, Thomas MW. Walking Speed: Reference Values and Correlates for Older Adults. *J Orthop Sports Phys Ther.* 1996;24(2):86-90.
13. Wolf SL, Catlin PA, Gage K, Gurucharri K, Robertson R, Stephen K. Establishing the reliability and validity of measurements of walking time using the Emory Functional Ambulation Profile. *Phys Ther.* 1999;79(12):1122-33.
14. Davis DHJ, Rockwood MRH, Mitnitski AB, Rockwood K. Impairments in mobility and balance in relation to frailty. *Arch Gerontol Geriatr.* 2011;53(1):79-83.
15. Shumway-Cook A, Brauer S, Woollacott M. Predicting the probability for falls in community-dwelling older adults using the Timed Up & Go Test. *Phys Ther.* 2000;80(9):896-903.
16. Butcher SJ, Pikaluk BJ, Chura RL, Walkner MJ, Farthing JP, Marciniuk DD. Associations between isokinetic muscle strength, high-level functional performance, and physiological parameters in patients with chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis.* 2012;7:537-42.
17. Marques A, Jácome C, Cruz J, Gabriel R, Figueiredo D. Effects of a Pulmonary Rehabilitation Program With Balance Training on Patients With COPD. *J Cardiopulm Rehabil Prev.* 2015;35(2):154-8.
18. Mancini M, Horak FB. The relevance of clinical balance assessment tools to differentiate balance deficits. *Eur J Phys Rehabil Med.* 2010;46(2):239-48.
19. Kim S, Yuk GC, Gak H. Effects of the horse riding simulator and ball exercises on balance of the elderly. *Phys Ther.* 2013;25(11):1425-8.

20. Picelli A, Melotti C, Origano F, Neri R, Waldner A, Smania N. Robot-assisted gait training versus equal intensity treadmill training in patients with mild to moderate Parkinson's disease: A randomized controlled trial. *Parkinsonism Relat Disord*. 2013;19(6):605-10.
21. Wirz M, Zemon DH, Rupp R, Scheel A, Colombo G, Dietz V, Hornby TG. Effectiveness of automated locomotor training in patients with chronic incomplete spinal cord injury: a multicenter trial. *Arch Phys Med Rehabil*. 2005;86(4):672-80.
22. Rozenberg D, Dolmage TE, Evans RA, Goldstein RS. Repeatability of Usual and Fast Walking Speeds in Patients With Chronic Obstructive Pulmonary Disease. *J Cardiopulm Rehabil Prev*. 2014;34(5):348-54.
23. Kennedy DM, Stratford PW, Wessel J, Gollish JD, Penney D. Assessing stability and change of four performance measures: a longitudinal study evaluating outcome following total hip and knee arthroplasty. *BMC Musculoskelet Disord*. 2005;6:3.
24. Morris S, Morris ME, Iansek R. Reliability of measurements obtained with the Timed "Up & Go" test in people with Parkinson disease. *Phys Ther*. 2001;81(2):810-8.
25. Ng SS, Hui-Chan CW. The Timed Up & Go Test: Its Reliability and Association With Lower-Limb Impairments and Locomotor Capacities in People With Chronic Stroke. *Arch Phys Med Rehabil*. 2005;86(8):1641-7.
26. Ries J, Echternach J, Nof L, Blodgett M. Test-Retest Reliability and Minimal Detectable Change Scores for the Timed "Up & Go" Test, the Six-Minute Walk Test, and Gait Speed in People With Alzheimer Disease. *Phys Ther*. 2009;89(6):569-79.

27. van Hedel HJ, Wirz M, Dietz V. Assessing walking ability in subjects with spinal cord injury: Validity and reliability of 3 walking tests. *Arch Phys Med Rehabil.* 2005;86(2):190-6.
28. Mesquita R, Janssen DJA, Wouters EFM, Schols JMGA, Pitta F, Spruit MA. 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.
29. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, Roberts C, Shoukri M, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol.* 2011;64(1):96-106.
30. Global Strategy for the Diagnosis Management and Prevention of COPD. Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2015. Available from: <http://www.goldcopd.org/>.
31. Doherty DE, Belfer MH, Brunton SA, Fromer L, Morris CM, Snader TC. Chronic Obstructive Pulmonary Disease: Consensus Recommendations for Early Diagnosis and Treatment. *J Fam Pract.* 2006;Suppl:S1-S8.
32. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, Crapo R, Enright P, Grinten CPMVD, Gustafsson P, Jensen R, Johnson DC, MacIntyre N, McKay R, Navajas D, Pedersen OF, Pellegrino R, Viegi G, Wanger J. Standardisation of spirometry. *Eur Respir J.* 2005;26(2):319-38.
33. Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault J-C. Lung volumes and forced ventilatory flows. *Eur Respir J.* 1993;6(Suppl 16):5-40.
34. Mathias S, Nayak US, Isaacs B. Balance in elderly patients: the "get-up and go" test. *Arch Phys Med Rehabil.* 1986;67(6):387-9.

35. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull.* 1979;86(2):420-8.
36. Portney L, Watkins M. *Foundations of Clinical Research: Applications to Practice.* 2nd ed. Upper Saddle River, NJ: Prentice Hall Health; 2000.
37. Steffen T, Seney M. Test-Retest Reliability and Minimal Detectable Change on Balance and Ambulation Tests, the 36-Item Short-Form Health Survey, and the Unified Parkinson Disease Rating Scale in People With Parkinsonism. *Phys Ther.* 2008;88(6):733-46.
38. Bland MJ, Altman DG. Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement. *The Lancet.* 1986;327(8476):307-10.
39. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res.* 2005;19(1):231-40.
40. Huang S-LH, Ching-Lin; Wu, Ruey-Meei; Tai, Chun-Hwei; Lin, Chin-Hsien; Lu, Wen-Shian. Minimal Detectable Change of the Timed "Up & Go" Test and the Dynamic Gait Index in People With Parkinson Disease. *Phys Ther.* 2011;91(1):114-21.
41. Gwet KL. *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring The Extent of Agreement Among Raters.* 4th ed. Gaithersburg, MD (USA): Advanced Analytics, LLC; 2014.
42. Bennie S, Bruner K, Dizon A, Fritz H, Goodman B, Peterson S. Measurements of Balance: Comparison of the Timed "Up and Go" Test and Functional Reach Test with the Berg Balance Scale. *J Phys Ther Sci.* 2003;15(2):93-7.
43. Flansbjer UB, Holmback AM, Downham D, Patten C, Lexell J. Reliability of gait performance tests in men and women with hemiparesis after stroke. *J Rehabil Med.* 2005;37(2):75-82.

Reliability and MDC of TUG and 10MWT in COPD

44. Beauchamp MK, Brooks D, Stein RSG. Deficits in postural control in individuals with COPD - emerging evidence for an important secondary impairment. *Multidisciplinary Respiratory Medicine*. 2010;5(6):417-21.
45. Haley SM, Fragala-Pinkham MA. Interpreting Change Scores of Tests and Measures Used in Physical Therapy. *Phys Ther*. 2006;86(5):735-43.
46. Hausdorff JM, Rios DA, Edelberg HK. Gait variability and fall risk in community-living older adults: A 1-year prospective study. *Arch Phys Med Rehabil*. 2001;82(8):1050-6.
47. Maki BE. Gait changes in older adults: predictors of falls or indicators of fear. *J Am Geriatr Soc*. 1997;45(3):313-20.
48. Studenski S, Perera S, Patel K, Rosano C, Faulkner K, Inzitari M, Brach J, Chandler J, Cawthon P, Connor EB, Nevitt M, Visser M, Kritchevsky S, Badinelli S, Harris T, Newman AB, Cauley J, Ferrucci L, Guralnik J. Gait speed and survival in older adults. *JAMA*. 2011;305(1):50-8.
49. Harada N, Chiu V, Damron-Rodriguez J, Fowler E, Siu A, Reuben DB. Screening for balance and mobility impairment in elderly individuals living in residential care facilities. *Phys Ther*. 1995;75(6):462-9.
50. van Iersel MB, Munneke M, Esselink RA, Benraad CE, Olde Rikkert MG. Gait velocity and the Timed-Up-and-Go test were sensitive to changes in mobility in frail elderly patients. *J Clin Epidemiol*. 2008;61(2):186-91.
51. Fritz S, Lusardi M. White Paper: "Walking Speed: the Sixth Vital Sign". *J Geriatr Phys Ther*. 2009;32(2):2-5.
52. Graham JE, Ostir GV, Fisher SR, Ottenbacher KJ. Assessing walking speed in clinical research: a systematic review. *J Eval Clin Pract*. 2008;14(4):552-62.

Reliability and MDC of TUG and 10MWT in COPD

53. Watson MJ. Refining the Ten-metre Walking Test for Use with Neurologically Impaired People. *Physiotherapy*. 2002;88(7):386-97.
54. Peters DM, Fritz SL, Krotish DE. Assessing the reliability and validity of a shorter walk test compared with the 10-Meter Walk Test for measurements of gait speed in healthy, older adults. *J Geriatr Phys Ther*. 2013;36(1):24-30.
55. Gajdosik RL, Bohannon RW. Clinical measurement of range of motion. Review of goniometry emphasizing reliability and validity. *Phys Ther*. 1987;67(12):1867-72.