BMJ Open

Measurement properties of the incremental step test for people with chronic obstructive pulmonary disease: a cross-sectional study

Tânia Gonçalves,[1] João Carlos Winck,[2,3] Fátima Silva,[4] Cátia Caneiras ,[5,6,7,8] António Mesquita Montes,[1,9] Rui Vilarinho [1,4]

To cite: Gonçalves T, Carlos Winck J, Silva F, *et al*. Measurement properties of the incremental step test for people with chronic obstructive pulmonary disease: a cross-sectional study. *BMJ Open* 2024;**14**:e078425. doi:10.1136/bmjopen-2023-078425

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2023-078425).

Received 01 August 2023
Accepted 22 January 2024

For numbered affiliations see end of article.

Correspondence to
Dr Rui Vilarinho;
ruivilarinho1@gmail.com

ABSTRACT

Objectives The new incremental step test (IST) is a field test that was developed for people with chronic obstructive pulmonary disease (COPD), based on the characteristics of the incremental shuttle walk test (ISWT); however, its measurement properties still need to be determined. We aimed, first, to assess the construct validity (through the comparison with the ISWT), within-day reliability and measurement error of the IST in people with COPD; and, second, to identify whether the participants have a learning effect in the IST.

Design Cross-sectional study, conducted according to COnsensus-based Standards for the selection of health status Measurement INstruments guidelines.

Setting A family health unit in Portugal, April 2022 to June 2023.

Participants and analysis 63 participants (67.5±10.5 years) attended two sessions to perform two IST and two ISWT, separately. Spearman's correlations were used to compare the best performances between the IST and the ISWT. Intraclass correlation coefficient ($ICC_{2,1}$) was used for reliability, and the SE of measurement (SEM), minimal detectable change at 95% CI (MDC95) and Bland and Altman 95% limits of agreement (LoA) were used for measurement error. The learning effect was explored with the Wilcoxon signed-rank test.

Results The IST was significant and strongly correlated with the ISWT ($0.72<\rho<0.74$, $p<0.001$), presented an $ICC_{2,1}$ of 0.95 (95% CI 0.92 to 0.97), SEM=11.7 (18.9%), MDC95=32.4 (52.2%) and the LoA were −33.61 to 31.48 for the number of steps. No difference was observed between the number of steps of the two attempts of the IST ($p>0.05$).

Conclusions The IST can be suggested as a valid and reliable test to assess exercise capacity in people with COPD, with no learning effect when two IST are performed on the same day. The measurement error of the IST is considered indeterminate.

Trial registration number NCT04715659.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ The assessment of the measurement properties of the incremental step test (IST) was conducted according to COnsensus-based Standards for the selection of health status Measurement INstruments guidelines, which provide general principles that should be considered in the design of all studies on measurement properties.

⇒ The IST was developed specifically for people with chronic obstructive pulmonary disease based on the characteristics of another field test also developed for this population (incremental shuttle walk test).

⇒ The IST has not yet been compared with the gold standard test to assess the exercise capacity (cardiopulmonary exercise testing), in order to determine its criterion validity.

INTRODUCTION

Exercise capacity is an important clinical measure to determine the functional status of people with chronic obstructive pulmonary disease (COPD).[1] It is also an important outcome domain reported in evidence for the development of core outcome sets in pulmonary rehabilitation (PR) programmes.[2]

The field tests are used to assess the exercise capacity to overcome the limitations of cardiopulmonary exercise testing (CPET) (high cost, training of personnel and medical contraindications).[3 4] Step tests are an interesting option of field test since they require minimal physical space and equipment, are portable, and, most importantly, they reflect one of the main activities of daily living, such as climbing and descending stairs.[5] These advantages allow their application in the new settings of PR according to the current perspectives to increase the implementation and access of PR programmes, through community-based and home-based PR programmes.[6] This strategy addresses the current inadequate number of PR programmes worldwide, in relation to the large needs of people with COPD.[7 8]

One of the options of step tests is the incremental step test (IST) which was developed specifically for people with COPD.[9] This test

was adapted based on the characteristics of the incremental shuttle walk test (ISWT): incremental profile, externally paced profile, same duration and same number of levels. This adaptation arose due to a systematic review with the aim to identify the step tests used in people with COPD, where most of them were adapted from the 6 min walk test (6MWT), but no identified step test was adapted from the ISWT.[5]

The development of a new field test requires the study of its measurement properties, namely its validity, reliability and measurement error, before its full implementation in clinical practice, to assure that its selection is evidence based.[10 11] The first results for this purpose were promising, showing correlation values of 0.50 and 0.46 for construct validity (compared with 6MWT and 1 min sit-to-stand test, respectively). Additionally, its application proved to be feasible in the home environment because these results were obtained from data collection performed at participants' homes, with no adverse events reported.[9] Despite these promising results on validity, analysis of construct validity by comparing the performance of the IST with the performance of the test on which it was based (ISWT) has not yet been studied.

Another measurement property previously assessed for the IST was the reliability, where an intraclass correlation coefficient (ICC) of 0.96 (95% CI 0.92 to 0.98) for between-days reliability (7 days apart) was found.[9] The presence of a learning effect between two IST performed on different days was also demonstrated in people with COPD, which was expected since this population presents a learning effect in most of the field tests (eg, walking tests).[12] However, unacceptable values were found for measurement error (between days), especially in the minimal detectable change at 95% CI (MDC95) and its percentage (27.9 steps and 45.8%, respectively). These results could be different if attempts are performed on the same day, but this data is not yet available.

Therefore, this study comprises two aims: (1) to assess the construct validity of the IST through the comparison with the ISWT, its within-day test–retest reliability and measurement error and (2) to identify whether the participants have a learning effect when two IST are performed on the same day. We hypothesise that the IST performance will show strong correlations with the ISWT performance, excellent within-day reliability and acceptable values for measurement error. We predict the presence of a learning effect between two IST performed on the same day.

METHODS

Study design and sample size

A cross-sectional study was conducted between April 2022 and June 2023. Reporting follows Strengthening The Reporting of Observational Studies in Epidemiology guidelines.[13]

The study was registered on ClinicalTrials.gov, NCT04715659.

The methodology and sample size to determine the validity, reliability and measurement error of the IST followed the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) guidelines.[11 14] Validity is a measure of how well a test measures what it sets out to measure, that is, if it relates to the gold standard measure (criterion validity) or other measures that assess the same construct (hypotheses testing for construct validity).[15] Considering the hypotheses testing for construct validity, a comparison with another outcome measurement instrument was assessed (convergent validity) in this study, by analysing the correlation between the number of steps and maximum cadence reached in the IST and distance, and maximum speed achieved in the ISWT. Reliability refers to the consistency of a measure and its ability to replicate the performance from one assessment or rater to another. Measurement error is a systematic and random error of the participant's performance that is not attributed to true changes in the construct to be measured.[15] Sample size was defined according to COSMIN guidelines, which recommend that a minimum of 50 individuals should be recruited to ensure the quality of studies assessing the measurement properties of instruments.[16]

Participants

The recruitment of people with COPD was carried out at the Family Health Unit of Joane, Vila Nova de Famalicão, Portugal. During a routine medical appointment with their general practitioner, these people were informed of the existence and nature of this research project through the reading of the informed consent. This document described all the conditions of the study and participation (from displacements and risks of data collection) and availability of contact with the researchers. Therefore, the people who agreed to participate signed the informed consent form and, afterwards, the research team contacted them.

Participants were considered eligible if they met the following criteria: diagnosis of COPD according to the Global initiative for chronic Obstructive Lung Disease (GOLD) criteria—postbronchodilator forced expiratory volume in one second (FEV1)/forced vital capacity ratio<70%,[17] stable COPD (absence of acute exacerbation of respiratory symptoms resulting in additional therapy in the last 4 weeks) and ability to provide informed consent. Participants were excluded if they had other lung diseases, presence of a significant cardiac, musculoskeletal or neuromuscular disease, signs of cognitive impairment, significant balance disorder, current neoplastic or immunological disease, any therapeutic intervention additional to usual care, and previous or current participation in PR programmes.

Data collection

The accepted participants were contacted by the research team to schedule two assessment visits, within 7–10 days apart.

In the first visit, the sociodemographic and clinical data, patient-reported outcomes measures (PROMs) and the performance of two IST (IST1 and IST2) were collected. In the second visit, the performance of two ISWT attempts (ISWT1 and ISWT2) were collected. Before this, the stability of the health status of the participants was ensured, questioning them about possible changes in their symptoms or the presence of new ones since the last visit. Both visits were carried out at the participants' homes, except in some cases, where there was no physical space to perform the ISWT. In these cases, the ISWT was performed at the Family Health Unit of Joane, Vila Nova de Famalicão, Portugal.

Sociodemographic and clinical data

On the first visit, the last report of the lung function tests (spirometry) were collected from all participants and, according to the GOLD guidelines, the airflow limitation of COPD (GOLD I, II, III, IV) of each participant was classified according to the FEV1(%) values.[17]

Sociodemographic (age, sex) and clinical data (medication, comorbidities, smoking status, long-term oxygen use, non-invasive ventilation, number of exacerbations, unscheduled consultations, emergency department admission and hospitalisations in the previous year) were collected. Anthropometric data (height, weight and body mass index) were collected using a measuring tape and bioelectrical impedance measure (HomeFashion, New York, USA), with Bluetooth connection via the Apple Store or Google Play app.

Patient-reported outcome measures (PROMs)

The modified Medical Research Council (mMRC) Scale and the COPD Assessment Test (CAT) were used to assess dyspnoea and the impact of COPD, respectively. The Portuguese versions of the tests, which are available in the Directorate-General of Health of Portugal website, were used.[18] The mMRC is a scale in which the individual classifies the severity of dyspnoea between 0 and 4, with higher scores corresponding to greater severity.[19] On the other hand, the CAT is a scale that assesses the impact of COPD through symptoms on people's lives: cough, sputum, chest tightness, breathlessness when climbing ramps or stairs, limitation in household activities, confidence in leaving the house, sleep and energy. On these topics, the participant scores from 0 to 5 points the intensity of the symptoms, being able to present a total score from 0 to 40 points, with higher values corresponding to a greater impact of the disease.[20 21]

The application of the mMRC and CAT, together with information on non-scheduled consultations, emergencies and hospitalisations due to acute exacerbations of COPD in the last year, also permitted the assessment of the symptoms and the risk of exacerbations through the GOLD ABE tool assessment on each participant.[17]

Incremental step test

The IST was designed to provide an incremental profile by using a digital recording with timed metronome step cadence, and with a 20 cm tall platform (Max Aerobic step, Mambo, Tisselt, Belgium). The number of levels and duration of each level (increment) were based on the characteristics of the ISWT.[9] The original protocol of the ISWT consists of 12 levels; however, as suggested by the literature, we can add more levels to the protocol (total of 15 levels) to allow its future application for other clinical populations, in order to prevent the ceiling effect.[22] Therefore, the IST consists of 15 levels, each of 1 min duration. The timed metronome set the step cadence which starts at 10 steps/min and increases 2 steps/min every 1 min, with a step cadence maximum of 38 steps/min (level 15). The maximum test duration is 15 min. Heart rate (HR) and peripheral oxygen saturation (%SpO_2) were monitored and registered during the test with a pulse oximeter (PalmSAT 2500 Series, Nonin Medical, Minnesota, USA). The perceived dyspnoea and leg fatigue during the test were also registered with the mBorg scale. The blood pressure was not assessed due to the difficulty of measuring during the stepping. The second test (IST2) was performed after 30 min rest or a return of the vital signs, %SpO_2 and mBorg to baseline values.

The criteria to stop the test were: the inability to maintain the required step cadence for 10 s, %SpO_2 falls to ≤85%, when requested by the participant, or when symptoms were reported (chest pain, intolerable dyspnoea, leg cramps, diaphoresis and a pale or ashen appearance). The total number of steps performed (main outcome measure of the IST), the maximal step cadence reached and the duration of the test were also collected. The predicted distance was calculated for each participant based on the available reference equation[23] to calculate the percentage predicted value (% predicted) using the following formula:

$$\%\ predicted = \left(distance\ performed \div distance\ predicted\right) \times 100.$$

The instructions to perform the IST and a reporting form are available as online supplemental material in Vilarinho *et al*.[9]

Incremental shuttle walk test

This test was conducted according to the American Thoracic Society/European Respiratory Society standards.[3] The main outcome measures of the ISWT were distance and maximum speed achieved. The predicted distance was calculated for each participant based on the available reference equation[24] to calculate the percentage predicted value (% predicted) using the following formula:

$$\%\ predicted = \left(distance\ performed \div distance\ predicted\right) \times 100.$$

HR, %SpO_2, perceived dyspnoea and leg fatigue (mBorg scale) were monitored before, during each level and after the test. Blood pressure was only recorded before and after the test. The second test (ISWT2) was performed after a 30 min rest or a return of the vital signs, %SpO_2 and mBorg to baseline values.

The ISWT stopping criteria were more than 0.5 m away from the cone when the beeper sounds for the second consecutive time, when requested by the participant, reaching 85% of their predicted maximum HR, $\%SpO_2$ falling to ≤85%, with symptoms of chest pain, mental confusion or lack of coordination, dyspnoea or fatigue intolerable, or any other relevant clinical factor.[25]

Data analysis

Data analysis was performed using IBM SPSS Statistics V.28.0 (International Business Machines). The level of significance was set at 0.05. Continuous variables were tested for normality using the Kolmogorov-Smirnov test. For descriptive statistics, data were presented by mean and SD, median and percentiles (percentile 25–75) or frequencies (percentage).

For the validity assessment, the construct validity was analysed through the correlation between the highest step cadence of the last completed level and the highest number of steps achieved (best IST), and the highest speed of the last completed level and the highest distance achieved (best ISWT), using the Spearman correlation coefficient. The use of the best tests is due to the fact that people with COPD usually present a learning effect for field tests when at least two tests are conducted and, therefore, it is recommended to use the best performance for clinical studies and clinical practice.[12] According to COSMIN recommendations, a 'positive' rating to qualify construct validity is determined if the correlation coefficient is equal to or above 0.5.[16 26] The strength of these correlations was also analysed based on the *British Medical Journal* guidelines: significant correlation coefficients of 0–0.19 as very weak, 0.2–0.39 as weak, 0.4–0.59 as moderate, 0.6–0.79 as strong and 0.8–1.0 as very strong.[27]

Within-day test–retest reliability (number of steps, duration and cadence of the IST) was calculated by the ICC model 2 (random two-way effects), with a single rater ($ICC_{2,1}$), absolute agreement and with 95% CI.[28] According to COSMIN guidelines, an ICC value above 0.70 can be considered 'positive'.[16 26]

The measurement error was determined by calculating the SE of measurement (SEM) and the minimal detectable change at 95% CI (MDC95).[29] To calculate the SEM the following formula was used:

$$SEM = SD \times \sqrt{1 - ICC},$$

where SD is the SD of the performances of the two attempts of the IST obtained from all participants. The %SEM was calculated using the equation:

$$\%SEM = \left(SEM \div mean\right) \times 100,$$

where the 'mean' is the mean of the performances of the two attempts of the IST obtained from all participants. For the calculation of MDC95 and %MDC95, the following equations were used, respectively:

$$MCD95 = 1.96 \times SEM \times \sqrt{2}$$

$$\%MCD95 = \left(MCD95 \div mean\right) \times 100,$$

where the mean is the mean of the performances of the two attempts of the IST obtained from all participants. Some authors consider a result lower than 30% an acceptable value of %MDC95.[30 31] Additionally, the Bland and Altman 95% limits of agreement (LoA) were calculated, using the equation:

$$LoA = mean_{difference} \pm 1.96 \times SD_{diff},$$

where $mean_{diff}$ and SD_{diff} are the mean and SD of the differences between IST1 and IST2, respectively.[32]

The learning effect of the IST was analysed using Wilcoxon signed-rank test to compare the performance (number of steps, duration and step cadence) between IST1 and IST2. The same test was used to compare the physiological response (HR, $\%SpO_2$, dyspnoea and leg fatigue) between IST1 and IST2 and between the best IST and the best ISWT.

Patient and public involvement

None.

RESULTS

In total, 85 participants with COPD were screened to be included in the study. Of those, 20 participants were excluded due to: dropped out with no reason given (n=4), no telephone answering (n=9), no booking available (n=3), presence of a significant musculoskeletal disease (n=2) or neurological disease (n=2). A total of 65 participants were eligible for the study; however, two participants did not appear for the data collection (n=2). Thus, 63 participants were included (figure 1). 45 participants received both visits at home to perform the IST and ISWT, and 18 participants performed the ISWT at the Family Health Unit due to the lack of sufficient physical space at the participants' homes.

The characteristics of the 63 participants included in the study are presented in table 1. Most of these participants were men (49 men, 77.8%), aged 67.5±10.5 years, had moderate airflow limitation (GOLD II, 31 participants, 49.2%) and belonged to GOLD A group (30 participants, 47.6%). Two participants used long-term oxygen therapy (3.2%) and they used it during the study. Nine participants used non-invasive ventilation (14.3%) (table 1).

Construct validity

The median of the best number of steps performed in the IST by the participants was 52 (24.0; 100.0) steps, with a % predicted of 37.0%, and the median of the best distance performed in the ISWT was 280 m (170.0; 440.0), with a % predicted of 58.9%. According to the ABE GOLD classification, participants in GOLD A performed a median of 97.50 (51.0; 113.50) steps with a % predicted of 50.7%, participants in GOLD B performed 25.0 (19.0; 56.0) steps with a % predicted of 23.7% and participants in GOLD E performed 34.0 (20.75; 57.75) steps with a % predicted of 24.1%. In the ISWT, participants in GOLD A performed a median of 410.0 m (270.0; 540.0) with a % predicted

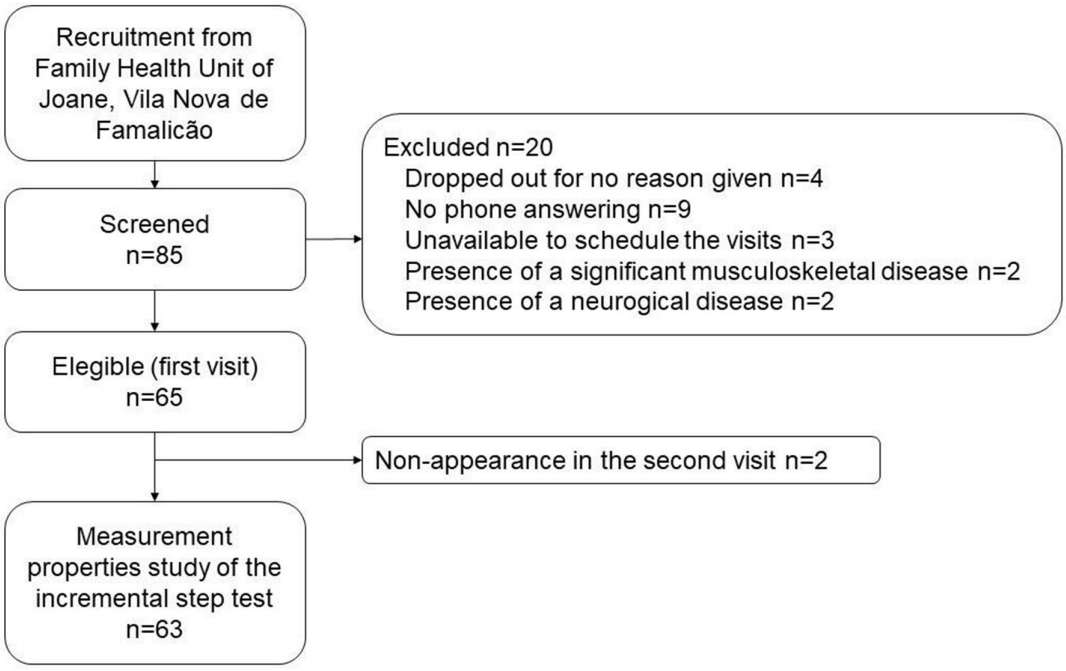

Figure 1 Study flow diagram.

of 66.5%, participants in GOLD B performed 180.0 m (130.0; 340.0) with a % predicted of 51.1% and participants in GOLD E performed 185.0 m (177.50; 447.50) with a % predicted of 56.6%.

The correlations between the number of steps of the best IST and the distance and speed of the best ISWT were significant, 'positive' and strong (ρ=0.74, p<0.001 and ρ=0.74, p<0.001, respectively). In addition, the correlation between the cadence of the best IST and the distance and speed of the best ISWT were also significant, 'positive' and strong (ρ=0.73, p<0.001 and ρ=0.72, p<0.001, respectively) (figure 2).

Within-day test–retest reliability, measurement error and learning effect

The IST presented a 'positive' value of $ICC_{2,1}$ for the number of steps and duration (0.95; 95% CI 0.92 to 0.97), and for cadence (0.94; 95% CI 0.91 to 0.97). The SEM presented a value of 11.7 steps (%SEM=18.9%) and the MDC95 showed a value of 32.4 steps (%MDC95=52.2%). For this reason, %MDC95 was considered unacceptable. The LoA plot showing the differences between IST1 and IST2 number of steps is presented in figure 3. The average bias (ie, the mean of the differences, $mean_{diff}$) was −1.06 number of steps and LoA ranged from −33.61 to 31.48 number of steps. Measurements from four participants fell outside the lower limit of the LoA. No evidence of systematic error was found (figure 3).

No significant differences were observed in the number of steps, duration or step cadence (p=0.070; p=0.056; p=0.582, respectively) between IST1 and IST2. No significant differences were found in HR, $\%SpO_2$, dyspnoea and leg fatigue in pretest and post-test between IST1 and IST2. Significant differences (p<0.001) were found in HR, $\%SpO_2$, dyspnoea and leg fatigue before and after the completion of each test (IST1 and IST2) (table 2).

Significant differences were found in HR, %SpO2, dyspnoea and leg fatigue between the best IST and ISWT, highlighting the higher dyspnoea and leg fatigue observed in the IST (online supplemental table 1).

DISCUSSION

This study demonstrated that the IST presented significant, 'positive' and strong correlations with the ISWT (correlation values between 0.72 to 0.74, p<0.001). Furthermore, the within-day test–retest reliability was 'positive' and no learning effect was observed. However, despite the absence of a systematic error in the Bland Altman analysis, the measurement error of the IST was considered unacceptable, mainly due to the SEM and MDC95 values and percentages.

Regarding the construct validity, the results were in accordance with the predefined hypotheses, with distance and speed of the ISWT presenting the strongest correlations with the number of steps of the IST (ρ=0.74, p<0.001). These results can be explained by the fact that the IST was based on the characteristics of the ISWT (incremental profile, the same number of levels and the same duration of each level).[3 25] This analysis was based on the methodology and sample size recommended by COSMIN guidelines, which support a high level of evidence for this measurement property of the IST.[11 16 26 33] In fact, other step tests applied in people with COPD showed a low level of evidence for their validity due to the small sample sizes included in the studies and inadequate methodologies that were not based on COSMIN guidelines.[5] Therefore, our results place the IST as a valid

Table 1 Participant baseline characteristics

Characteristics	Eligible participants (n=63)
Age, years	67.5±10.5
Sex, male (%)	49 (77.8)
BMI, kg/m^2	26.8±4.6
mMRC (total score)	1.0 (1.0; 2.0)
CAT (total score)	12.1±8.1
GOLD stages (I, II, II, IV), n (%)	10, 31, 13, 9 (15.9, 49.2, 20.6, 14.3)
GOLD group (A, B, E), n (%)	30, 25, 8 (47.6, 39.7, 12.6)
FEV1, %predicted	62.2±19.3
FEV1/FVC%	56.5±10.6
Long-term oxygen therapy, n (%)	2 (3.2)
Non-invasive ventilation, n (%)	9 (14.3)
Number of exacerbations (previous year), n	31
Unscheduled consultations (previous year), n	20
Emergency department admission (previous year), n	6
Hospitalisations (previous year), n	7
Comorbidities, n (%)	
Cardiac disease	7 (11.1)
Arrhythmia	1 (1.6)
Heart failure	11 (17.5)
Hypertension	41 (65.1)
Diabetes	16 (25.4)
Musculoskeletal	13 (20.6)
OSAS	2 (3.2)
Medication, n (%)	
SABA	9 (14.3)
LABA	24 (38.1)
SAMA	11 (17.5)
LAMA	8 (12.7)
LABA+LAMA	6 (9.5)
LABA+ICS	17 (27.0)
ICS	2 (3.2)

Data are expressed as mean±SD or median (percentile 25–75), unless otherwise stated.
BMI, body mass index; CAT, COPD Assessment Test; COPD, chronic obstructive pulmonary disease; FEV1, forced expiratory volume in one second; FVC, forced vital capacity; GOLD, Global Initiative for Chronic Obstructive Lung Disease; ICS, inhaled corticosteroids; LABA, long-acting β2-agonists; LAMA, long-acting muscarinic antagonists; mMRC, modified Medical Research Council; OSAS, obstructive sleep apnoea syndrome; SABA, short-acting β2-agonists; SAMA, short-acting muscarinic antagonists.

field test to assess exercise capacity in people with COPD in clinical practice.

An important observation from our results was the lower % predicted values for the IST in the total sample and in all subgroups (GOLD A, B and E) compared with the % predicted values for the ISWT. In fact, these lower performances in IST were expected since it is demonstrated in evidence that, in people with COPD, stair-climbing results in more blood lactate production, lung hyperinflation and dyspnoea when compared with walking.[34] For these reasons, the stepping training should be included in PR programmes to improve patients' exercise capacity, especially for patients who express limitations in climbing stairs. The information collected on perceived dyspnoea during stair climbing activity in PROMs (eg, CAT and London Chest Activity of Daily Living) can be considered for this purpose.[20 21 35]

Another interesting observation from our study is that most participants performed both tests (IST and ISWT) at home, but we believe that these data may not be representative of all people with COPD in the community. The decision to perform the ISWT was made during the first visit, where the IST was performed, allowing the researcher to assess the home environment and to decide whether the necessary conditions for carrying out this walking test were guaranteed. We believe that the opportunity to perform both tests in participants' homes may have ensured participants' adherence to complete this study since one of the benefits of home care is that patients feel more comfortable receiving care services in their own homes.[36] Furthermore, it was possible to perform the ISWT in the majority of participants' homes since only a 10 m corridor is needed to perform this walking test.[3] Although exercise capacity can be assessed with other walking tests, such as the 6MWT, which is largely used in clinical studies and clinical practice, it requires a longer corridor.[3] Therefore, its applicability in the home environment is more limited and step tests can be an important alternative in these circumstances.

In the within-day reliability analysis, the present study obtained a 'positive' classification with an $ICC_{2,1}$ higher than 0.70 (0.95; 95% CI 0.92 to 0.97), which indicates that the IST provides consistent results when two tests are applied on the same day.[16 26] These results were similar to those previously presented on the between-day reliability of the IST ($ICC_{2,1}$=0.96; 95% CI 0.92 to 0.98),[9] showing once again the consistency of the performance of the test during two attempts, regardless of the time application mode between attempts.

On the other hand, in this study, the SEM and the MCD95 found for the number of steps of the IST were high, indicating a large variation in measurement error. In fact, these values were similar to the values found in the between-day measurement error and both are considered unacceptable (between-day results: 10.1 steps (%SEM=16.6%) and 27.9 steps (%MDC95=45.8%).[9] We suppose that these findings were influenced by the high heterogeneity of groups of participants included in our sample based on the ABE assessment tool (GOLD A, B, E), thus increasing the measurement error. The ABE classification appears to be important to discriminate people with the worst outcomes,[17 37] and, therefore, participants from each ABE group present heterogeneity of symptoms and exercise capacity levels, despite their stable COPD

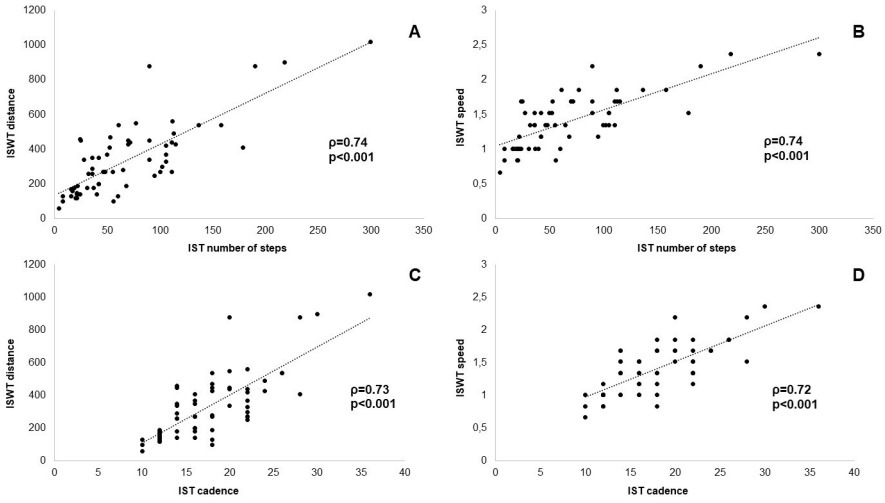

Figure 2 Correlations between the incremental step test (number of steps and cadence) and the incremental shuttle walk test (distance and speed). IST, incremental step test; ISWT, incremental shuttle walk test.

condition during the study. Besides, different comorbidities presented by the target population of the present study (table 1), such as diabetes, obesity and cardiovascular disease, can also influence the measurement error.[1] On the other hand, other potential factors that may have influenced these results should be considered, including the differences observed in the perception of dyspnoea and fatigue between the IST and the ISWT. Evidence suggests that a higher perception of effort, in which tests are perceived as more effortful, can have negative effects on motivation.[38 39] According to the mBorg ratings in our study, participants experienced more dyspnoea and leg fatigue in the IST than in the ISWT (online supplemental table 1), p<0.001), which may have caused different levels of motivation between the tests. This suggestion should be considered and confirmed in future studies through the assessment of motivation and, additionally, the level of comfort experienced by the participants and preference for the test modality (stepping or walking). This will allow us to analyse whether these factors can contribute to more acceptable results in the measurement error of the IST. All the explanations presented above may also justify the wide range of the LoA in our Bland and Altman analysis; however, no evidence of systematic bias was observed. As a limitation of our study, the quality of the measurement error of this study cannot be determined according to COSMIN guidelines. To rate the quality of an error measurement, MDC95 or LoA must be compared with the

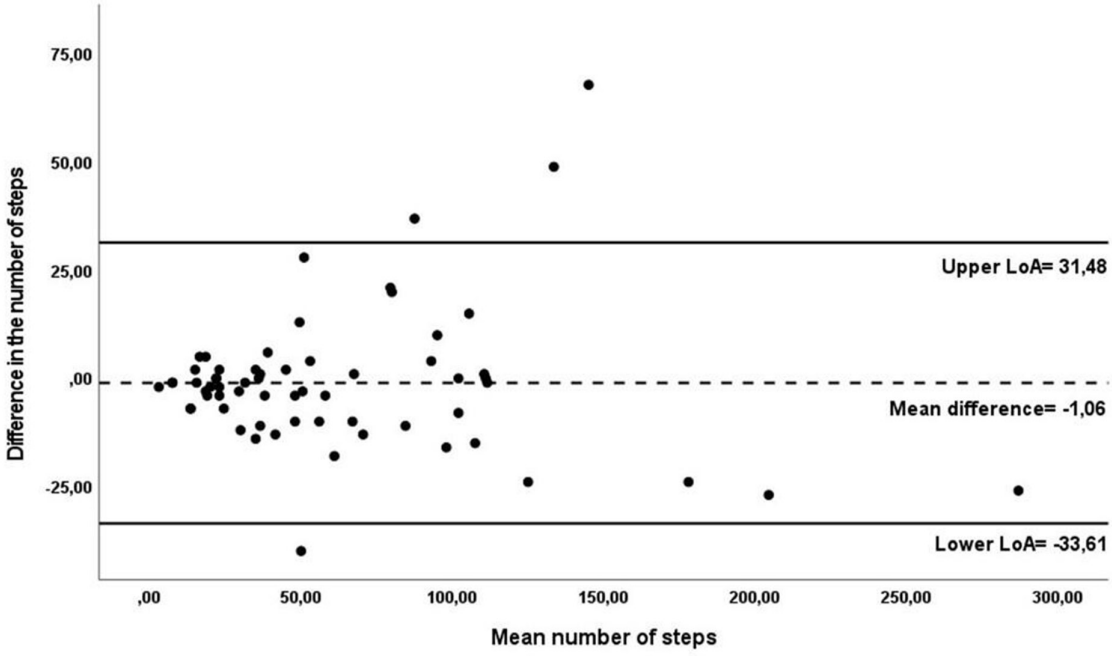

Figure 3 Bland and Altman plot of the difference between number of steps in the incremental step test-1 and the incremental step test-2 against the mean of the number of steps in the incremental step test-1 and the incremental step test-2. LoA, limits of agreement.

Table 2 Performance and response of the two incremental step tests (IST1 and IST2)

(n=63)	IST1		IST2	
Steps, n	46.0 (22.0; 95.0)		48.0 (24.0; 90.0)	
Duration (s)	219.0 (117.0; 381.0)		236.0 (119.0; 365.0)	
Cadence (steps/min)	16.0 (12.0; 22.0)		16.0 (14.0; 20.0)	
	Pre	**Post**	**Pre**	**Post**
Heart rate (bpm)	77.1±13.7	105.3±17.9*	76.2±13.2	106.2±20.1*
SpO_2 (%)	95.2±2.2	92.4±4.3*	95.5±2.1	92.2±4.4*
Dyspnoea (rating)	0.0 (0.0; 1.0)	6.0 (5.0; 6.0)*	0.0 (0.0; 1.0)	6.0 (5.0; 6.0)*
Fatigue (rating)	0.0 (0.0; 0.0)	3.0 (2.0; 5.0)*	0.0 (0.0; 0.0)	3.0 (2.0; 5.0)*

Data are expressed as mean±SD or median (percentile 25–75).
*P<0.001 for comparisons within tests (pretest vs post-test).
IST, incremental step test; SpO_2, peripheral oxygen saturation.

minimal important change (MIC), which is defined as the smallest change in the outcome of interest that patients perceive as important, either beneficial or harmful, and that would lead the patient to consider a change in management.[40] However, no MIC is yet available for the IST (eg, based on PR programmes) and, consequently, its measurement error is indeterminate.[26]

Lastly, the results of this study showed an absence of learning effect when two IST are applied on the same day, suggesting that a single IST may be sufficient in people with COPD. Even so, this result was different when two IST were applied on different days (7 days apart), where a learning effect was observed (5% more in the number of steps in the second test, p<0.05).[9] These differences may be due to the different settings where participants were recruited for the studies. In the between-day reliability study, participants were recruited from hospitals and clinics, and are more likely to present more symptoms and risk of exacerbations, confirmed according to their ABE classification (20% in GOLD A, 60% in GOLD B and 20% in GOLD E).[9] On the other hand, the participants in the present study were recruited from a family health unit and had fewer symptoms and risk of exacerbations, which may explain the absence of a learning effect in the IST. Therefore, given the lack of agreement between the results, we suggest the performance of one or two IST can be influenced by the participants' health status and the setting where the recruitment is performed. Even so, to follow the technical standards of field tests for people with COPD (eg, walking tests)[3] the performance of two IST and the record of the best one is also recommended, in general, in clinical practice to determine the exercise capacity of people with COPD.

This study presents some strengths and limitations that need to be acknowledged. As mentioned before, the methodology and sample size used were defined following COSMIN guidelines, which provide general principles that should be considered in the design of all studies on measurement properties. Another important strength is that we determine the construct validity by comparing the performance of the IST with the performance of the test on which it was based (ISWT).

As limitations, we only had the opportunity to compare our test with other field tests and not with the CPET, the gold standard test to assess the exercise capacity.[41] This comparison would allow the determination of criterion validity of the IST by correlating the performance and cardiorespiratory variables between the tests, especially the VO_2peak. Additionally, this would allow us to analyse whether IST can have a maximal cardiorespiratory response in people with COPD, similar to CPET, supporting its capacity to be considered a maximal and symptom-limited test. If confirmed, this will contribute to the application of a new alternative as the basis for individualised prescription of endurance training (step training) intensity in this population.

Another limitation concerns the fact that most participants were male and classified in the GOLD A and GOLD B groups (according to the ABE assessment tool); however, despite the fact that people with COPD in these groups present fewer symptoms and/or less risk of exacerbation, the decrease in exercise capacity level was notorious in our participants according to the % predicted values in the IST (total sample 37.0%, GOLD A 50.7% and GOLD B 23.7%) and ISWT (total sample 58.9%, GOLD A 66.5% and GOLD B 51.1%). Even so, further studies are important to determine the measurement properties based on each GOLD ABE group, recruiting more participants, especially in the most advanced stage of the disease (GOLD E). This will allow us, for example, to confirm whether measurement error results are similar or different between groups. Studies with participants referred to PR programmes will also lead us to determine the responsiveness of the IST, as another important measurement property considered by the COSMIN guidelines.[15]

In addition, more settings are important in future studies for the recruitment of participants, since the participants of our study were only recruited at a family health unit which also hinders the generalisation of the findings.

CONCLUSION

The IST can be suggested as a valid test to assess exercise capacity in people with COPD, based on its 'positive' and strong correlations with the ISWT. This test also showed 'positive' within-day reliability with no learning effect. The measurement error of the IST is considered indeterminate.

Author affiliations
[1]Center for Rehabilitation Research (CIR), School of Health, Polytechnic Institute of Porto, Porto, Portugal
[2]Department of Medicine, Faculty of Medicine, University of Porto, Porto, Portugal
[3]Instituto CUF Porto, Porto, Portugal
[4]FP-I3ID, Escola Superior de Saúde Fernando Pessoa, Porto, Portugal
[5]Microbiology Research Laboratory on Environmental Health (EnviHealthMicroLab), Institute of Environmental Health (ISAMB), Associate Laboratory TERRA, Faculty of Medicine, Universidade de Lisboa (ULisboa), Lisbon, Portugal
[6]Institute for Preventive Medicine and Public Health, Faculty of Medicine, Universidade de Lisboa (ULisboa), Lisboa, Portugal
[7]Healthcare Department, Nippon Gases Portugal, Maia, Portugal
[8]Egas Moniz Center for Interdisciplinary Research (CiiEM), Egas Moniz School of Health and Science, Almada, Portugal
[9]Department of Physiotherapy, Santa Maria Health School, Porto, Portugal

Twitter João Carlos Winck @joaowinck

Acknowledgements We would like to thank the general practitioners at the family health unit of Joane, Vila Nova de Famalicão, who helped us with the recruitment and sharing of clinical data from the participants.

Contributors Guarantor of the study: RV. Conceptualisation: RV and AMM. Methodology: TG, RV and JCW. Validation: RV, CC and AMM. Formal analysis: TG and RV. Investigation: TG, RV, JCW, CC and AMM. Resources: TG and RV. Data curation: CC and AMM. Writing—original draft preparation: TG and RV. Writing—review and editing: JCW, FS, CC and AMM. Supervision: RV, CC and AMM. All authors read and agreed to the published version of the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by ethics committee of the School of Health of the Polytechnic of Porto with the registration number E0134 and the ethics committee of the Northern Regional Health Administration of Portugal (ARS Norte, registration number: CE/2022/121). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

ORCID iDs
Cátia Caneiras http://orcid.org/0000-0002-3735-8554
Rui Vilarinho http://orcid.org/0000-0002-4422-8440

REFERENCES

1 Bui K-L, Nyberg A, Maltais F, *et al*. Functional Tests in Chronic Obstructive Pulmonary Disease, Part 1: Clinical Relevance and Links to the International Classification of Functioning, Disability, and Health. *Ann Am Thorac Soc* 2017;14:778–84.
2 Souto-Miranda S, Vaes AW, Gloeckl R, *et al*. International perspectives on outcome measurement in pulmonary rehabilitation of people with COPD: A qualitative study. *Respir Med* 2022;201:106936.
3 Holland AE, Spruit MA, Troosters T, *et al*. An official European Respiratory Society/American Thoracic Society technical standard: field walking tests in chronic respiratory disease. *Eur Respir J* 2014;44:1428–46.
4 Spruit MA, Singh SJ, Garvey C, *et al*. An official American Thoracic Society/European Respiratory Society statement: key concepts and advances in pulmonary rehabilitation. *Am J Respir Crit Care Med* 2013;188:e13–64.
5 Vilarinho R, Caneiras C, Montes AM. Measurement properties of step tests for exercise capacity in COPD: A systematic review. *Clin Rehabil* 2021;35:578–88.
6 Holland AE, Cox NS, Houchen-Wolloff L, *et al*. Defining Modern Pulmonary Rehabilitation. An Official American Thoracic Society Workshop Report. *Ann Am Thorac Soc* 2021;18:e12–29.
7 Desveaux L, Janaudis-Ferreira T, Goldstein R, *et al*. An international comparison of pulmonary rehabilitation: A systematic review. *COPD* 2015;12:144–53.
8 Rochester CL, Vogiatzis I, Holland AE, *et al*. An Official American Thoracic Society/European Respiratory Society Policy Statement: Enhancing Implementation, Use, and Delivery of Pulmonary Rehabilitation. *Am J Respir Crit Care Med* 2015;192:1373–86.
9 Vilarinho R, Serra L, Águas A, *et al*. Validity and reliability of a new incremental step test for people with chronic obstructive pulmonary disease. *BMJ Open Respir Res* 2022;9:e001158.
10 Terwee CB, Bot SDM, de Boer MR, *et al*. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
11 Mokkink L, Prinsen C, Patrick D, *et al*. COSMIN study design checklist for patient-reported outcome measurement instruments. 2019. Available: https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf
12 Singh SJ, Puhan MA, Andrianopoulos V, *et al*. An official systematic review of the European Respiratory Society/American Thoracic Society: measurement properties of field walking tests in chronic respiratory disease. *Eur Respir J* 2014;44:1447–78.
13 von Elm E, Altman DG, Egger M, *et al*. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Journal of Clinical Epidemiology* 2008;61:344–9.
14 Prinsen CAC, Mokkink LB, Bouter LM, *et al*. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1147–57.
15 Mokkink LB, Terwee CB, Knol DL, *et al*. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Med Res Methodol* 2010;10:22.
16 Mokkink LB, de Vet HCW, Prinsen CAC, *et al*. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res* 2018;27:1171–9.
17 Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management and prevention of COPD. 2023. Available: https://goldcopd.org/
18 Direção Geral da Saúde. Orientação Técnica - Programas de Reabilitação Respiratória NOS Cuidados de Saúde Primários. Orientação N° 014/2019 de 07/08/2019. Direção Geral DA Saúde Portugal. n.d. Available: https://www.dgs.pt/directrizes-da-dgs/orientacoes-e-circulares-informativas/orientacao-n-0142019-de-07082019.aspx
19 Crisafulli E, Clini EM. Measures of dyspnea in pulmonary rehabilitation. *Multidiscip Respir Med* 2010;5:202–10.
20 Jones PW, Harding G, Berry P, *et al*. Development and first validation of the COPD Assessment Test. *Eur Respir J* 2009;34:648–54.
21 Dodd JW, Hogg L, Nolan J, *et al*. The COPD assessment test (CAT): response to pulmonary rehabilitation. A multicentre, prospective study. *Thorax* 2011;66:425–9.
22 Dourado VZ, Vidotto MC, Guerra RLF. Reference equations for the performance of healthy adults on field walking tests. *J Bras Pneumol* 2011;37:607–14.

23 Vilarinho R, Toledo A, Silva C, *et al*. Reference Equation of a New Incremental Step Test to Assess Exercise Capacity in the Portuguese Adult Population. *JCM* 2023;12:271.
24 Marques A, Rebelo P, Paixão C, *et al*. Enhancing the assessment of cardiorespiratory fitness using field tests. *Physiotherapy* 2020;109:54–64.
25 Singh SJ, Morgan MD, Scott S, *et al*. Development of a shuttle walking test of disability in patients with chronic airways obstruction. *Thorax* 1992;47:1019–24.
26 Terwee CB, Prinsen CAC, Chiarotto A, *et al*. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018;27:1159–70.
27 The British Medical Journal. Correlation and regression. n.d. Available: https://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/11-correlation-and-regression
28 Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 2016;15:155–63.
29 Bruton A, Conway JH, Holgate ST. Reliability: What is it, and how is it measured? *Physiotherapy* 2000;86:94–9.
30 Huang S-L, Hsieh C-L, Wu R-M, *et al*. Minimal detectable change of the timed “up & go” test and the dynamic gait index in people with Parkinson disease. *Phys Ther* 2011;91:114–21.
31 Smidt N, van der Windt DA, Assendelft WJ, *et al*. Interobserver reproducibility of the assessment of severity of complaints, grip strength, and pressure pain threshold in patients with lateral epicondylitis. *Arch Phys Med Rehabil* 2002;83:1145–50.
32 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–10.
33 Terwee CB, Mokkink LB, Knol DL, *et al*. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012;21:651–7.
34 Dreher M, Walterspacher S, Sonntag F, *et al*. Exercise in severe COPD: Is walking different from stair-climbing? *Respiratory Medicine* 2008;102:912–8.
35 Garrod R, Bestall JC, Paul EA, *et al*. Development and validation of a standardized measure of activity of daily living in patients with severe COPD: the London Chest Activity of Daily Living scale (LCADL). *Respir Med* 2000;94:589–96.
36 Leff B, Burton L, Mader S, *et al*. Satisfaction with hospital at home care. *J Am Geriatr Soc* 2006;54:1355–63.
37 Vanfleteren LEGW, Lindberg A, Zhou C, *et al*. Exacerbation Risk and Mortality in Global Initiative for Chronic Obstructive Lung Disease Group A and B Patients with and without Exacerbation History. *Am J Respir Crit Care Med* 2023;208:163–75.
38 Marcora SM. Do we really need a central governor to explain brain regulation of exercise performance? *Eur J Appl Physiol* 2008;104:929–31;
39 Furlan L, Sterr A. The Applicability of Standard Error of Measurement and Minimal Detectable Change to Motor Learning Research-A Behavioral Study. *Front Hum Neurosci* 2018;12:95.
40 Schünemann HJ, Puhan M, Goldstein R, *et al*. Measurement properties and interpretability of the Chronic respiratory disease questionnaire (CRQ). *COPD* 2005;2:81–9.
41 AmericanCollege of Sports Medicine. *American College of Sports Medicine’s Guidelines for Exercise Testing and Prescription*. Lippincott Williams & Wilkins, 2013.