PULMONOLOGY

www.journalpulmonology.org

ORIGINAL ARTICLE

Reliability and validity of the Chester step test in patients with interstitial lung disease

A. Alves[a,e], A. Oliveira[a,b,c,d], P.G. Ferreira[f,g], V. Martins[h], A. Marques[a,b,*]

[a] Lab3R — Respiratory Research and Rehabilitation Laboratory, School of Health Sciences, University of Aveiro (ESSUA), Aveiro, Portugal
[b] iBiMED — Institute of Biomedicine, Department of Medical Sciences, University of Aveiro, Aveiro, Portugal
[c] Respiratory Medicine, West Park Healthcare Centre, Toronto, ON, Canada
[d] School of Rehabilitation Science, McMaster University, Hamilton, ON, Canada
[e] REMEO Home Department, Linde Healthcare, Porto, Portugal
[f] ILD Outpatient Clinic, Pulmonology Department — Centro Hospitalar Universitário de Coimbra (CHUC), Coimbra, Portugal
[g] Faculty of Medicine of the University of Coimbra (FMUC), Coimbra, Portugal
[h] Pulmonology Department — Hospital Distrital da Figueira da Foz, Figueira da Foz, Portugal

Abstract
Background: The Chester Step Test (CST) is a simple and inexpensive field test, which requires minimal physical space to assess exercise capacity. Such characteristics make the CST suitable to be used in different settings, however, its measurement properties in patients with interstitial lung diseases (ILD) are unknown.
Methods: A cross-sectional study was conducted in patients with ILD. First, a CST-1 and a 6-minute walk test (6MWT) were performed. After 48-72 hours, a CST-2 was repeated. A 2nd rater was present in one of the sessions. Relative reliability was measured using intraclass correlation coefficient ($ICC_{1,1}$ and $ICC_{2,1}$). Absolute reliability was determined using standard error of measurement (SEM), minimal detectable change at 95% confidence interval (MDC_{95}) and the Bland-Altman method. The values of SEM and MDC_{95} were also expressed as a percentage of the mean. Construct validity was explored using Spearman correlation coefficient (r_s) between the number of steps taken in the best CST and the distance performed in the 6MWT.
Results: Sixty-six patients with ILD (65.5±12.9 years; 48.5%men; FVC 79.4±18.8pp; DL_{CO} 49.0±18.3pp) participated in the study. Relative (ICC 0.95-1.0) and absolute reliability were excellent without evidence of systematic bias. The SEM and MDC_{95} were 11.8 (14.7%) and 32.6 steps (40.7%), respectively. The correlation between CST and 6MWT was significant, positive, and high (r_s=0.85, p=0.001).

* Corresponding author at: Lab3R - Respiratory Research and Rehabilitation Laboratory, School of Health Sciences (ESSUA) and Institute of Biomedicine (iBiMED), University of Aveiro, Agras do Crasto - Campus Universitário de Santiago, Edifício 30, 3810-193 Aveiro, Portugal.
E-mail address: amarques@ua.pt (A. Marques).

Conclusion: The CST is a reliable and valid test and might be especially useful to assess exercise capacity in patients with ILD in limited space environments.

Introduction

Interstitial lung diseases (ILD) are a heterogeneous group of diffuse parenchymal lung disorders which, despite presenting diverse etiologies, share several clinical features.[1,2] Although comprising various degrees of inflammation and/or fibrosis, a percentage of ILD patients can develop a progressive self-sustaining fibrosis, namely those with idiopathic nonspecific interstitial pneumonia, unclassifiable idiopathic interstitial pneumonia, connective tissue disease-associated ILDs, hypersensitivity pneumonitis, sarcoidosis and ILDs related to other occupational exposures (e.g. silicosis, asbestosis).[3] This progressive-fibrosing phenotype may lead to worsening of symptoms (mainly exertional dyspnea, dry cough and fatigue), progressive impairment in gas exchange and lung function decline, reduced exercise capacity, muscle dysfunction and reduced quality of life.[4,5] Some of these patients can also present a median survival rate below 5 years after diagnosis and generally exhibit high associated socioeconomic costs, high dependency on others to perform daily living activities and a massive burden on healthcare systems.[6-9]

In patients with ILD, reduced exercise capacity is associated with poor health-related quality of life and increased hospital admissions.[10,11] Finding valid, feasible and standardized tests which enable comparison with standard values and provide agreement between healthcare professionals is, thus, a priority in order to provide the best quality of care to this population.[11] The gold standard to evaluate exercise capacity is cardiopulmonary exercise testing (CPET).[12] However, CPET is not easily available in clinical practice as it requires expensive equipment, the presence of specialized human resources and is time-consuming.[13] To overcome these limitations, field tests such as the 6-minute walk test (6MWT), the incremental shuttle walk test (ISWT) and, more recently, the Chester Step Test (CST) have been used to assess exercise capacity in patients with chronic respiratory diseases.[13-15] Field tests are more affordable and simpler to apply than CPET and are better related to patients' demands during activities of daily living.[13,16] Particularly, the CST requires less space than the other field tests, which allows it to be easily applied in different settings, including inpatient, outpatient and home-based settings.[16]

The CST is an externally paced, incremental and multi-stage test, designed to assess exercise capacity in healthy individuals.[15,16] Recently, it has been validated to assess exercise capacity in patients with chronic obstructive pulmonary disease (COPD).[15] However, the physiological mechanisms of exercise limitation in patients with COPD differ significantly from those in patients with ILD, in whom exercise intolerance is mostly due to impaired gas exchange and circulation limitation.[10,17] Thus, it is imperative to test the measurement properties of the CST in ILD to assure that the

selection of this instrument for research and clinical practice is evidence-based and its results can be reliably interpreted.[18]

The authors hypothesize that the number of steps in the CST will present: (1) excellent intra and inter-rater reliability; (2) significant, positive, and high correlation with the distance covered in the 6MWT (6MWD). The main purpose of this study was to assess the reliability (relative and absolute) and construct validity of the CST in patients with ILD.

Materials and methods

Study design and population

This cross-sectional study was integrated in a larger trial (POCI-01-0145-FEDER-007628, POCI-01-0145-FEDER-028806 and PTDC/SAU-SER/28806/2017), with ethical approval from the *Unidade de Investigação em Ciências da Saúde: Enfermagem* (UICISA: E) of the *Escola Superior de Enfermagem de Coimbra*, Portugal (N°P517-08/2018), from the Ethics Committee for Health of the *Centro Hospitalar do Baixo Vouga*, EPE, Aveiro, Portugal (N/Ref 0863926) and from the *Hospital Distrital da Figueira da Foz*, EPE, Leiria, Portugal (March 15th 2019). All participants signed an informed consent.

The study initiated in January 2019, and it was completed in November 2020. Patients were considered eligible if they were diagnosed by their pulmonologist with any ILD, according with the internationally accepted guidelines,[1,19,20] and were clinically stable over the past month (i.e., no hospital admissions, exacerbations — i.e., acute, clinically significant respiratory deterioration, typically less than 1 month in duration, with new bilateral glass opacity and/or consolidation superimposed on a background pattern consistent with fibrosing ILD[21] - or changes in their pharmacological treatment strategy). They were excluded if they had other lung diseases, signs of cognitive impairment or substance abuse, or presence of a significant cardiovascular, neurological, or musculoskeletal disease that precluded their participation in data collection.

This study was conducted and reported according to the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative and a minimum of fifty participants were aimed to be included in order to achieve a good sample size for measurement properties assessment.[18,22]

Data collection

Participants were asked to attend to two assessment sessions, with at least 48-72 hours apart. Sociodemographic, anthropometric, and clinical data were first obtained to characterize the sample. Lung function tests (spirometry and diffusion capacity for carbon monoxide - DL_{CO}) were conducted during patients' routine medical appointments and collected from

their most recent clinical records. The Self-Administered Comorbidities Questionnaire (SCQ) was used to score the severity of comorbidities.[23,24] This questionnaire is composed of 12 medical and 3 optional conditions and attributes a maximum of 3 points to each condition (1 point for the presence of the problem, 1 point if receiving treatment for it and 1 point if the medical condition limits the person activities). Scores range from 0 to 45, with higher scores indicating more severe comorbidities.[23,24] Then participants performed a CST-1 and a 6MWT, in this specific order. A resting period of at least 30 minutes between tests was given to allow for recovery of vital signs, fatigue, and dyspnea to their baseline values (a longer period was given, after the 30 min, if participants had not returned to their baselines levels). In the second session, a CST-2 was performed. A 2nd rater was present in one of the sessions to assess the CST.

The CST was performed using a digital recording with timed metronome rhythms and a 20 cm tall single-step device.[16] The digital recording also gives the standardized instructions of the test and the chance to practice the test briefly.[16] The CST has 5 stages, lasting 2 minutes each. The timed metronome set the step cadence, which starts at 15 steps/minute and increases 5 steps/minute every 2 minutes: stage 1 (15 steps/minute); stage 2 (20 steps/minute); stage 3 (25 steps/minute); stage 4 (30 steps/minute); stage 5 (35 steps/minute). The maximum test duration is 10 minutes, corresponding to the final stage 5. Heart rate (HR) and peripheral arterial oxygen saturation (SpO_2) were monitored continuously with a pulse oximeter (Konica Minolta, Pulsox-300i, Japan) and recorded on paper by the rater every minute or at the end of each 2-minute stage, respectively.[25] The perceived dyspnea and fatigue were recorded using the modified Borg scale every minute.[25] The CST ended when: the participant reached 80% of the reserve HR[25]; if the SpO_2 dropped below 85%; or if the participant was unable to maintain the step cadence for 15 seconds. Moreover, if the participant showed signs of intolerable dyspnea, being over-tired or dizzy, the CST was immediately terminated. The main outcome measure of the CST was the total number of steps taken. The best of the two CST, where the participant performed the highest number of steps, was selected for validity analysis.

The 6MWT was performed on a flat, straight, 30 meters length corridor with a hard surface, according to the European Respiratory Society/American Thoracic Society guidelines.[26] Before each 6MWT, participants rested on a chair, located near the starting position, for at least 10 minutes. Then, participants were instructed to walk as fast as possible, without running or jogging, for 6 minutes.[26] If participants requested to pause during the test or their SpO_2 dropped below 85%, they could sit on the chairs placed along the corridor.[26] Participants were encouraged to resume walking as soon as they could or when their SpO_2 reached at least 88%. Criteria for immediately terminating the test included chest pain, intolerable dyspnea, leg cramps, staggering, diaphoresis and pale or ashen appearance.[26] Standard encouragement was given each minute.[26] The 6MWD was the main outcome measure.

Statistical analysis

Data analysis was performed using IBM SPSS Statistics (version 25.0, IBM Corporation, Armonk, NY, USA).

Descriptive statistics, i.e., relative frequencies (percentage), mean ± standard deviation (SD) or median [interquartile range] were used to describe the sample. The Kolmogorov-Smirnov test (KS) was used to determine the normality of data distribution.[27] Outliers were identified, through the inspection of extreme points on the plotted graphs of the variables in study and analysis were performed with and without their presence. We decided not to remove outliers since their presence did not affect results significantly. Tests with a $p<0.05$ were considered statistically significant.

Relative reliability was measured using intraclass correlation coefficient (ICC)(29). $ICC_{1,1}$ and $ICC_{2,1}$ models were used to determine intra-rater and inter-rater reliability,[29] respectively, according to the following equations:

$$ICC1,1 = \frac{BMS - WMS}{BMS + (k - 1)WMS} \qquad \text{(equation 1)}$$

$$ICC2,1 = \frac{BMS - WMS}{BMS + (k - 1)WMS} \qquad \text{(equation 2)}$$

where BMS is between-subjects mean squares, WMS is within-subjects mean squares, EMS is the error (residual) of mean squares, RMS is between raters mean squares, k is the number of measurements/raters (k=1), and n is the number of participants. An ICC lower than 0.50 was considered of poor reliability, 0.5-0.75 moderate, 0.75-0.90 good and greater than 0.9 excellent reliability.[18,28]

Absolute reliability was determined by calculating the standard error of measurement (SEM) and the minimal detectable change at 95% confidence interval (MDC_{95})(29). The SEM was measured according to the following equation: $SEM = SD_{difference}/\sqrt{2}$), where $SD_{difference}$ is the SD of the differences between the CST-1 and CST-2.[18,30] The MDC_{95} was calculated as follows: $MDC_{95} = 1.96 \times \sqrt{2} \times SEM$.[18] The values of SEM and MDC_{95} were also expressed as a percentage of the mean and calculated as follow: $SEM\% = (SEM/mean) \times 100$ and $MDC_{95}\% = (MDC_{95}/mean) \times 100$, where mean is the mean of the number of steps taken in CST-1 and CST-2. A $MDC_{95}\%$ of less than 30% was considered acceptable.[31]

The Bland-Altman method was applied to calculate absolute reliability.[18,22] First, we plotted the difference between the number of steps taken in CST-1 and CST-2 against the mean of the number of steps taken in CST-1 and CST-2.[32] Then, we calculated the mean and SD of the differences between CST-1 and CST-2, the closer the mean difference is to zero and the smaller the SD of the differences, the more reliable is the measure.[32] Finally, we calculated the 95% limits of agreement (LoA95) as follows: $LoA_{95} = mean_{difference} \pm 1.96 \times SD_{differences}$.[32]

Construct validity was assessed by analysing the relationship between the number of steps taken in the best CST and the 6MWD using Spearman correlation coefficient (r_s)(18). A correlation of 0-0.3 was considered poor, 0.3-0.5 weak, 0.5-0.7 moderate, 0.7-0.9 strong, and 0.9-1.0 excellent.[18,27]

Results

Seventy-eight patients with ILD were screened to be included in the study. Sixty-six were eligible to participate and twelve were excluded for the following reasons: decline to

participate (n=4), drop out for no reason given (n=1), presence of a significant cardiovascular disease (ischemic cardiomyopathy associated with myocardial infarction, n=4), presence of a significant musculoskeletal disease (severe gonarthrosis, n=2), presence of cognitive impairment (Alzheimer's disease, n=1). Sixty-six participants were included for the construct validity study and fifty-three for the reliability study, since thirteen individuals did not attend the second assessment session due to participants unavailability. A flow diagram of recruitment is provided in Fig. 1.

Eligible participants were on average 65.5±12.9 years old, slightly overweight (body mass index= 28.9±5.2kg/m^2) and 48.5% were male (n=32). Most prevalent types of ILD were chronic hypersensitivity pneumonitis (n=29, 43.9%), followed by idiopathic pulmonary fibrosis (n=16, 24.2%) and sarcoidosis (n=6, 9.1%). Participants presented a mean forced expiratory volume in one second (FEV$_1$) of 81.9± 19.9% of predicted and mean forced vital capacity (FVC) of 79.4±18.8% of predicted. Nineteen participants (28.8%) presented mild ILD (DL$_{CO}$ >60% predicted), 18 participants (27.3%) moderate ILD (40%≤DL$_{CO}$≤60% predicted) and 19 participants (28.8%) severe ILD (DL$_{CO}$<40% predicted). We were unable to access DL$_{CO}$ values for 10 patients (15.1%). Thirty-one participants used long-term oxygen therapy (47%) and five participants used non-invasive ventilation during sleep (7.6%). The mean number of steps taken in CST-1 and CST-2 were 77.7±50.2, and 82.4±55.7, respectively. The main reason for CST termination was the inability to maintain the required step cadence. The mean 6MWD was 399.4±128.2 meters (83.1±26.4% of predicted). A detailed sample characterization is summarized in Table 1.

Reliability

The CST demonstrated excellent relative reliability, for both intra-rater reliability (ICC$_{1,1}$=0.95; 95%CI 0.91-0.97) and inter-rater reliability (ICC$_{2,1}$=1.0; 95%CI 0.99-1.0). Regarding absolute reliability, SEM and MDC$_{95}$ values were 11.8 steps (SEM%=14.7%) and 32.6 steps (MDC$_{95}$%=40.7%), respectively. The Bland-Altman plot was created and a mean difference of -4.72 steps was observed with the LoA95 ranging from -37.28 and 27.84 steps (Fig. 2).

Validity

The correlation between the number of steps of the best CST and the 6MWD was significant, positive, and strong (r$_s$=0.85, p=0.001) (Fig. 3).

Discussion

Excellent intra-rater and inter-rater reliability were found for the CST. A similar study, which focused on patients with COPD and assessed the reliability of the CST also showed excellent relative reliability (ICC of 0.99; 95%CI 0.97-0.99) (15). This finding indicates that CST provides consistent results and excellent agreement between healthcare professionals, which allows for comparisons of patients' results even when CST is applied by different raters on different occasions, i.e., pre and post PR program.[11,33] Moreover, it also suggests that only minimal training is required for healthcare professionals to apply the CST.
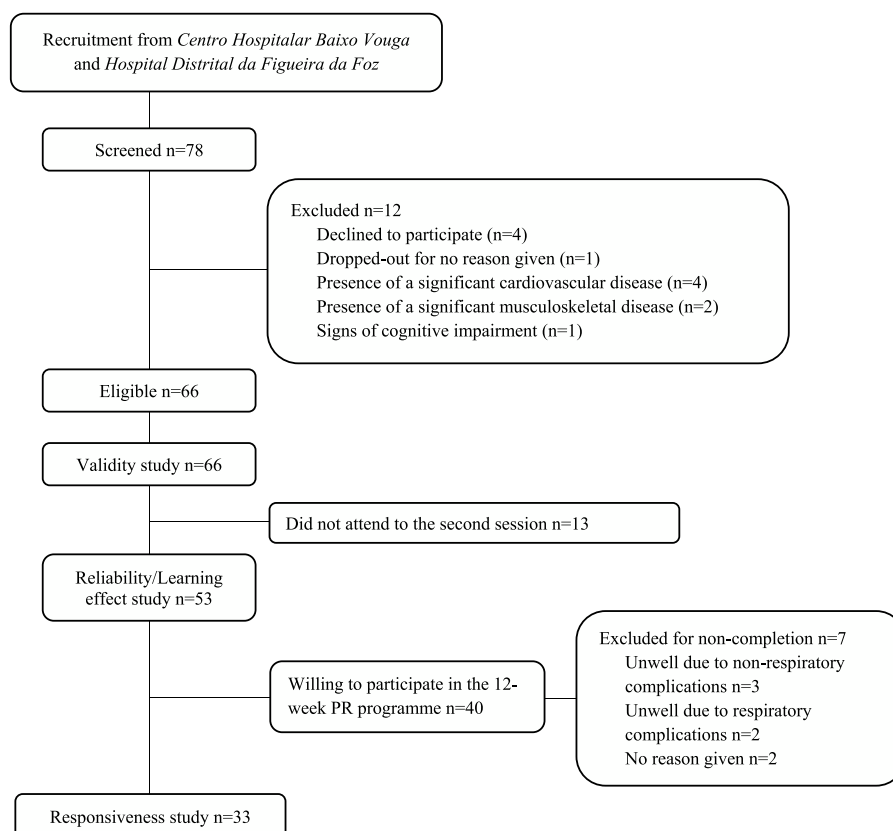

Fig. 1 Flow diagram of participants with interstitial lung disease included in the study.

Table 1 Sample characterization (n=66).

Characteristics	Eligible participants (n=66)
Age, years	65.5±12.9
Gender, male n (%)	32 (48.5)
BMI, kg/m^2	28.9±5.2
Smoking status, n (%)	
Current	2 (3)
Former	24 (36.4)
Never	42 (63.6)
Packs/year	35.0 [7.0-55.8]
Exacerbations/year, n (%)	
0	48 (72.7)
1	14 (21.2)
≥2	4 (6)
Lung function	
FEV$_1$, L	2.1±0.7
FEV$_1$, %predicted [*1]	83.7±20.8
FVC, L	2.5±0.8
FVC, %predicted [*1]	79.4±18.8
FEV$_1$/FVC, %	82.7±9.2
DL$_{CO}$, %predicted [*1]	49.0±18.3
DL$_{CO}$ >60%predicted, n (%)	19 (28.8)
40%≤DL$_{CO}$ ≤60%predicted, n (%)	18 (27.3)
DL$_{CO}$ <40%predicted, n (%)	19 (28.8)
Long-term oxygen therapy, n (%)	31 (47)
Non-invasive ventilation, n (%)	5 (7.6)
Pharmacological treatment for ILD, n (%)	
Glucocorticoids	43 (65.2)
Immunosuppressant	30 (45.5)
Antifibrotics	5 (7.6)
ILD types, n (%)	
IPF	16 (24.2)
Sarcoidosis	6 (9.1)
Chronic hypersensitivity pneumonitis	29 (43.9)
NSIP secondary to systemic sclerosis	2 (3)
UIP secondary to systemic sclerosis	4 (6.1)
UIP secondary to rheumatoid arthritis	2 (3)
Anti-synthetase syndrome	2 (3)
Desquamative interstitial pneumonia	1 (1.5)
LIP related to Sjogren's syndrome	1 (1.5)
Silicosis	1 (1.5)
Respiratory bronchiolitis ILD	1 (1.5)
Follicular bronchiolitis related to Sjogren's syndrome	1 (1.5)
SCQ	9±3.9
CST-1	77.7±50.2
CST-2	82.4±55.7
6MWT	399.4±128.2
6MWT, % of predicted [*2]	83.1±26.4

Notes: Values are presented as mean±standard deviation or median [interquartile range].

Legend: BMI, body mass index; CST, Chester Step Test; FEV$_1$, forced expiratory volume in one second; FVC, forced vital capacity; DL$_{CO}$, diffusion capacity for carbon monoxide; ILD, interstitial lung disease; IPF, idiopathic pulmonary fibrosis; LIP, lymphocytic interstitial pneumonia; NSIP, non−specific interstitial pneumonia; UIP, usual interstitial pneumonia; SCQ, self-administered comorbidities questionnaire; 6MWT, 6-minute walk test.

[*1] computed according to Standardization of Spirometry 2019 Update[41];

[*2] computed according to Marques et al 2020[42]

Our findings suggest that it is necessary to improve above 32.6 steps to assume that the change observed was above the measurement error.[18] Although this cut-off is informative, whether that change is clinically meaningful remains unknown. Moreover, our MDC$_{95}$% was above the 30% acceptable limit.[31] It is likely that this finding was influenced by the high heterogeneity of symptoms and exercise capacity observed in our sample of patients within disease subgroups (mild DL$_{CO}$ 28.8%; moderate DL$_{CO}$ 27.3% and severe DL$_{CO}$ 28.8%), thus increasing the MDC$_{95}$ value.[1,34] Therefore, future studies determining the minimal clinically important differences of the CST in patients with ILD, ideally using homogeneous samples, and using both anchor (i.e., mean change, the receiver operating characteristic curves and linear regression analysis) and distribution-based methods (i.e., 0.5 times the SD, SEM, 1.96 times the SEM; MDC$_{95}$ and ES)[14] are needed.

In the present study, no evidence of systematic bias was observed. In patients with COPD, the CST showed smaller and narrower values of mean difference and LoA95 (mean difference of -1.1 steps with LoA95 ranging from -20.2 to 17.9 steps).[15] It is likely that the high heterogeneity observed in our sample of patients within disease subgroups may have contributed to the wider range of the LoA95, especially those patients that achieved higher stages on the CST.

The CST has adequate construct validity as shown by the strong correlation between the number of steps and the 6MWD. Similar results were found for the ISWT in the same population (r=0.76, p<0.0001), while a lower correlation was observed in patients with COPD (r=0.60, p=0.001).[13,15] In addition to the adequate validity, the CST presents some advantages over the ISWT and the 6MWT, such as the apparent absence of a learning effect and requiring minimal space for its application.[35] As reported previously, the 6MWT requires a 30 meters length corridor, and the ISWT a 10 meters corridor, which is often difficult to find in home and clinical settings.[35,36]

The CST may also be useful in other contexts and populations. During the COVID-19 pandemic, the need to transfer patients to rehabilitation facilities or discharge them to home increased exponentially and so did the demand for home-
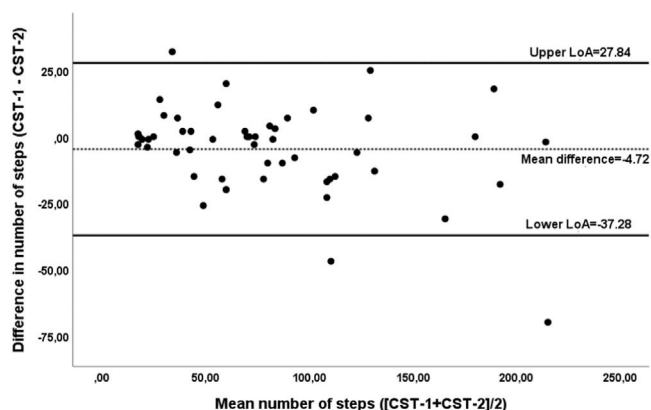

Fig. 2 Bland and Altman plot of the difference between number of steps in the Chester Step Test-1 (CST-1) and CST-2 against the mean of the number of steps in test -1 and test -2 in patients with interstitial lung disease (n=66). The dashed horizontal line represents the mean difference, and the solid horizontal lines represent the 95% upper and lower limits of agreement.
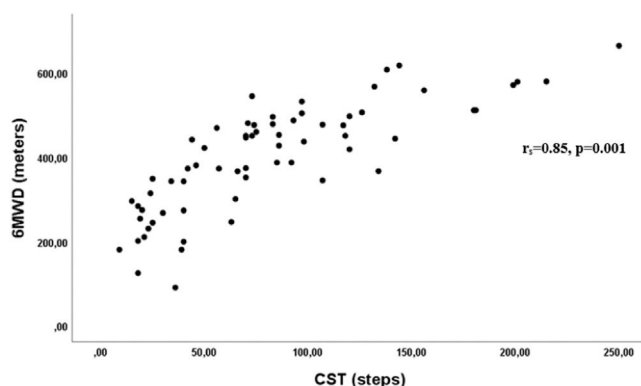
A. Alves, A. Oliveira, P.G. Ferreira et al.

Fig. 3 Correlation between the number of steps of the best Chester Step Test (CST) and the 6-minute walk distance (6MWD) in patients with interstitial lung disease (n=66).

based and/or remote rehabilitation.[37,38] Worldwide, healthcare professionals and researchers were looking for validated, easy and practical to implement exercise tests that allowed assessing and prescribing exercise in non-clinical settings.[39] Step tests emerged as validated measures, however up until now only the 3-minute step test had been used in such settings.[39] To the best of the authors' knowledge, the current study is the first to show that the CTS is safe, reliable, and valid to be conducted in a community setting. Futures studies should be conducted to assess its applicability in other populations, such as patients post-COVID-19.

Assessment of exercise capacity is of paramount importance as exercise capacity in patients with ILD has been associated with a more reserved prognosis and a poorer health-related quality of life.[11,17,1] Clinicians and researchers now have at their disposal a field test that is valid and reliable to assess exercise capacity in patients with ILD even in limited space settings.

Limitations

This study has some limitations that need to be acknowledged. First, we only assessed construct and not criterion validity (i.e., correlation with the gold standard − CPET). Second, we did not evaluate the effects of the disease severity and variability on the MDC_{95}, since our sample size was too small to perform subgroup analysis. Third, our study only included patients with clinical stable ILD and participants with any other significant impairment or disease were excluded, thus our results cannot be generalized to all patients. Modified versions of the CST may be needed either to assess patients only with mild ILD or to safely assess patients with more severe impairments.

Conclusions

The CST is a reliable and valid test to evaluate exercise capacity in patients with clinical stable ILD. Due to its characteristics, the CST may constitute an appropriate alternative to the 6MWT and the ISWT in limited space environments.

Conflicts of interest

The authors report no financial, or non-financial, conflicts of interest in this work.

Acknowledgements

References

1. Antoniou KM, Margaritopoulos GA, Tomassetti S, Bonella F, Costabel U, Poletti V. Interstitial lung disease. Eur Respirat Rev. 2014;23(131):40−54.
2. Olson AL, Gifford AH, Inase N, Fernández Pérez ER, Suda T. The epidemiology of idiopathic pulmonary fibrosis and interstitial lung diseases at risk of a progressive-fibrosing phenotype. Eur Respir Rev. 2018;27(150):180077.
3. Cottin V, Hirani NA, Hotchkin DL, Nambiar AM, Ogura T, Otaola M, et al. Presentation, diagnosis and clinical course of the spectrum of progressive-fibrosing interstitial lung diseases. Eur Respirat Rev. 2018;27(150).
4. Wells AU, Brown KK, Flaherty KR, Kolb M, Thannickal Victor J. What's in a name? That which we call IPF, by any other name would act the same. Eur Respir J. 2018;51(5):1800692.
5. Cottin V, Wollin L, Fischer A, Quaresma M, Stowasser S, Harari S. Fibrosing interstitial lung diseases: knowns and unknowns. Eur Respirat Rev. 2019;28(151):180100.
6. Algamdi M, Sadatsafavi M, Fisher JH, Morisset J, Johannson KA, Fell CD, et al. Costs of workplace productivity loss in patients with fibrotic interstitial lung disease. Chest. 2019;156 (5):887−95.
7. Belkin A, Albright K, Swigris JJ. A qualitative study of informal caregivers' perspectives on the effects of idiopathic pulmonary fibrosis. BMJ Open Respira Res. 2014;1(1):e000007.
8. Holtze C, Flaherty K, Kreuter M, Luppi F, Moua T, Vancheri C, et al. Healthcare utilisation and costs in the diagnosis and treatment of progressive-fibrosing interstitial lung diseases. Eur Respirat Rev. 2018;27(150):180078.
9. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med. 2011;183(6):788−824.
10. Holland AE. Exercise limitation in interstitial lung disease-mechanisms, significance and therapeutic options. Chronic Respirat Dis. 2010;7(2):101−11.
11. Puente-Maestu L. Physiological rationale of commonly used clinical exercise tests. Pulmonol. 2020;26(3):159−65.

12. American Thoracic Society. ATS/ACCP statement on cardiopulmonary exercise testing. Am J Respir Crit Care Med. 2003;167 (2):211−77.

13. Singh S, Moiz JA, Ali MS, Talwar D. Reliability, validity, and responsiveness of the incremental shuttle walk test in patients with interstitial lung disease. J Cardiopulmonary Rehabil Prevent. 2018;38(6):425−9.

14. RMd Bois, D Weycker, Albera C, Bradford WZ, Costabel U, Kartashov A, et al. Six-minute-walk test in idiopathic pulmonary fibrosis: test validation and minimal clinically important difference. Am J Respir Crit Care Med. 2011;183(9):1231−7.

15. de Camargo AA, Justino T, de Andrade CHS, Malaguti C, Dal Corso S. Chester step test in patients with COPD: reliability and correlation with pulmonary function test results. Respir Care. 2011;56(7):995−1001.

16. Sykes K, Roberts A. The Chester step test—a simple yet effective tool for the prediction of aerobic capacity. Physiotherapy. 2004;90(4):183−8.

17. Spruit MA, Singh SJ, Garvey C, ZuWallack R, Nici L, Rochester 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(8): e13−64.

18. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34−42.

19. Hutchinson J. Statement on sarcoidosis. Joint Statement of the American Thoracic Society (ATS), the European Respiratory Society (ERS) and the World Association of Sarcoidosis and Other Granulomatous Disorders (WASOG) adopted by the ATS Board of Directors and by the ER. Am J Respir Crit Care Med. 1999;160 (2):736−55.

20. Travis WD, Costabel U, Hansell DM, King TE Jr, Lynch DA, Nicholson AG, et al. An official American Thoracic Society/European Respiratory Society statement: update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. Am J Respir Crit Care Med. 2013;188(6):733−48.

21. Kolb M, Bondue B, Pesci A, Miyazaki Y, Song JW, Bhatt NY, et al. Acute exacerbations of progressive-fibrosing interstitial lung diseases. Eur Respirat Rev. 2018;27(150):180071.

22. Mokkink LB, Prinsen CA, Patrick DL, Alonso J, Bouter LM, de Vet HC, et al. COSMIN Study Design checklist for Patient-reported outcome measurement instruments. 2019.

23. Molto A, Dougados M. Comorbidity indices. Clin Exp Rheumatol. 2014;32(5 Suppl 85):S131−S4.

24. Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. Arthritis Care Res. 2003;49(2):156−63.

25. Sykes K. Chester aerobic test software. Cardiac rehabilitation ed. UK: Cartwright Fitness Ltd 2013 - 2018; 2016.

26. Holland AE, Spruit MA, Troosters T, Puhan MA, Pepin V, Saey D, et al. An official European Respiratory Society/American Thoracic Society technical standard: field walking tests in chronic respiratory disease. p 1428, 6th Eur Respir J; 2014. p. 1428−46.

27. Cohen J. Statistical power analysis for the behavioral sciences. Routledge; 2013.

28. Bruton A, Conway JH, Holgate ST. Reliability: what is it, and how is it measured? Physiotherapy. 2000;86(2):94−9.

29. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;86(2):420.

30. De Vet HC, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine: a practical guide. Cambridge University Press; 2011.

31. Jácome C, Cruz J, Oliveira A, Marques A. Validity, reliability, and ability to identify fall status of the Berg Balance Scale, BESTest, Mini-BESTest, and Brief-BESTest in patients with COPD. Phys Ther. 2016;96(11):1807−15.

32. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8(2):135−60.

33. Coolican H. Research methods and statistics in psychology. Routledge; 2018.

34. Dowman L, McDonald CF, Hill C, Lee A, Barker K, Boote C, et al. The benefits of exercise training in interstitial lung disease: protocol for a multicentre randomised controlled trial. BMC Pulmonary Med. 2013;13(1):1−8.

35. Coll F, Hill K, Burrows S, Watson C, Edgar D. Modified Chester step test in a healthy adult population: measurement properties and development of a regression equation to estimate test duration. Phys Ther. 2020;100(8):1411−8.

36. Maddocks M, Nolan CM, Man WD. Simple functional tests in COPD: stand up and be counted!. Eur Respirat J. 2017;49 (3):1700104.

37. Polastri M, Lazzeri M, Jácome C, Vitacca M, Costi S, Clini E, et al. Rehabilitative practice in Europe: roles and competencies of physiotherapists. Are we learning something new from COVID-19 pandemic? 2021.

38. Polastri M, Ciasca A, Nava S, Andreoli E. Two years of COVID-19: trends in rehabilitation. Elsevier; 2022.

39. Holland AE, Malaguti C, Hoffman M, Lahham A, Burge AT, Dowman L, et al. Home-based or remote exercise testing in chronic respiratory disease, during the COVID-19 pandemic and beyond: a rapid review. Chronic Respirat Dis. 2020;17:1479973120952418.

40. Graham BL, Steenbruggen I, Miller MR, Barjaktarevic IZ, Cooper BG, Hall GL, et al. Standardization of spirometry 2019 update. An official American thoracic society and European respiratory society technical statement. Am J Respir Crit Care Med. 2019;200(8):e70−88.

41. Marques A, Rebelo P, Paixão C, Almeida S, Jácome C, Cruz J, et al. Enhancing the assessment of cardiorespiratory fitness using field tests. Physiotherapy. 2020;109:54−64.