Performance-based outcome measures to assess functionality in hospitalised patients with COPD exacerbations: a systematic review of the measurement properties

Naiara Tais Leonardi [1], Débora Mayumi Oliveira Kawakami[1], John R. Hurst [2], Joana Cruz [3] and Renata Gonçalves Mendes[1]

[1]Department of Physical Therapy, Federal University of Sao Carlos, São Paulo, Brazil. [2]UCL Respiratory, University College London, London, UK. [3]School of Health Sciences of the Polytechnic of Leiria, Leiria, Portugal.

Corresponding author: Renata Gonçalves Mendes (renatamendes@ufscar.br)

Shareable abstract (@ERSpublications)
Measurement properties of performance-based outcome measures to assess functionality in inpatients with exacerbations of COPD are still scarce. There is a low quality of evidence supporting validity and lack of evidence of responsiveness and reliability. https://bit.ly/41IauH3

Cite this article as: Leonardi NT, Kawakami DMO, Hurst JR, *et al*. Performance-based outcome measures to assess functionality in hospitalised patients with COPD exacerbations: a systematic review of the measurement properties. *Eur Respir Rev* 2023; 32: 230013 [DOI: 10.1183/16000617.0013-2023].

Received: 13 Jan 2023
Accepted: 12 May 2023

Abstract

Introduction Hospitalised patients with exacerbations of COPD (ECOPD) may have physical and functional impairments that impact morbidity and readmission. Therefore, it is crucial to properly identify reduced functionality in these patients to support a personalised rehabilitation. The objective of this study is to summarise and compare the measurement properties of functionality performance-based outcome measures for hospitalised patients with ECOPD.

Methods A systematic review based on the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) was performed. The PubMed, Embase, PEDro and Cochrane databases were searched using terms related to functionality, hospitalised patients with ECOPD and measurement properties. Studies were selected and extracted by two researchers. The COSMIN Risk of Bias checklist was applied to assess the methodological quality of the studies and measurement property results were compared with the criteria for good measurement properties. Quality of evidence was graded using a modified Grades of Recommendation, Assessment, Development and Evaluation approach.

Results 13 studies were included with nine outcome measures, namely the 6-min pegboard ring test, the de Morton mobility index, the incremental shuttle walk test (ISWT), the 6-min walk test (6MWT), maximum inspiratory pressure (MIP), the Berg balance scale, 4-m gait speed, handgrip strength and the 6-min stepper test. Construct validity was rated as sufficient, except for the ISWT. Responsiveness, assessed only for MIP, was considered insufficient and measurement errors for the ISWT and 6MWT were insufficient, with a very low quality of evidence for all measurement properties.

Conclusion Measurement properties of performance-based outcome measures to assess functionality in patients hospitalised with ECOPD are still scarce, with very low evidence supporting validity and a lack of evidence of responsiveness and reliability. Further studies are needed to address this topic and guide assertive and personalised management.

Introduction

COPD is characterised by chronic respiratory symptoms and persistent, often progressive, airflow limitation [1]. Exacerbations of COPD (ECOPD) are the major cause of morbidity [2] and mortality in these patients [3]. Among other factors, an increase in systemic inflammatory markers, use of glucocorticoids [4], energy imbalance and hospitalisation culminate in greater physical and functional impairment [5].

ECOPD may lead to short-and long-term impact on functionality, which involves body function and activities, as defined by the International Classification of Functioning Disability and Health (ICF) [6].

These include reduced exercise tolerance, quadriceps muscle strength and daily physical activity [7–9], which may not fully recover after an event [10]. Reduced physical activity after an ECOPD may also contribute to home isolation and higher rates of depression and physical impairments, generating a vicious cycle of increased exacerbations, re-admissions and mortality [11].

An individualised and multidimensional assessment identifying treatable issues is essential to guide personalised patient management using "treatable traits" [12]. By definition, treatable traits must 1) have clinical relevance, 2) be identifiable and measurable, and 3) be treatable. In a patient with ECOPD, the impairment of functionality – *e.g.* muscle weakness, sarcopenia, exercise intolerance and physical inactivity – may be considered as a treatable trait [13] and must be adequately measured with good-quality measures to support the development and implementation of targeted and effective treatment. However, there are no guidelines on the best functionality markers in the hospital setting for patients with ECOPD. Functionality markers using tests, professionals and instructions are known as performance-based outcome measures [14].

In a previous study, Oliveira and Marques [15] carried out a systematic review on the outcome measures considered in different rehabilitation scenarios (*i.e.* hospital-based, home-based) involving patients with ECOPD, identifying several measures of functional exercise capacity in acute events, specifically the 6-min walk test (6MWT), the incremental shuttle walk test (ISWT), the endurance shuttle walk test, the 3-min step test, the 3-min walk test and the 2-min step-in-place test, which highlights the importance of properly identifying functionality in these patients. However, in addition to having a range of options, it is essential to know whether these measures are demonstrably adequate for their purpose, how they compare with similar measures and how to interpret the results they produce [16].

Therefore, there is a need to identify measures to assess functionality (*i.e.* body function and activities) that are feasible and have appropriate measurement properties for hospitalised patients with ECOPD to facilitate healthcare professional decision-making. Thus, the aim of this study was to summarise and compare the measurement properties of functionality performance-based outcome measures for hospitalised patients with ECOPD.

Methods

Study design

This systematic review was carried out according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses [17] and COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) recommendations [14, 18–20] and prospectively registered in PROSPERO (CRD42021281879).

Search strategy and databases

The key elements recommended by COSMIN were used to define the search strategy as follows [18–20]: 1) construct – functionality (body functions and activities, as defined by the ICF [6]); 2) population – hospitalised patients with ECOPD; and 3) measurement properties – the filter developed by Terwee *et al.* [21], which is a highly sensitive and precise search filter for finding studies on measurement properties. Health descriptors (MeSH and Emtree) and other free terms were used. The PubMED, Embase, PEDro and Cochrane databases were used, including English, Spanish and Portuguese entries (until August 2021), and those updated on PubMed (September 2021–July 2022) (see supplementary material).

Eligibility criteria

According to the COSMIN recommendations [18–20], the inclusion criteria were as follows: 1) performance-based outcome measures should aim to assess one construct within the concept of functionality (body functions and activities, according to ICF definitions [6]); 2) the study sample should represent hospitalised patients with ECOPD (>50%, according to the COSMIN suggestion [18–20]); and 3) the study should evaluate one or more measurement property, feasibility or interpretability of the instrument. Exclusion criteria consisted of the following: 1) studies involving patients with neurological or orthopaedic impairments, patients who were intubated, and outpatients; 2) book chapters, congress abstracts, letters to the editor, comments, unpublished articles and protocol studies; and 3) studies that only used the performance-based outcome measure as an outcome measurement instrument to measure the outcome in an intervention.

Selection of studies

Studies were selected in the StArt tool (State of the Art through Systematic Review, version 2.3.4.2, São Carlos, São Paulo, Brazil) by two independent researchers (N.T.L. and D.M.O.K.) and any cases of

disagreement were resolved by consensus with a third reviewer (R.G.M). First, studies were screened according to the title and abstract, and if a study appeared relevant by at least one reviewer based on the abstract, or in case of doubt, the full-text article was retrieved and assessed for eligibility.

Data extraction

Data were extracted to an Excel sheet by two researchers to avoid errors or loss of important information. The data included summaries of study characteristics (authors, year, country, study design and sample characteristics), performance-based outcome measures (name, protocol, variable and equipment) and results of the measurement properties analysis (measurement property, type of analysis, comparator and quantitative result). Furthermore, feasibility data, such as test application time, adverse events, causes of interruption and patient's required mental and physical ability level to perform the test, and interpretability data, such as minimal important difference, distribution of the score and floor and ceiling effects, were also collected.

ICF classification

Variables obtained from the performance-based outcome measures and analysed in the studies were classified into two components of functionality, according to ICF definitions, *i.e.* body functions and activities. Body functions are the physiological functions of body systems and activities are the execution of a task or action by an individual [6]. Based on these classifications, each performance-based outcome measure was classified into functionality categories by consensus of the review team, considering analysed variables in the included studies and previous literature [22–24].

Methodological quality of studies

The methodological quality of studies was assessed by two independent researchers using the COSMIN Risk of Bias checklist, in which the measurement properties were identified and corresponding boxes filled in [18–20]. They were classified as very good, adequate, doubtful or inadequate. Measurement properties assessed included reliability, measurement error, criterion validity, hypothesis testing for construct validity and responsiveness, according to the COSMIN taxonomy framework. Reliability and measurement error boxes were adapted for performance-based outcome measures, as recommended by Mokkink *et al*. [14]. In cases where studies did not present a pre-established hypothesis for construct validity and responsiveness, we formulated a set of hypotheses about the expected relationships between the performance-based outcome measure under review and another comparator according to the literature to meet COSMIN recommendations [18–20]. For hypotheses testing for construct validity and responsiveness, expected direction (positive or negative association) and strength (small (0.00–0.25), low (0.26–0.49), moderate (0.50–0.69), high (0.70–0.89) and very high (0.90 and 1.00) correlations [25]), based on the literature, were considered (table S3). Structural validity, internal consistency and cross-cultural validity were not analysed, considering that performance-based outcome measures are clinometric measures and these analyses would not be appropriate.

Criteria for good measurement properties

Results of the measurement properties of the performance-based outcome measures were evaluated against the updated criteria for good measurement properties [18–20]. Each measurement property was classified as sufficient (+), insufficient (−) or indeterminate (?) according to the Prinsen consensus [19]. According to the COSMIN guidelines [18–20], to be considered sufficient (+), hypotheses testing for construct validity must have results in accordance with the hypothesis, reliability must have an intraclass correlation coefficient (ICC) or weighted Kappa $\geqslant 0.70$, measurement error must have a smallest detectable change or limit of agreement <minimal important change [14], and responsiveness must have results in accordance with the hypothesis or area under the curve <0.70.

Summarising evidence and grading quality of evidence

To reach an overall conclusion about the quality of the instrument, the consistency of the results from all available studies were considered for each measurement property. Consistent results were qualitatively summarised (*e.g.* providing the range of the values found or the number of confirmed hypotheses for construct validity) and compared against the criteria for good measurement properties to determine whether, in general, the measurement property of an instrument was sufficient (+), insufficient (−), inconsistent (±) or indeterminate (?) [18–20]. To rate the qualitatively summarised results as sufficient (or insufficient), in principle 75% of the results should meet the criteria and should be analysed by the authors [18–20].

Finally, the quality of evidence was graded based on qualitative syntheses as high, moderate, low or very low using a modified Grades of Recommendation, Assessment, Development and Evaluation (GRADE)

approach recommended by COSMIN, which uses four factors: 1) risk of bias (*i.e.* the methodological quality of the studies); 2) inconsistency (*i.e.* unexplained inconsistency of results across studies); 3) imprecision (*i.e.* total sample size of the available studies); and 4) indirectness (*i.e.* evidence from different populations than the population of interest in the review) [18–20]. The starting point is always the assumption that the pooled or overall result is of high quality. The quality of evidence is subsequently downgraded by one or two levels per factor to moderate, low or very low evidence. If the results were inconsistent among studies, we used the following as possible strategies: 1) find explanations and summarise by subgroup; 2) not summarising the results and not classifying the evidence; or 3) basing the conclusion on the most consistent results and downgrading for inconsistency. The review team decided which strategy was most appropriate for use in each specific situation [18–20].

Results

Study selection and characteristics

The search strategy for functionality performance-based outcome measures identified 8699 records. After screening and eligibility criteria, 12 studies were included in the first search and one study in the updated search, totalling 13 studies (figure S1). Table 1 presents the outcomes, performance-based outcome measures, characteristics of studies, variables analysed and classified ICF functionality components. In general, patients were older, diagnosed with COPD according to the Global Initiative for Chronic Obstructive Lung Disease criteria [26] and there was no common time point for the functionality assessment of the patients during hospitalisation, ranging from the first 24 h from admission to discharge.

Outcomes, outcome measures and ICF components

We identified nine performance-based outcome measures assessing functionality outcomes. Functionality categories were classified as the body function category, which included "functions of the cardiovascular and respiratory systems" and "neuromusculoskeletal and movement-related functions", and the activities category, which included "mobility" (table S1). Outcome measures were divided into eight functionality outcomes, as follows: 1) lower limb function and exercise tolerance using the 6-min stepper test (6MST) [27]; 2) peripheral muscle strength using handgrip strength (HGS) [28]; 3) respiratory muscle strength using maximal inspiratory pressure (MIP) [29, 30]; 4) upper limb function and exercise tolerance using the 6-min pegboard and ring test (6PBRT) [31, 32]; 5) balance using the Berg balance scale (BBS) [33]; 6) mobility using the de Morton mobility index (DEMMI) [34]; 7) walking capacity using 4-m gait speed (4MGS) [35]; and 8) walking capacity and exercise tolerance using the ISWT [36] and 6MWT [37–39]. The characteristics of the outcome measures are presented in table S2.

Measurement properties

In general, most of the studies evaluated the validity of the instruments, mainly by hypothesis testing [27–30, 32–35, 38], two studies evaluated reliability [36, 39] and only two studied responsiveness [29, 37]. Results of the measurement properties are presented in table S3, while results related to COSMIN methodological quality of the studies and measurement properties are shown in table 2. Some studies had a purpose other than assessing instrument validity [28–30, 33, 38]; however, our review team reached a consensus to include studies in which assessing the measurement properties of instruments was not the purpose of the study but still included data relevant to analyse measurement properties. Results of these analysis and the hypotheses formulated by the review team are presented in table S3.

Measurement properties analysis of performance-based outcome measures according to the COSMIN checklist

Body functions

Lower limb function and exercise tolerance

6MST: one study assessed the validity of the 6MST [27]. Construct validity was assessed by Ribeiro *et al.* [27], specifically convergent validity, and they reported a correlation with the 6MWT of rho=0.87 ($p<0.001$). Quality of evidence was graded as very low due to there being only one study of adequate quality with reduced sample size (tables 2 and 3).

Peripheral muscle strength

HGS: one study assessed the validity of HGS [28]. Construct validity was assessed by Turan *et al.* [28], with an unclear objective of analysing the measurement property. A convergent validity with the 6MWT (rho=0.516 ($p<0.001$)), age (rho=−0.250 ($p=0.012$)) and length of hospital stay (rho=−0.247 ($p=0.015$)) was found. We formulated a hypothesis that a positive moderate correlation with the 6MWT ⩾0.50 [40] and with another dissimilar construct would be a positive small correlation of 0.20–0.30 [41]. The known-group validity assessed the difference between ECOPD *versus* COPD stable *versus* non-COPD of dominant hand 0.47±0.17 bar ($p<0.001$) *versus* 0.57±0.16 bar *versus* 0.55±0.16 bar; nondominant hand

https://doi.org/10.1183/16000617.0013-2023

TABLE 1 Outcomes, performance-based outcome measures, population and disease characteristics and analysis of the included studies (n=13)

Functionality outcome	Performance-based outcome measure	Study, year, country; study design	Population	Disease		Analysis		
			Target; n; age; sex; FEV_1	Inclusion criteria	Exclusion criteria	Time point	Measurement properties	Variables (BF; AC)
Body functions								
Lower limb function and exercise tolerance	6-min stepper test	Ribeiro *et al.* [27], 2022, Brazil; cross-sectional observational	Hospitalised patients with ECOPD; n=16; 69.4±11.4 years; 56% female; 49.4±9.9%	No cognitive or motor deficit, no previous cardiovascular disease, no previous thoracoabdominal surgery within 1 month, BMI <30 $kg \cdot m^{-2}$, and no use of vasoactive drugs	Inability to perform the evaluations and cardiorespiratory instability during the tests	As soon as possible: use of NIV less than 2 h per period of 6 h, dyspnoea at rest <7 on the mBorg scale, RR <25 $breaths \cdot min^{-1}$, S_{pO_2}>88%.	Construct validity (convergent validity); reliability	BF: inspiratory capacity, S_{pO_2}, RR, HR, dyspnoea, fatigue, BP AC: cycles performed
Peripheral muscle strength	Handgrip strength	Turan *et al.* [28], 2019, Turkey; cross-sectional observational	Hospitalised patients with ECOPD; n=101; 68.3±9.1 years; 26% female; 38.9±14.6%	ECOPD and stable COPD	Cancer, neurological or rheumatological disease, advanced heart disease or other conditions that might influence upper limb functions, and the cooperation of the individual	NR	Construct validity (known-group, convergent validity)	BF: isometric grip strength AC: NR
Respiratory muscle strength	Maximal inspiratory pressure	Mesquita *et al.* [29], 2013, Brazil; prospective observational	Hospitalised patients with ECOPD; n=19; 67±11 years; 36.8% female; 26 (19–32)%	Spontaneous breathing on hospital admission, no important comorbidities, no recent hospitalisation due to COPD exacerbation and no participation in any exercise training in the previous 6 months	Death, withdrew consent or missed assessment on more than 1 assessment day (discharge and 1 month after discharge)	Day 1 of hospitalisation; at discharge and 1 month after	Construct validity (convergent validity); responsiveness	BF: maximum inspiratory strength AC: NR
		Tudorache, *et al.* [30], 2010, Romania; cross-sectional observational	Hospitalised patients with ECOPD; n=121; 60±12 years; 10% female; 42.5±19.5%	Smokers; during their hospital admission received antibiotics and also systemic corticotherapy treatment for at least 10 days	Myopathy, significant hyperinflation and severe comorbidities	Day 10 of hospitalisation	Construct validity (convergent validity)	BF: maximum inspiratory strength AC: NR

Continued

TABLE 1 Continued

Functionality outcome	Performance-based outcome measure	Study, year, country; study design	Population	Disease		Analysis		
			Target; n; age; sex; FEV_1	Inclusion criteria	Exclusion criteria	Time point	Measurement properties	Variables (BF; AC)
Upper limb function and exercise tolerance	6-min pegboard and ring test	de Barros *et al.* [31], 2020, Brazil; cross-sectional observational	Hospitalised patients with ECOPD; n=18; 71.3±5.1 years; 61% female; 43.2±18.3%	No cognitive or motor deficit, no heart or other pulmonary disease, no recent thoracoabdominal surgery within 1 month, BMI <35 $kg \cdot m^{-2}$, and no use of vasoactive drugs	Inability to perform the evaluations and cardiorespiratory instability during the tests	As soon as possible: use of NIV for less than 2 h per period of 6 h, resting dyspnoea <7 on the mBorg scale, RR <25 $breaths \cdot min^{-1}$, S_{pO_2}>88%	Feasibility	BF: dynamic hyperinflation; dyspnoea; fatigue; HR; BP AC: total number of rings displaced
		Felisberto *et al.* [32], 2018, Brazil; cross-sectional observational	Hospitalised patients with ECOPD; n=17; 70.9±5.1 years; 59% female; 41.8±17.9%	No cognitive or motor deficit, no heart or other pulmonary disease, no recent thoracoabdominal surgery within 1 month, BMI <35 $kg \cdot m^{-2}$, and no use of vasoactive drugs	Inability to perform the evaluations and cardiorespiratory instability during the tests	As soon as possible: use of NIV for less than 2 h per period of 6 h, resting dyspnoea <7 on the mBorg scale, RR <25 $breaths \cdot min^{-1}$, S_{pO_2}>88%	Construct validity (convergent validity, discriminative validity, known-group)	BF: dyspnoea, fatigue, RR, S_{pO_2}, BP, HR AC: total number of rings displaced
Activities								
Balance	Berg balance scale	Oliveira *et al.* [33], 2017, Australia; cross-sectional observational	Hospitalised patients with ECOPD; n=26; age 72±7 years; 50% female; 48±11%	Admitted with ECOPD	Severe neurological or musculoskeletal disease, visual or vestibular problems, inability to perform the testing procedures, and inability to understand spoken English	Prior to discharge (on average, at day of hospitalisation)	Construct validity (convergent validity, known-group)	BF: NR AC: score
Mobility	de Morton mobility index	Camp *et al.* [34], 2019, Canada; cross-sectional observational	Hospitalised patients with ECOPD; n=22; 60±10 years; 36% female; 57±37%	Admitted to an acute care ward for an ECOPD, no intubation or invasive ventilation, and able to understand verbal English instructions	Unable to give informed consent, confused or agitated, and activity contraindicated	Day 3 of hospitalisation	Construct validity (convergent validity, known-group, discriminative validity)	BF: NR AC: total score
Walking capacity	4-m gait speed	Nakano *et al.* [35], 2021, Japan; retrospective observational	Patients with COPD treated at hospital; exacerbation (n=62) and other causes (n=16), n=78; 76.3±0.9 years; 12% female; 47.2±2.6%	NR	NR	Before hospital discharge once patient's condition was stable	Construct validity (convergent validity)	BF: NR AC: gait speed

Continued

https://doi.org/10.1183/16000617.0013-2023

TABLE 1 Continued

Functionality outcome	Performance-based outcome measure	Study, year, country; study design	Population	Disease		Analysis		
			Target; n; age; sex; FEV_1	Inclusion criteria	Exclusion criteria	Time point	Measurement properties	Variables (BF; AC)
Body function and activities								
Exercise tolerance and walking capacity	Incremental shuttle walk test	Johnson-Warrington *et al.* [36], 2015, UK; cross-sectional observational	Hospitalised patients with ECOPD; n=39; 67.7±7.8 years; 48% female; 42.5±13.2%	Established diagnosis of COPD FEV_1/FVC <70%	Locomotive, cardiac, neurological or cognitive comorbidity or had attended PR in the previous 6 months	As close to discharge as possible	Reliability (test–retest)	BF: V'_{O_2}, dyspnoea, HR, S_{pO_2} AC: distance
	6-min walk test	Blankenburg *et al.* [37], 2012, Germany; prospective observational	Hospitalised elderly patients with ECOPD; n=82; 67±9.5 years; 26% female; 40%	Elderly patients that underwent inpatient treatment for ECOPD	Other diseases that might affect physical performance and/or dyspnoea	First 2 days of hospitalisation, 2 days before discharge and 2 days after discharge	Responsiveness	BF: dyspnoea AC: distance
		Liao *et al.* [38], 2020, Taiwan; cross-sectional observational	Hospitalised patients with ECOPD; n=83; 74.0±6.9 years; 10% female; NR	Able to walk, conscious and able to communicate in Mandarin or Taiwanese, and a score of 8 or more on the Short Portable Mental Status Questionnaire	Stroke, myocardial infarction, acute mental illness, dementia, depression, alcohol or drug addiction, NIV or IMV, and history of fracture in the lower extremity	Day 2 of hospitalisation	Construct validity (convergent validity, known-group)	BF: NR AC: distance
		Osadnik *et al.* [39], 2016, Brazil and Australia; cross-sectional observational	Hospitalised patients with ECOPD; n=46; 67.2±11.1 years; 46% female; 43.0±16.2%	No hospitalisation in the last 30 days; no musculoskeletal or neurological conditions; no participation in PR in the last 6 months and no other pulmonary diseases	Transfer to the ICU before the second day of hospitalisation, changes in mental status and worsening of hypoxaemia	Day of hospital discharge	Reliability (test–retest); measurement error	BF: dyspnea, S_{pO_2}, HR AC: distance

Data for age and FEV_1 are presented as mean±SD or median (interquartile range). AC: activities; BF: body function; BMI: body mass index; BP: blood pressure; ECOPD: exacerbation of COPD; FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity; HR: heart rate; ICU: intensive care unit; IMV: invasive mechanical ventilation; NIV: noninvasive ventilation; NR: not reported; PR: pulmonary rehabilitation; RR: respiratory rate; S_{pO_2}: peripheral oxygen saturation; V'_{O_2}: oxygen uptake.

https://doi.org/10.1183/16000617.0013-2023

TABLE 2 Results of the measurement properties analysis of the included performance-based outcome measures according to the COSMIN, including methodological quality, criteria for good measurement property and overall rating

Performance-based outcome measure	Study, year	Validity							Reliability						Responsiveness		
		Criterion validity			Hypothesis testing				Measurement error								
		n	Meth qual[#]	Result[¶]	n	Type	Meth qual[#]	Result[¶,+]	n	Meth qual	Result[¶,§]	n	Meth qual[#]	Result[¶,ƒ]	n	Meth qual	Result[¶,##]
6-min step test	Ribeiro *et al.* [27], 2022				16	Convergent	Adequate	Results in line with one hypothesis (1+)	16	Inadequate	*t*-test						
Summary result (overall rating)					16			1+ (+)									
Handgrip strength	Turan *et al.* [28], 2019				101	Convergent	Doubt	Results in line with three hypotheses (3+); results not in line with two hypotheses (2–)									
					101	Known group	Doubt	(?)									
Summary result (overall rating)					101			3+ and 2– (+)									
Maximal inspiratory pressure	Mesquita *et al.* [29], 2013				19	Convergent	Doubt	Results in line with one hypothesis (1+); results not in line with one hypothesis (1–)							19	Doubt	Spearman rho=0.58 (p=0.01) (+)
	Tudorache *et al.* [30], 2010				121	Convergent	Doubt	Results in line with two hypotheses (2+).									
Summary result (overall rating)					140			3+ and 1– (+)							19		0.58 (+)
Six-peg board ring test	Felisberto *et al.* [32], 2018				17	Convergent	Adequate	Results in line with six hypotheses (6+); results not in line with 10 hypotheses (10–)									
					17	Discriminative	Adequate	Results in line with one hypothesis (1+)									
					17	Known group	Very good	Results in line with one hypothesis (1+)									
Summary result (overall rating)					17			8+ and 10– (+)									
Berg balance scale	Oliveira *et al.* [33], 2017				26	Convergent	Doubt	Results in line with two hypotheses (2+)									
					26	Known group	Doubt	Results not in line with hypotheses (1–)									
Summary result (overall rating)					26			2+ and 1– (+)									
De Morton mobility index	Camp *et al.* [34], 2019				22	Convergent	Adequate	Results in line with two hypotheses (2+); results not in line with one hypothesis (1–)									
					22	Known group	Doubt	(?)									
					22	Discriminative	Very good	Results in line with two hypotheses (2+)									
Summary result (overall rating)					22			4+ and 1– (+)									

Continued

https://doi.org/10.1183/16000617.0013-2023

TABLE 2 Continued

Performance-based outcome measure	Study, year	Validity							Reliability						Responsiveness		
		Criterion validity			Hypothesis testing				Measurement error								
		n	Meth qual[#]	Result[¶]	n	Type	Meth qual[#]	Result[¶,+]	n	Meth qual	Result[¶,§]	n	Meth qual[#]	Result[¶,f]	n	Meth qual	Result[¶,##]
4-m gait speed	Nakano *et al.* [35], 2021				78	Convergent	Adequate	Results in line with two hypotheses (2+)									
Summary result (overall rating)					78			2+ (+)									
Incremental shuttle walk test	Johnson-Warrington *et al.* [36], 2015											39	Inadequate	LoA=(−41.6; 69.8 m) >MCID (47.5 m) (−)			
Summary result (overall rating)												39		LoA=41.6; 69.8 m (−)			
6-min walk test	Blankenburg *et al.* [37], 2012														82	Inadequate	*t*-test: 97±114 −290±190 m (p<0.05) (+)
	Liao *et al.* [38], 2020				83	Convergent	Doubt	Results in line with two hypotheses (2+); results not in line with two hypotheses (2−)									
					83	Known groups	Doubt	(?)									
	Osadnik *et al.* [39], 2016								46	Adequate	$ICC_{2,1}$= 0.885 (0.801–0.934) (+)	46	Adequate	LoA=(−92.2; 104.5 m) >MCID (30 m) (−)			
Summary result (overall rating)					83			2+ and 2− (+)	46		0.885 (+)	46		LoA=92.2; 104.5 m (−)			

ICC: intraclass correlation coefficient; LoA: limit of agreement; MCID: minimum clinically important difference; meth qual: methodological quality; n: sample size. Empty cells in grey represent measurement properties that were not found for the respective performance-based outcome measure in the present study. [#]: Methodological quality was classified as very good, adequate, doubtful or inadequate, according to the COSMIN Risk of Bias Checklist [18–20]. [¶]: Results of the measurement properties were evaluated against the criteria for good measurement properties and classified as sufficient (+), insufficient (–) or indeterminate (?) [18–20]. [+]: Hypotheses testing for construct validity: (+) the result is in accordance with the hypothesis; (?) no hypothesis defined (by the review team); (–) the result is not in accordance with the hypothesis [18–20]. [§]: Reliability: (+) ICC or (weighted) Kappa ⩾0.70; (?) ICC or (weighted) Kappa not reported; (–) ICC or (weighted) Kappa <0.70 [14]. [f]: Measurement error: (+) smallest detectable change (SDC) or limits of agreement (LoA) or coefficient of variation (CV)*$\sqrt{2}$*1.96<MCID, % specific agreement >80%; (?) minimal important change not defined; (–) SDC or LoA or CV*$\sqrt{2}$*1.96>MCID, % specific agreement <80% [14]. [##]: Responsiveness: (+) the result is in accordance with the hypothesis odds ratio area under the curve (OR AUC) ⩾0.70; (?) no hypothesis defined (by the review team); (–) the result is not in accordance with the hypothesis OR AUC <0.70 [18–20].

https://doi.org/10.1183/16000617.0013-2023

TABLE 3 Summary of findings

Outcome/performance-based outcome measure		Measurement property	Summary or pooled result	Overall rating	Quality of evidence
Body function					
Lower limb and exercise tolerance	6MST	Construct validity	One out of one hypothesis confirmed	Sufficient	Very low (one study of adequate quality (−1), imprecision, n<50 (−2))
Peripheral muscle strength	HGS	Construct validity	Three out of five hypotheses confirmed	Sufficient	Very low (one study of doubtful quality (−2), inconsistency (−1))
Respiratory muscle strength	MIP	Construct validity	Three out of four hypotheses confirmed	Sufficient	Very low (multiple studies of doubtful quality (−1), inconsistency (−1), indirectness (−1))
		Responsiveness	Rho: 0.58; sample size: 19	Sufficient	Very low (one study of doubtful quality (−2), imprecision, n<50 (−2))
Upper muscle strength and exercise tolerance	6PBRT	Construct validity	Eight out of 18 hypotheses confirmed	Sufficient	Very low (only one study of adequate quality (−1), Inconsistency in results (−1); imprecision with n<50 (−2))
Activities					
Balance	BBS	Construct validity	Two out of three hypotheses confirmed	Sufficient	Very low (one study of doubtful quality (−2), imprecision, indirectness))
Mobility	DEMMI	Construct validity	Four out of five hypotheses confirmed	Sufficient	Very low (only one study of adequate quality (−1), imprecision with n<50 (−2))
Walking capacity	4mGS	Construct validity	Two out of two hypotheses confirmed	Sufficient	Low (one study of adequate quality (−1), imprecision, n<100 (−1))
Body function and activities					
Exercise tolerance and walking capacity	ISWT	Measurement error	LoA 41.6–69.8 m; greater than MCID (47.5 m)	Insufficient	Very low (only one study of inadequate quality (−3); imprecision, n<50 (−2))
	6MWT	Construct validity	Two out of four hypotheses confirmed	Sufficient	Very low (only one study of doubt quality (−2); inconsistency in results (−1) imprecision, n<100 (−1))
		Reliability	ICC: 0.885; sample size: 46	Sufficient	Very low (only one study of adequate quality (−1); imprecision, n<50 (−2))
		Measurement error	LoA of 92.2–104.5 m; greater than MCID 30 m	Insufficient	Very low (only one study of adequate quality (−1); imprecision, n<50 (−2))

4mGS: 4-m gait speed; 6MST: 6-min stepper test; 6MWT: 6-min walk test; 6PBRT: 6-min peg board ring test; BBS: Berg balance scale; DEMMI: de Morton mobility index; HGS: handgrip strength; ICC: intraclass correlation coefficient; ISWT: incremental shuttle walk test; LoA: limit of agreement; MCID: minimum clinically important difference; MIP: maximal inspiratory pressure.

0.44±0.16 bar (p<0.001) *versus* 0.55±0.16 bar *versus* 0.52±0.16 bar. The mean HGS was significantly lower in the ECOPD subjects than the stable COPD and non-COPD subjects; however, we did not find a cut-off point for this comparison in the literature. Thus, we rated construct validity as sufficient and downgraded the quality of evidence for inconsistency (tables 2 and 3).

Respiratory muscle strength

MIP: two studies assessed the validity and one study assessed responsiveness of MIP. Construct validity was assessed by Mesquita *et al*. [29] and Tudorache *et al*. [30]. They found a correlation between MIP and maximal expiratory muscle strength (rho=0.49 (p=0.04)), quadriceps peak torque (rho=0.57 (p=0.01)), 6MWT (rho=0.53 (p=0.0003)) and forced expiratory volume in 1 s (FEV_1) (rho=0.45 (p<0.01)). The hypotheses proposed by our group was a positive moderate correlation with MEP, quadriceps strength and 6MWT ⩾0.50 [40] and a positive small correlation with FEV_1 of 0.2–0.3. There were three results in line with the hypotheses. Responsiveness was assessed and a correlation with MEP of rho=0.58 (p=0.01) was found, in line with our hypothesis. Construct validity and responsiveness were rated as sufficient, with a very low quality of evidence (tables 2 and 3).

Upper limb function and exercise tolerance

6PBRT: one study assessed the validity of the 6PBRT [32]. Felisberto *et al*. [32] assessed the construct validity, with 16 established hypotheses for convergent validity and six results in line with the hypotheses (HGS: rho=0.70 (p=0.002); modified Pulmonary Functional Status Dyspnea Questionnaire (PFSDQ-M) dyspnoea: rho=−0.66 (p<0.001); PFSDQ-M fatigue: rho=−0.60 (p=0.01); PFSDQ-M change in activities of daily living: rho −0.51 (p=0.03); COPD Assessment Test: rho=−0.51 (p=0.03); fatigue upper limbs: rho=−0.76 (p<0.001)) and one result in line with the hypothesis to discriminative validity (the 6-PBRT did not correlate with height). The hypotheses to the known-group validity were defined by the authors of the present study, such as a difference higher than 24% in the performance of the 6-PBRT between ECOPD and healthy elderly patients [42], and the 6PBRT showed an adequate difference between the groups: 248.7±63.0 (number of rings moved) *versus* 361.6±49.9 number of rings moved (p<0.001). Other variables with similar construct to the 6-PBRT, namely elbow flexor torque peak, elbow extensor torque peak, total muscular work of the elbow flexor muscular, total muscular work of elbow extensor muscles and endurance of elbow flexors and extensors, showed a moderate correlation with the 6-PBRT, and the review authors considered these results to conclusion. Thus, although only eight out of 18 hypotheses were confirmed, we summarised the construct validity as sufficient and downgraded the evidence for inconsistency (tables 2 and 3).

Activities

Balance

BBS: one study assessed the validity of the BBS [33]. The construct validity was assessed by Oliveira *et al*. [33], although they did not have a clear purpose of assessing construct validity. A correlation with dyspnoea (modified Medical Research Council (mMRC)) and maximal isometric quadriceps muscle strength of rho=−0.33 (p>0.05) and rho=0.51 (p<0.05) were found, respectively, and a difference between patients with ECOPD *versus* healthy controls of 50.7±4.3 *versus* 55.2 ±1.4 (p<0.05). Our hypotheses were a positive low correlation with quadriceps strength ⩾0.26 [43] and with the mMRC ⩾0.26 [44] and a difference of 5–7 points between groups [45]. We found twos result consistent with the hypothesis and one result not consistent with the hypothesis. Thus, the construct validity was rated as sufficient with a very low quality of evidence (tables 2 and 3).

Mobility

DEMMI: one study assessed the validity of the DEMMI [34]. Camp *et al*. [34] assessed the construct validity of the DEMMI; specifically, convergent validity, known-group validity and discriminative validity. The convergent validity had a hypothesis of correlation of 0.60 with other functional measurements for the lower limb and we found two results in line with the hypotheses (6-min walk distance (6MWD): rho=0.69 (p=0.0006); gait velocity: rho=0.61 (p=0.0028)). The known-group validity assessed the difference between patients using a gait aid and not using a gait aid, without an established hypothesis, and found a score of 58.5±18.0 *versus* 79.5±16.2, respectively. Finally, regarding discriminative validity, there was no correlation between the DEMMI and resting heart rate and St George Respiratory Questionnaire total score. The overall rating was sufficient and evidence was graded as very low due to having only one study and its reduced sample size (tables 2 and 3).

Walking capacity

4mGS: one study assessed the validity of the 4mGS [35]. The construct validity was assessed by Nakano *et al*. [35], who found a convergent validity with the 6MWT and mMRC (rho=0.70 (p<0.0001) and

rho=0.68 (p<0.0001), respectively), without a defined hypothesis. Results were in line with the hypotheses of our review team with a positive high correlation of ⩾0.70 with the 6MWT [46] and positive moderate correlation of ⩾0.50 with the mMRC [47]. The overall rating was sufficient, and the evidence was graded as low due one study with adequate validity and reduced sample size (tables 2 and 3).

Body functions and activities

Exercise tolerance and walking capacity

ISWT: one study assessed the reliability of the ISWT [36]. Johnson-Warrington *et al*. [36] assessed the test–retest reliability of the ISWT using Bland–Altman analysis and found ISWT1 88.2±96.7 m and ISWT2 102.3±100.4 m, with a mean difference of 14.1±28.4 m. Limits of agreement (LoAs) were calculated from Bland–Altman analysis: −41.56–69.76 m. However, the LoAs were higher than the minimal clinically important difference (MCID) of 47.5 m [48] and the measurement error was rated as insufficient. Evidence was graded as very low due to including only one study of inadequate quality and the reduced sample size (tables 2 and 3).

6MWT: one study assessed validity [38], one study assessed reliability [39] and one study assessed responsiveness [37] of the 6MWT. Liao *et al*. [38] did not present a clear objective to assess validity; however, the authors tested the association between 6MWD and functionality measurements which were considered a construct validity. They found a positive correlation between 6MWD and maximal inspiratory strength, maximal expiratory strength, lower limb muscle strength and lower limb muscle endurance. However, considering a hypothesis of a positive moderate correlation of ⩾0.50 [40], we had two results in line with the hypotheses, in which one was for maximal inspiratory muscle strength (rho=0.54) and another for lower limb muscle endurance (rho=0.64), while maximal expiratory muscle strength had a correlation of 0.49. Regarding the known-groups analysed by ANOVA the review authors did not find hypotheses for the analysed groups. Considering that we had two of four hypotheses confirmed and the results were inconsistent, the review authors rated the summarised results of construct validity as "sufficient" and downgraded the quality of the evidence by inconsistency, according to COSMIN recommendations [18–20] (table S2). The test–retest reliability was assessed by Osadnik *et al*. [39], who found an $ICC_{2,1}$ of 0.885 (0.801–0.934) and measurement error using Bland–Altman analysis with an LoA of −92.2–104.5 m, higher than the MCID of 30 m [49]. Blankenburg *et al*. [37] investigated if the 6MWD would improve under effective exacerbation therapy and assessed responsiveness using the *t*-test. They found that 6MWD increased from 97±114 m to 290±190 m (p<0.05). Although there was an increase higher than an MCID of 30 m after treatment [49], the paired t-test is not considered appropriate [41] to analyse responsiveness and, therefore, the methodological quality was rated as inadequate. Thus, the overall rating of construct validity and reliability was sufficient and measurement error was insufficient with very low evidence (tables 2 and 3).

Feasibility and interpretability

We found few studies assessing feasibility and interpretability of the performance-based outcome measures included in the review (table S4). Only the 6PBRT and 6MST reported safety and did not present records of adverse events. All patients were able to complete the 6PBRT; however, 55.5% of patients needed to interrupt the test for a few seconds because of symptoms such as dyspnoea and fatigue [31], 15.8% were not able to perform the 6MWT due to mobility limitations [37], and all patients were able to perform the 6MST without interruptions [27].

Discussion

This study aimed to summarise the measurement properties of functionality performance-based outcome measures for hospitalised patients with ECOPD. Limited number of studies assessing measurement properties of the functionality performance-based outcome measures during ECOPD was found and the methodological quality must be classified as doubtful. Eight outcome measures (6MST, HGS, MIP, 6PBRT, BBS, DEMMI, 4mGS and 6MWT) were rated as sufficient to construct validity. Although the validity was sufficient, the low number and low methodological quality of studies resulted in a very low quality of evidence in almost all performance-based outcome measures. In terms of responsiveness and reliability, only one was tested in each property, the MIP and 6MWT, respectively and, two were assessed in terms of measurement error (ISWT and 6MWT) and rated as insufficient with a very low quality of evidence.

ECOPD are characterised by increased symptoms, particularly dyspnoea and/or cough and sputum, which may be accompanied by tachypnoea and/or tachycardia [1] and the low number of studies analysing measurement properties of functionality performance-based outcome measures in an inpatient scenario could be explained by the common assumption that these patients suffer from increased exercise

intolerance. However, whilst data on adverse effects are limited, it is important to report that feasibility studies included in this review did not report important adverse events.

It is important to point out that most functionality measures applied in inpatient rehabilitation [15, 50], as well as those found in this study, were adapted from stable patients [51, 52] with measurement properties already tested in this different context. However, measurement properties should consider the acute setting of patients with ECOPD to have a more accurate measure to optimise health outcomes through early individualised intervention.

Regarding walking capacity and exercise tolerance, we found a very low level of evidence for the insufficient measurement error of the 6MWT and ISWT, with an LoA higher than the MCID, possibly related to a learning effect [36, 39] and effect of symptoms [53]. Both tests generate physiological stress [54] and the ISWT may lead to a symptom-limited maximal performance [55]. In addition, no evidence of validity was found for the ISWT. Another performance-based outcome measure able to assess exercise tolerance is the 6MST, which presented sufficient validity with a very low level of evidence due to only one study found; however, it had a strong correlation with the 6MWT (rho=0.87 ($p<0.001$)) [27]. Thus, this study suggests that this test may be an alternative to the 6MWT for use in limited physical spaces. However, we should also consider balance and lower limb muscle strength to climb stairs when applying this measure. All these tests should be used to support the exercise prescription considering limitations, such as dyspnoea, heart rate, fatigue and arterial oxygen saturation.

Walking capacity may be assessed by 4mGS, which promotes lower physiological stress compared to the 6MWT and ISWT, can be performed in a short corridor, and, therefore, may be a more feasible option. We found that construct validity was rated as sufficient with low evidence in hospitalised patients with ECOPD, with a strong correlation observed with 6MWD (rho=0.70 ($p<0.0001$)). Gait speed may be a simple and easy tool for detecting poor and very poor 6MWT performance [46] and when associated with other measures, such as the Short Physical Performance Battery (SPPB), it may detect functional impairment in patients with stable COPD [56]. Previous studies with the SPPB involved older hospitalised patients [57]; however, it was not limited to ECOPD patients. Therefore, this battery of tests needs to be explored in this population.

In terms of peripheral muscle strength, sufficient construct validity with very low evidence was found for HGS. As shown previously, muscle strength is a relevant treatable trait in patients with ECOPD and lower HGS is related to an increased likelihood of death and poorer health-related quality of life in COPD patients [58]. In terms of lower limb function, another performance-based outcome measure option is the 6MST, as previously mentioned. An option to assess lower limb strength would be a sit-to-stand test (STS) [59], such as the five-repetition STS test [60] and the 30-min STS test [61]; however, no study was found that assessed the measurement properties of STS tests in hospitalised patients with ECOPD. Considering low-resource settings and feasibility, it would be interesting to further explore the STS test for inpatients. Upper limb function may be assessed by the 6PBRT, which showed sufficient construct validity with very low evidence. Although upper limb function is not commonly addressed, both in assessment and intervention, in patients with COPD, especially inpatients, it may affect activities of daily living [62]. Thus, patient participation in clinical decision-making is essential to properly choose the performance-based outcome measure and rehabilitation focus [63].

Respiratory muscle weakness may lead to increased dyspnoea [64] and inspiratory muscle training may improve exercise capacity and decrease dyspnoea, mainly in stable patients [65]. MIP is an easy-to-use measure of respiratory muscle strength; however, it also requires equipment (spirometer or manuvacuometer). We found sufficient construct validity and sufficient responsiveness with very low evidence for hospitalised patients with ECOPD. Regarding balance, patients with COPD may present with clinically meaningful impairment, which may be related to reduced muscle strength, physical activity and exercise capacity [43], and which may be worse during hospitalisation contributing to a high incidence of falls as an aggravation in the disease process [33].

Finally, patients with ECOPD are more physically inactive during hospitalisation and these low physical activity levels remain after discharge [9], which may be related to symptoms, using oxygen therapy and musculoskeletal and cardiovascular comorbidities. In this review, in the study of Blakenburg *et al.* [37], we found that 15.8% of patients were not able to perform the 6MWT due to being bed bound or being too weak. Thus, it is important to assess mobility during hospitalisation to set a rehabilitation goal. The DEMMI is an alternative tool to assess the performance of activities of daily living such as lying to sitting, sitting to standing from a chair, walking independence and picking up a pen from the floor, for example [66].

Clinical implications and considerations

Pulmonary rehabilitation (PR) is widely recognised as an important component in the clinical management of people with COPD. Exercise-based PR is able to improve health-related quality of life, exercise capacity and functionality [67, 68], and reduce mortality and readmissions [67]. Despite this, the body of evidence on its application in the acute phase of COPD is still developing. Regardless of the setting, to achieve an assertive and personalised diagnosis and rehabilitation management, it is crucial to apply performance-based outcome measures with good measurement properties. The choice of an instrument to assess the functionality should take into consideration the patients' needs, hospital space, need for equipment, time, validity and accuracy of measure, and professional experience [63]. This review is helpful for practitioners and researchers to point out the current knowledge on performance-based outcome measures. In terms of construct validity, it is important to consider measures rated as sufficient, as it means the ability of the instrument to validly measure exactly the construct to be measured [16]. This review found a low quality of evidence, suggesting that if these tests are applied, the results should be carefully interpreted, as further studies are needed to confirm them. However, the ISWT, with no evidence of validity, is not recommended for use in this specific population until further investigations have been conducted. Considering responsiveness as the "ability of an instrument to detect change over time in the construct to be measure" [16], the results point to a lack of evidence, making it difficult to apply the measures to analyse the effects of the intervention in this population.

In addition, there is no evidence of reliability and measurement error of functionality performance-based outcome measures, except for the 6MWT and ISWT, which were rated as insufficient. Since the reliability is the "ability of the instrument to distinguish between patients" while measurement error is a "systematic error of a patient that is not attributed to the true changes in the construct to be measure" [14], it is currently difficult to interpret results of functionality levels and state whether the changes in outcome are results of real or error changes. The lack of studies may involve factors specific to the inpatient population, such as rapid and dynamic changes, symptomatology, and patient demotivation to repeat the tests. Further studies are recommended and should consider training and standardising raters, avoiding time-of-day variation, averaging or repeated measurement, and reviewing the MCID for this specific population [16].

Limitations

This study has some limitations, as follows: the COSMIN checklist as a guideline for patient-reported measures was adapted to performance-based outcome measures; the CINAHL and Scopus databases were not used; although recommended, the lack of clear objective and analysis in some of the studies could have influenced the classification of quality, requiring development of hypotheses; there was difficulty in finding studies exclusively on ECOPD (inclusion of studies with greater than 50% of patients with ECOPD); and, finally, only studies in Portuguese, Spanish and English languages were included.

Conclusion

Measurement properties of performance-based outcome measures to assess functionality in hospitalised patients with ECOPD are still scarce, with a very low quality of evidence supporting validity and a lack of evidence of responsiveness and reliability. The construct validity was rated as sufficient in the 6MST, HGS, MIP, 6PBRT, BBS, DEMMI and 6MWT, with a very low quality of evidence, suggesting caution in interpreting results obtained by these measures. In addition, there is a lack of evidence of responsiveness, reliability and measurement error, making it difficult to use these measures to assess intervention effectiveness, discriminative assessment and outcome changes over time in this population. This study highlights the need for further studies addressing functionality instruments and their measurement properties in an inpatient setting with ECOPD to guide professionals in clinical decision-making more assertively.

Questions for future research

- What functional performance-based outcome measure, adequate in terms of measurement properties, should be assessed in hospitalised patients with ECOPD?
- Is clinical decision-making based on functionality performance-based outcome measures feasible in an acute setting with ECOPD?
- Is an individualised rehabilitation protocol based on adequate functional performance-based outcome measures able to improve outcomes following hospitalised ECOPD?

Provenance: Submitted article, peer reviewed.

Conflict of interest: N.T. Leonardi reports support for the present manuscript from FAPESP (São Paulo Research Foundation) in the form of English review and article processing charges, as well as support in attending meetings and travels to present products/results of the project. J.R. Hurst reports consulting fees from AstraZeneca, GSK; lecture honoraria from Boehringer Ingelheim, Chiesi, Sanofi, Takeda; travel support and advisory board participation from AstraZeneca; donation of oximeters from Nonin; outside the submitted work. J. Cruz reports travel support to attend the European Respiratory Society Congress 2022 as the representative of the Early Career Members of Assembly 9; and is a member of the Editorial Board of *ERJ Open Research* (unpaid); outside the submitted work. R.G. Mendes reports support for the present manuscript from FAPESP (São Paulo Research Foundation), in the form of project funding providing support for buying laboratory consumables materials and equipment and also article processing charges and maintenance costs, as well as attending meetings and travels to present products/results of the project. All other authors have nothing to disclose.

Support statement: This work was supported by Fundação de Apoio à Pesquisa do Estado de São Paulo (FAPESP) (numbers process: 2015/12763-4; 2015/26501-1 (Project grant); 2021/03493-4 (NTL scholarship)). The funding source had no involvement in this analysis or in the writing of this manuscript. Funding information for this article has been deposited with the Crossref Funder Registry.

References

1 Global Initiative for Chronic Obstructive Lung Disease. Global strategy for diagnosis, management, and prevention of chronic obstructive pulmonary disease: 2023 report. Date last accessed: 6 September 2022. Date last updated: July 2022. https://goldcopd.org/2023-gold-report-2/
2 Halpin DMG, Miravitlles M, Metzdorf N, *et al.* Impact and prevention of severe exacerbations of COPD: a review of the evidence. *Int J COPD* 2017; 12: 2891–2908.
3 Soler-Cataluña JJ, Martínez-García MÁ, Román Sánchez P, *et al.* Severe acute exacerbations and mortality in patients with chronic obstructive pulmonary disease. *Thorax* 2005; 60: 925–931.
4 Abdulai RM, Jensen TJ, Patel NR, *et al.* Deterioration of limb muscle function during acute exacerbation of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2018; 197: 433–449.
5 Torres-Sánchez I, Cabrera-Martos I, Díaz-Pelegrina A, *et al.* Physical and functional impairment during and after hospitalization in subjects with severe COPD exacerbation. *Respir Care* 2017; 62: 209–214.
6 Organização Panamericana da Saúde, Organização Mundial da Saúde. Classificação Internacional de Funcionalidade, Incapacidade e Saúde. Date last accessed: 6 September 2022. Date last updated: 2008. https://apps.who.int/iris/bitstream/handle/10665/42407/9788531407840_por.pdf
7 Machado A, Barusso M, De Brandt J, *et al.* Impact of acute exacerbations of COPD on patients' health status beyond pulmonary function: a scoping review. *Pulmonology* 2022; in press [https://doi.org/10.1016/j.pulmoe.2022.04.004]
8 Spruit MA, Gosselink R, Troosters T, *et al.* Muscle force during an acute exacerbation in hospitalised patients with COPD and its relationship with CXCL8 and IGF-I. *Thorax* 2003; 58: 752–756.
9 Pitta F, Troosters T, Probst VS, *et al.* Physical activity and hospitalization for exacerbation of COPD. *Chest* 2006; 129: 536–544.
10 Ibrahim W, Harvey-Dunstan TC, Greening NJ. Rehabilitation in chronic respiratory diseases: in-hospital and post-exacerbation pulmonary rehabilitation. *Respirology* 2019; 24: 889–898.
11 Hurst JR, Skolnik N, Hansen GJ, *et al.* Understanding the impact of chronic obstructive pulmonary disease exacerbations on patient health and quality of life. *Eur J Intern Med* 2020; 73: 1–6.
12 McDonald VM, Fingleton J, Agusti A, *et al.* Treatable traits: a new paradigm for 21st century management of chronic airway diseases: Treatable Traits Down Under International Workshop report. *Eur Respir J* 2019; 53; 1802058.
13 McDonald VM, Osadnik CR, Gibson PG. Treatable traits in acute exacerbations of chronic airway diseases. *Chron Respir Dis* 2019; 16: 1479973119867954.
14 Mokkink LB, Boers M, Van Der VC, *et al.* COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. *BMC Med Res Methodol* 2020; 20: 293.
15 Oliveira AL, Marques AS. Outcome measures used in pulmonary rehabilitation in patients with acute exacerbation of chronic obstructive pulmonary disease: a systematic review. *Phys Ther* 2018; 98: 191–204.
16 de Vet HCW, Terwee CB, Mokkink LB, *et al.* Measurement in Medicine. Cambridge, Cambridge University Press, 2011.
17 Page MJ, McKenzie JE, Bossuyt PM, *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021; 372: n71.
18 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: 1171–1179.

19 Prinsen CAC, Bouter LBMLM, Patrick JADL. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018; 27: 1147–1157.
20 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: 1159–1170.
21 Terwee CB, Jansma EP, Riphagen II, *et al.* Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009; 18: 1115–1123.
22 Bui KL, 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: 778–784.
23 D'Amore C, O'Hoski S, Griffith LE, *et al.* Factors associated with participation in life situations in people with COPD. *Chron Respir Dis* 2022; 19: 14799731221079305.
24 Oliveira CC, Lee A, Granger CL, *et al.* Postural control and fear of falling assessment in people with chronic obstructive pulmonary disease: a systematic review of instruments, international classification of functioning, disability and health linkage, and measurement properties. *Arch Phys Med Rehabil* 2013; 94: 1784–1799.e7.
25 Munro BH. Statistical Methods for Health Care Research. 4th Edn. Philadelphia, Lippincott Williams & Wilkins, 2001.
26 Global Initiative for Chronic Obstructive Lung Disease. Global Strategy for Prevention, Diagnosis and Management of COPD. Date last accessed: 4 February 2022. Date last updated: 22 November 2021. https://goldcopd.org/wp-content/uploads/2021/12/GOLD-REPORT-2022-v1.1-22Nov2021_WMV.pdf
27 Ribeiro DB, Terrazas AC, Yamaguti WP. The six-minute stepper test is valid to evaluate functional capacity in hospitalized patients with exacerbated COPD. *Front Physiol* 2022; 13: 853434.
28 Turan Z, Özyemişçi Taşkiran Ö, Erden Z, *et al.* Does hand grip strength decrease in chronic obstructive pulmonary disease exacerbation? A cross-sectional study. *Turkish J Med Sci* 2019; 49: 802–808.
29 Mesquita R, Donária L, Genz ICH, *et al.* Respiratory muscle strength during and after hospitalization for COPD exacerbation. *Respir Care* 2013; 58: 2142–2149.
30 Tudorache V, Oancea C, Mlădinescu OF. Clinical relevance of maximal inspiratory pressure: determination in COPD exacerbation. *Int J Chron Obstruct Pulmon Dis* 2010; 5: 119–123.
31 de Barros CF, Felisberto RM, Nucci KCA, *et al.* Dynamic hyperinflation induced by the 6-minute pegboard and ring test in hospitalized patients with exacerbated COPD. *PLoS One* 2020; 15: e0241639.
32 Felisberto RM, de Barros CF, Nucci KCA, *et al.* Is the 6-minute pegboard and ring test valid to evaluate upper limb function in hospitalized patients with acute exacerbation of COPD? *Int J COPD* 2018; 13: 1663–1673.
33 Oliveira CC, Lee AL, McGinley J, *et al.* Balance and falls in acute exacerbation of chronic obstructive pulmonary disease: a prospective study. *COPD* 2017; 14: 518–525.
34 Camp PG, Sima CA, Kirkham A, *et al.* The de Morton mobility index is a feasible and valid mobility assessment tool in hospitalized patients with an acute exacerbation of chronic obstructive pulmonary disease. *Chron Respir Dis* 2019; 16: 1479973119872979.
35 Nakano T, Kimura S, Yamashita T, *et al.* Correlation of 4-meter gait speed with clinical indicators of chronic obstructive pulmonary disease. *Respir Investig* 2021; 59: 505–512.
36 Johnson-Warrington V, Mitchell KE, Singh SJ. Is a practice incremental shuttle walk test needed for patients with chronic obstructive pulmonary disease admitted to hospital for an acute exacerbation? *Respiration* 2015; 90: 206–210.
37 Blankenburg T, Guettel A, Busch C, *et al.* Six-minute walk distance and dyspnoea scores to assess the course of COPD exacerbation in elderly patients. *Clin Respir J* 2013; 7: 261–267.
38 Liao LY, Chen KM, Hsu HF. Cardiopulmonary endurance of hospitalized older adults with chronic obstructive pulmonary disease. *Nurs Res* 2020; 69: E1–E8.
39 Osadnik CR, Borges RC, McDonald CF, *et al.* Two 6-minute walk tests are required during hospitalisation for acute exacerbation of COPD. *COPD J Chronic Obstr Pulm Dis* 2016; 13: 288–292.
40 Gosselink R, Troosters T, Decramer M. Peripheral muscle weakness contributes to exercise limitation in COPD. *Am J Respir Crit Care Med* 1996; 153: 976–980.
41 Mokkink LB, Prinsen CA, Patrick DL, *et al.* COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). Date last accessed: 19 May 2022. Date last updated: February 2018. www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf
42 Zhan S, Cerny FJ, Gibbons WJ, *et al.* Development of an unsupported arm exercise test in patients with chronic obstructive pulmonary disease. *J Cardiopulm Rehabil* 2006; 26: 180–187.
43 Loughran KJ, Atkinson G, Beauchamp MK, *et al.* Balance impairment in individuals with COPD: a systematic review with meta-analysis. *Thorax* 2020; 75: 539–546.
44 Kaygusuz MH, Oral Tapan O, Tapan U, *et al.* Balance impairment and cognitive dysfunction in patients with chronic obstructive pulmonary disease under 65 years. *Clin Respir J* 2022; 16: 200–207.
45 Beauchamp MK, Harrison SL, Goldstein RS, *et al.* Interpretability of change scores in measures of balance in people with COPD. *Chest* 2016; 149: 696–703.
46 Karpman C, DePew ZS, LeBrasseur NK, *et al.* Determinants of gait speed in COPD. *Chest* 2014; 146: 104–110.

47 Karpman C, Benzo R. Gait speed as a measure of functional status in COPD patients. *Int J COPD* 2014; 9: 1315–1320.
48 Singh SJ, Jones PW, Evans R, *et al.* Minimum clinically important improvement for the incremental shuttle walking test. *Thorax* 2008; 63: 775–777.
49 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: 1447–1478.
50 Machado A, Matos Silva P, Afreixo V, *et al.* Design of pulmonary rehabilitation programmes during acute exacerbations of COPD: a systematic review and network meta-analysis. *Eur Respir Rev* 2020; 29: 200039.
51 Bui KL, Nyberg A, Maltais F, *et al.* Functional tests in chronic obstructive pulmonary disease, Part 2: Measurement properties. *Ann Am Thorac Soc* 2017; 14: 785–794.
52 Johnston KN, Potter AJ, Phillips A. Measurement properties of short lower extremity functional exercise tests in people with chronic obstructive pulmonary disease: systematic review. *Phys Ther* 2017; 97: 926–943.
53 Lee J, Nguyen HQ, Jarrett ME, *et al.* Effect of symptoms on physical performance in COPD. *Heart Lung* 2018; 47: 149–156.
54 Vagaggini B, Taccola M, Severino S, *et al.* Shuttle walking test and 6-minute walking test induce a similar cardiorespiratory performance in patients recovering from an acute exacerbation of chronic obstructive pulmonary disease. *Respiration* 2003; 70: 579–584.
55 Singh SJ, Morgan MDL, Scott S, *et al.* Development of a shuttle walking test of disability in patients with chronic airways obstruction. *Thorax* 1992; 47: 1019–1024.
56 Patel MS, Mohan D, Andersson YM, *et al.* Phenotypic characteristics associated with reduced short physical performance battery score in COPD. *Chest* 2014; 145: 1016–1024.
57 Volpato S, Cavalieri M, Guerra G, *et al.* Performance-based functional assessment in older hospitalized patients: feasibility and clinical correlates. *J Gerontol A Biol Sci Med Sci* 2008; 63: 1393–1398.
58 Holden M, Fyfe M, Poulin C, *et al.* Handgrip strength in people with chronic obstructive pulmonary disease: a systematic review and meta-analysis. *Phys Ther* 2021; 101: pzab057.
59 Vaidya T, Chambellan A, de Bisschop C. Sit-to-stand tests for COPD: a literature review. *Respir Med* 2017; 128: 70–77.
60 Jones SE, Kon SSC, Canavan JL, *et al.* The five-repetition sit-to-stand test as a functional outcome measure in COPD. *Thorax* 2013; 68: 1015–1020.
61 Zanini A, Aiello M, Cherubino F, *et al.* The one repetition maximum test and the sit-to-stand test in the assessment of a specific pulmonary rehabilitation program on peripheral muscle strength in COPD patients. *Int J COPD* 2015; 10: 2423–2430.
62 Velloso M, Garcia Stella S, Cendon S, *et al.* Metabolic and ventilatory parameters of four activities of daily living accomplished with arms in COPD patients. *Chest* 2003; 123: 1047–1053.
63 Machado A, Almeida S, Burtin C, *et al.* Giving voice to people – experiences during mild to moderate acute exacerbations of COPD. *Chronic Obstr Pulm Dis* 2022; 9: 336–348.
64 Larson JL, Covey MK, Corbridge S. Inspiratory muscle strength in chronic obstructive pulmonary disease. *AACN Clin Issues* 2002; 13: 320–332.
65 Beaumont M, Forget P, Couturaud F, *et al.* Effects of inspiratory muscle training in COPD patients: a systematic review and meta-analysis. *Clin Respir J* 2018; 12: 2178–2188.
66 de Morton NA, Davidson M, Keating JL. The de Morton mobility index (DEMMI): an essential health index for an ageing world. *Health Qual Life Outcomes* 2008; 6: 63.
67 Ryrsø CK, Godtfredsen N, Kofod LM, *et al.* Lower mortality after early supervised pulmonary rehabilitation following COPD-exacerbations: a systematic review and meta-analysis. *BMC Pulm Med* 2018; 18: 154.
68 Zhang D, Zhang H, Li X, *et al.* Pulmonary rehabilitation programmes within three days of hospitalization for acute exacerbation of chronic obstructive pulmonary disease: a systematic review and meta-analysis. *Int J COPD* 2021; 16: 3525–3538.