Systematic Review

Psychometric properties of the sit-to-stand test for patients with pulmonary hypertension: A systematic review

Natália L. Cardoso [a] [iD], Joceline F. de Sá [b], Larissa F.E. do Nascimento [a] [iD], Luciana A. Mendes [c] [iD], Selma Bruno [b], Rodrigo Torres-Castro [d,e,f], Isabel Blanco [e,f], Guilherme A.F. Fregonezi [a] [iD], Vanessa R. Resqueti [a,*] [iD]

[a] *Laboratory of Technological Innovation in Rehabilitation and PneumoCardioVascular Lab/HUOL, Onofre Lopes University Hospital, Brazilian Company of Hospital Services (EBSERH), Universidade Federal do Rio Grande do Norte (UFRN), Natal, Rio Grande do Norte, Brazil*
[b] *Laboratory of Cardiorespiratory and Metabolic Assessment - CORE/HUOL, Onofre Lopes University Hospital, Brazilian Company of Hospital Services (EBSERH), Universidade Federal do Rio Grande do Norte (UFRN), Natal, Rio Grande do Norte, Brazil*
[c] *Department of Biomedical Engineering, Universidade Federal do Rio Grande do Norte (UFRN), Natal, Rio Grande do Norte, Brazil*
[d] *Department of Physical Therapy, Faculty of Medicine, University of Chile, Santiago, Chile*
[e] *Department of Pulmonary Medicine, Hospital Clínic–Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, Barcelona, Spain*
[f] *Biomedical Research Networking Center on Respiratory Diseases (CIBERES), Madrid, Spain*

ARTICLE INFO

ABSTRACT

Background: Pulmonary hypertension (PH) is defined as an increase in mean pulmonary arterial pressure, often accompanied by indicators such as dyspnea on exertion, exercise intolerance, and systemic muscle dysfunction. Various protocols exist that can indirectly assess these indicators through the sit-to-stand test (STST).
Objective: Assess the psychometric properties of different STST protocols in patients with PH.
Methods: This study is a systematic review. We searched the PubMed, EMBASE, SciELO, Cochrane Central Register of Controlled Trials (CENTRAL), and Web of Science databases. The risk of bias was assessed using the COSMIN tool and the certainty of evidence using the modified Recommendations, Assessment, Development, and Evaluation (GRADE) classification. Two investigators evaluated independently, and a third evaluator was consulted as needed.
Results: Out of a total of 7933 articles identified, only 5 articles met the criteria for inclusion in the analysis. Four psychometric properties were assessed across the five protocols used. The 1-STST protocol provided high-quality evidence for both convergent validity and responsiveness. The 30-STST protocol showed moderate-quality evidence for convergent validity and responsiveness, while the 5-STST also demonstrated moderate-quality evidence for responsiveness. The between-groups validity and reliability of the 30-STST protocol were considered to be low and very low, respectively.
Conclusion: Despite the limited number of studies, we can infer that the most commonly used protocol is the 1-STST, which has a high degree of convergent validity and responsiveness when compared to other assessment tools.

Introduction

Pulmonary hypertension (PH) is defined by increased mean pulmonary arterial pressure (mPAP). This health condition considers clinical, pathophysiological, and hemodynamic factors. It is also categorized into five groups, namely pulmonary arterial hypertension (PAH), PH associated with left heart disease, PH associated with lung disease, PH associated with pulmonary artery obstruction, and PH with multifactorial mechanisms.[1] It is regarded as a life-threatening condition, and it is typically associated with progressive deterioration of function and increased mortality. Moreover, it affects approximately 1 % of the global population.[2]

The principal and earliest symptom is exercise-related dyspnea, which is progressive and reflects the inability of the cardiovascular system to increase cardiac output during exercise.[1] Other symptoms may be present, including fatigue, presyncope, chest pain, and ankle edema,

which may develop from right ventricular failure.[1,3] At the muscular level, these individuals tend to suffer from systemic muscle dysfunction due to an increased risk of functional decline as a result of loss of muscle function, decreased trophism, and an increase in type II fibers. Additionally, proteins regulating mitochondrial fusion in skeletal muscle are expressed at lower levels and associated with exercise intolerance compared to healthy individuals.[4] As observed through pathophysiological mechanisms in high-intensity exercise such as cardiopulmonary exercise testing (CPET),[5] more feasible tests such as the submaximal sit-to-stand test (STST) are useful and practical tools which provide indirect and non-invasive evaluations of the lower limb skeletal muscle strength[6] and exercise tolerance.[7,8] There are various protocols such as the 5-repetitions STST (5-STST), the 30-second STST (30-STST), and the 1-minute STST (1-STST).

The 5-STST protocol is a cost-effective tool,[9] reliable for the older population,[10] and for assessing lower limb strength and balance control in healthy individuals and adults with diseases.[11] Moreover, it is a marker of the low functional performance of patients with chronic obstructive pulmonary disease (COPD).[12,13] The 30-STST is considered a safe and feasible method for assessing lower limb function and strength[14] like the poor exercise tolerance in patients with COPD.[15] The 1-STST is regarded as a reliable, valid, and responsive alternative to measure physical capacity in healthy people and for some pathologies. It is correlated with the 6-minute walk test (6MWT), making it an ideal alternative for use in the office,[16,17] and was recently shown of detecting exertional desaturation in patients post COVID-19.[18]

Understanding the psychometric properties and evaluation methodologies of exercise tolerance assessments in PH is crucial for obtaining valid measures.[19,20] This approach ensures the selection of reliable, safe, and reproducible tests for patients, supporting clinical investigation and intervention applicability. Despite its clinical importance in patients with PH, there are few systematic reviews on the STST. This systematic review is crucial for guiding scientific decisions and developing optimal STST protocols for PH and its subtypes. Consequently, we aimed to identify the available evidence on the STST when used for patients with different forms of PH and present the psychometric properties and the number of repetitions achieved with different STST protocols.

Methods

This systematic review was conducted conformed to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).[21] This review also followed the PICOT mnemonic strategy, with the research question, "What are the psychometric properties of the different sit-to-stand test measures in individuals with pulmonary hypertension?". The risk of bias was assessed using the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN).[22,23] The protocol was registered in the International Prospective Register of Systematic Reviews (PROSPERO, no. CRD42021244271), and detailed methods have been previously published.[24]

Eligibility criteria

Randomized or quasi-randomized controlled clinical trials were included, as well as observational studies published in English. We considered studies conducted in inpatient, outpatient, or primary care settings.

This review included studies of adult participants with a clinical diagnosis of one of the five types of PH.[1] Exclusion criteria were: systematic reviews, in vitro studies, conference abstracts, theses, dissertations, literature reviews, studies carried out with children or mixed populations, and studies in which the STST protocol did not meet the criteria described by Kahraman et al.[25] Furthermore, studies were excluded if the data from patients with PH were not analyzed separately or could not be extracted or obtained even after contacting the authors.

Type of intervention

Included studies were those that assessed exercise tolerance using the STST (30-STST, 1-STST, or 5-STST) following the methodology proposed by Kahraman et al.,[25] as to perform as many sit-to-stand cycles as possible from a standard chair without arm support, with arms crossed at the wrists against the chest, starting on the command "go" to full standing and back to the seated position.

Type of outcome measures

Primary results

Psychometric properties (validity [criterion validity, construct validity], reliability [internal consistency and measurement error], and responsiveness) and repeatability achieved in STST protocols by participants with PH.

Secondary results

- Association between STST and patient-reported outcome measures (PROMs);
- Symptoms reported during STST protocols using the Borg scale;
- Changes in the respiratory or cardiovascular system response to STST protocols before and after the intervention.

Search strategy

A comprehensive search of the PubMed/MEDLINE, EMBASE, SciELO, Cochrane Central Register of Controlled Trials (CENTRAL), and Web of Science databases was conducted in April 2023, with no defined publication time restriction. The search strategy for all the databases was performed with keywords such as pulmonary hypertension OR pulmonary arterial hypertension AND sit-to-stand test OR exercise tolerance (Supplemental online material).

Data collection and analysis

Two researchers (NC-LN) reviewed the selected studies that were entered into the Rayyan software[26] and independently analyzed the titles and abstracts and removed irrelevant studies. A third researcher (JS) assessed any discrepancies as necessary. The full texts of the eligible studies were attached to the Rayyan software and analyzed by the same researchers (NC-LN). In case of conflict, the third researcher (JS) was consulted. The reasons for exclusion were recorded, and the screening process was summarized in a PRISMA flowchart (Fig. 1).

Data extraction and management

Data on measurement properties were extracted using the COSMIN[23] data extraction form, which qualitatively summarizes the results of studies in checklist boxes, based on general observations, data such as intraclass correlation coefficients (ICC), or those required to determine that measurement property. For secondary outcomes, a previously tested word extraction form was used. All extractions were performed by two researchers (NC-JS), and the third evaluator (VR) intervened to resolve discrepancies.

Risk of bias (RoB) assessment

Two investigators (NC-JS) independently analyzed the RoB using the COSMIN.[27,28] Disagreements were resolved by a third investigator (VR). Each consistent result of the individual study was summarized qualitatively and compared against criteria for good measurement properties, interpreted as very good, adequate, doubtful, or inadequate.[28,29] The results of each study were classified according to specific criteria and updated for each measurement property as follows: sufficient (+),
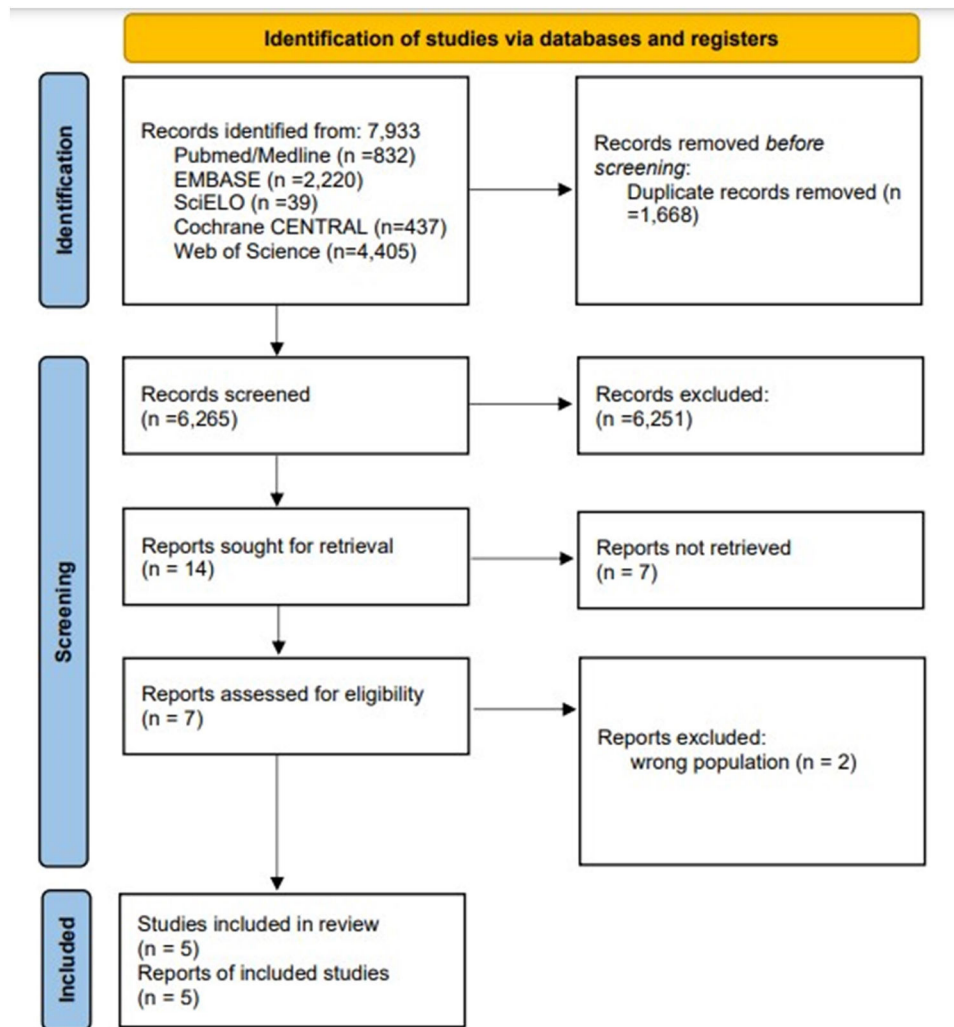
Fig. 1. Flow diagram - PRISMA 2020.

insufficient (-), or indeterminate (?).[2]

Evidence certainty assessment

The Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) rating was employed to assess the certainty of the evidence. The evidence was classified as high, moderate, low, or very low,[27] based on four domains: RoB, which evaluates the methodological quality of studies; inconsistency, which verifies unexplained inconsistency of results across studies; imprecision, which analyzes the total sample size of available studies and; indirection, which evaluates evidence from populations other than the population of interest in the review.[30] This evaluation was conducted by two independent evaluators (NC-JS), with the involvement of a third evaluator (VR) in case of discrepancies.

Data synthesis

We used Odds Ratio, Relative Risk, or Risk Difference for dichotomous data analysis for continuous variables and to carry out a quantitative synthesis (meta-analysis), the software RevMan V.5.3.528. In the only case of missing data or ongoing clinical trials, the authors were contacted via email. The objective was to conduct a subgroup analysis, by type of PH or by STST protocols. If necessary, we would perform sensitivity analyses to examine methodological quality's effects on the pooled estimate by removing studies classified as having a high RoB.

Results

A total of 7933 articles were selected. Of these, 1668 were excluded due to duplication, leaving 6265 titles and abstracts. Of these, 6251 were excluded. The remaining 14 articles proceeded on to the full-text search stage, during which seven were excluded. Two were conference abstracts, two by wrong publication, two were clinical trials still in progress, and ultimately, an article could not be accessed in its entirety. The final two papers excluded, were studies not on the population of interest (Fig. 1).

The five studies included[25,31-34] were conducted in diverse geographical locations, with a predominance of the 1-STST protocol, high expressiveness of group 1, PAH, and female subjects in the studies. However, the majority of studies included a low number of individuals in the sample. There was one RCT,[32] and four observational studies,[25,31,33,34] included. The summary of the characteristics of the included studies populations are presented in Table 1.

Four studies assessed convergent validity (Table 2).[25,31,33,34] Kahraman et al.[25] demonstrated a moderate correlation with the strength of the knee extensor muscles. With regard to age and the New York Heart Association (NYHA) classification, moderate negative correlations were found. Moreover, all correlations were statistically significant. In the study by Nakazato et al.[31] moderate correlation was observed between the frequency reached in the test with the accelerometer for the number of steps and the time achieved, and $p < 0.05$. Kronberger et al.[33] found a

Table 1

Population characteristics of included studies.

Authors	Population			Disease characteristic Disease	Test	Country
	n	Age	% Feminine gender			
González-Saiz et al[32]	40	IG = 46 ± 11 CC= 45 ± 12	60 %	PAH or CTEPH	5-STST	Spain/ USA
Kahraman et al[25]	38	50.3 ± 18.0	81 %	PAH or CTEPH	30-STST	Turkey
Nakazato et al[31]	20	44.3 ± 13.2	80 %	IPAH or PAH—CTD	1-STST	Brazil
Kronberger et al[33]	106	66 ± 15	57 %	All PH	1-STST	Austria
Keen et al[34]	75	52 ± 16.8	77 %	PAH or CTEPH	1-STST	UK

n, numbers of individuals; IG, intervention group; CG, control group; PAH, pulmonary arterial hypertension; CTEPH, chronic thromboembolic pulmonary hypertension; IPAH, idiopathic pulmonary arterial hypertension; PAH—CD, PAH associated with connective tissue disease; All PH, all groups pulmonary hypertension; 5- STST, 5-repetition sit-to-stand test; 30- STST, 30-second sit-to-stand test; 1- STST, one minute sit-to-stand test.

high correlation with the Borg dyspnea score, while Keen et al.[34] showed a moderate correlation with WHO Functional Class (WHO-FC), both with p < 0.001.

Kahraman et al.[25] investigated the validity between known groups, comparing the number of repetitions achieved in the test and the functional class and found in FC II and FC III (Table 2). Reliability was only assessed in the study by Kahraman et al.[25] in the test-retest, demonstrating an excellent ICC.

Responsiveness was divided into responsiveness before and after training (assessed only by Gonzalez-Saiz et al.[32]) and the responsiveness between the STST and other tests like the 6MWT[25,33] or incremental shuttle walking test (ISWT)[34] (Table 2). Gonzalez-Saiz et al.[32] demonstrated a positive change in the intervention group after eight weeks of exercise, with p < 0.001. Kahraman et al.[25] and Kronberger et al.[33] observed a moderate and high correlation between STST protocols *versus* 6MWT, respectively, and both were statistically significant. With regard to the ISWT, only the study by Keen et al. exhibited a strong correlation and p < 0.001.

Concerning the repetitions performed during the test, variability was observed due to the diversity of protocols. Kahraman et al.,[25] in their 30-STST study, reached the result of the reliability of the first evaluation of 12.23 ± 3.77 repetitions and the second evaluation of 12.07 ± 3.87 repetitions. González-Saiz et al.[32] with the 5-STST protocol, it was seen that the group that initially performed the intervention lasted 7.5 ± 1.4 s and post-intervention managed to make it in 6.0 ± 1.1 s (p < 0.001). The studies that employed the 1-STST protocol reported the following results, Nakazato et al.[31] achieved 23.8 ± 6.1 repetitions, Kronberger et al.[33] reported 17 ± 7 repetitions, and Keen et al.[34] recorded 20.1 repetitions.

Table 3 presents the results of the RoB assessment for the studies included in this review. The studies that evaluated convergent validity were rated as very good, as they met all the criteria determined by COSMIN.[23] Responsiveness was assessed in the studies by Gonzalez-Saiz et al.,[32] Kahraman et al.,[25] Kronberger et al.,[33] and Keen et al.[34] The reliability of the study by Kahraman et al.[25] was deemed inadequate due to the evaluator being aware of the previous measurement during the second test. Moreover, the validity between known groups was considered doubtful because the author did not define the characteristics of

Table 2

Characterization of the psychometric properties of the studies.

Types of tests	Authors	Tests	Psychometric properties evaluated	Results
STST	González-Saiz et al[32]	5-STST	Responsiveness	Before and after 8 weeks **IG:** 5- STST (performance time, s): 7.5 ± 1.4 to 6.0 ± 1.1 (p < 0.001) **CG:** 5- STST (performance tempo, s): 7.0 ± 1.6 to 6.9 ± 1.4 (p > 0.05)
	Kahraman et al[25]	30-STST	Test-rest reliability/ Convergent Validity/ Validity between groups/ Responsiveness	Test-retest reliability 30- STST (number): first evaluation 12.23 ± 3.77 / second evaluation 12.07 ± 3.87 / Difference = 0.15 ± 1.19 ICC (95 % CI) 0.95 (0.90, 0.97) Convergent Validity 30- STST Age r=−0.61, p < 0.001; NYHA-FC r= −0.45, p = 0.004; knee extensor muscle strength r = 0.54, p < 0.001; IPAQ-SF r = 0.37, p = 0.02. Validity between groups: Comparison p = 0.004 30- STST (number) and NYHA Class II 13.68 (3.34); 30- STST (number) versus NYHA Class III 10.25 (3.49). Responsiveness (comparison with other outcome measurement instruments) 6MWT r = 0.66, p < 0.001.
	Nakazato et al[31]	1-STST	Convergent Validity	1- STST was 23.8 ± 6.1 Accelerometer: steps per day, n. 4280.2 ± 2351.7 Activity time (min) 41.6 ± 19.3 Convergent Validity 1- STST Accelerometer number of steps per day r = 0.59, p = 0.006 Uptime accelerometer (min) r = 0.58, p = 0.007
	Kronberger et al[33]	1-STST	Convergent Validity/ Responsiveness	1- STST was 17 ± 7 Borg:5.0 ± 2.3 Convergent Validity 1- STST WHO-FC r = −0.59, p < 0.001 NT-proBNP r = −0.40, p < 0.001 mPAP r = −0.28, p < 0.001 BDS r = 0.70, p < 0.001. Responsiveness

Table 2 (*continued*)

Types of tests	Authors	Tests	Psychometric properties evaluated	Results
	Keen et al[34]	1-STST	Convergent Validity/ Responsiveness	(comparison with other outcome measurement instruments) 6MWD r = 0.71, p < 0.001 1-STST was 20.1 Borg: 3.6 ± 1.8 Convergent Validity 1- STST WHO-FC r =−0.50, p < 0.001 NT-proBNP r= −0.26, p = 0.02 emPHasis10 r= −0.43, p < 0.001 Age r= −0.39, p < 0.001 Responsiveness (comparison with other outcome measurement instruments) ISWT r = 0.70, p < 0.001

5- STST, 5-repetition sit-to-stand test; 30-STST, 30-second sit-to-stand test; 1-STST, one minute sit-to-stand test; IG, intervention group; CG, control group; NYHA-FC, New York Heart Association - functional class; ICC, intraclass correlation coefficient; IPAQ-SF, international physical activity questionnaire; 6MWT, 6-minute walk test; WHO-FC, WHO functional class; NT-proBNP, natriuretic peptide; mPAP, mean pulmonary arterial pressure; BDS, Borg dyspnea score; emPHasis10, quality of life questionnaire in Phasis10; ISWT, Incremental Shuttle Walking Test.

each subgroup studied. The COSMIN evaluation defines the updated criteria for good measurement properties, which can be seen in Table 3. The majority of the study results were determined to be sufficient, as they align with the proposal of each measurement property by Mokkink et al.[27] except for Kronberger et al.[33] on convergent validity, and Keen et al.,[34] on convergent validity and responsiveness. Both studies lack the hypothesis of their work.

The certainty of the evidence for the psychometric properties of each study is presented in Table 4, which was evaluated based on the four modified GRADE criteria.[29] The 30-STST and 1-STST studies that assessed the convergent validity are considered of moderate and high quality, respectively. The certainty of the evidence of the first one was downgraded due to imprecision as there is only one study conducted with a low number of included individuals. In terms of evaluating the validity between known groups, the 30-STST was classified as low certainty rating due to imprecision small sample size, and risk of bias (only one study of doubtful quality is available). Still, for the 30-STST, the reliability was classified as very low certainty of evidence due to imprecision (low number of individuals involved) and risk of bias (only one study of inappropriate quality is available). Responsiveness in the 5-STST was classified as moderate certainty of evidence, due to imprecision as there is only one study conducted with a low number of included individuals. As well as the responsiveness when compared to other tests, with the 30-STST, which also presented a moderate certainty of evidence for the same reason. Finally, the 1-STST had high certainty of evidence, as it fulfilled the necessary criteria.

Regarding the secondary outcomes, dyspnea was assessed using the modified Borg scale (0–10), with final results as follows: Nakazato et al.[31] reported 4.5 ± 1.5, Kronberger et al.[33] reported 5.0 ± 2.3, and Keen et al.[34] reported 3.6 ± 1.8. Cardiovascular alterations were reported in the study by Kronberger et al. ,[33] which demonstrated an increase in systolic BP of 12 ± 20 mmHg, diastolic BP of 2.3 ± 10 mmHg, HR of 21 ± 16 beats/min, and a decrease in SpO2 of 4.6 ± 5.9 %.

Table 3
Results of measurement properties risk of bias assessments and updated criteria for good measurement properties according to COSMIN.

Author/age test	Reliability (test-retest)			Convergent validity			Validity between groups			Responsiveness			Responsiveness Before and after)		
	n	Qualification	Result	n	Qualification	Result	n	Qualification	Result	n	Qualification	Result	n	Qualification	Result
González-Saiz et al., 2017 5- STST													40	Very good	+
Summary results 5- STST													40	Very good	+
Kahraman et al., 2020 30- STST	38	inappropriate	+	38	Very good	+	38	doubtful	+	38	Very good	+			
Summary results 30- STST	38	inappropriate	+	38	Very good	+	38	doubtful	+	38	Very good	+			
Nakazato et al. 2021 1- STST				20	Very good	+									
Kronberger et al., 2023 1- STST				106	Very good	?				106	Very good	+			
Keen et al., 2023 1- STST				75	Very good	?				75	Very good	?			
Summary results 1- STST				201	Very good	?				181	Very good	?			

5- STST, 5-repetition sit-to-stand test; 30- STST, 30-second sit-to-stand test; 1- STST, one minute sit-to-stand test; ICC, intraclass correlation coefficient; "+", sufficient; "?", indeterminate.

Table 4
Certainty of evidence by modified GRADE.

Properties	Summarized or grouped results	General evaluation	Certainty of evidence
Convergent validity 30- STST	**Very good** – 30- STST vs age r= −0.61 p < 0.001; 30- STST vs NYHA r= −0.45 p = 0.004; 30- STST vs knee extensor muscle strength r = 0.54 p < 0.001; 30- STST vs IPAQ-SF r = 0.37 p = 0.02 (Kahraman et al., 2020).	sufficient	**Moderate**[1] duo to imprecision
1- STST	**Very good** 1- STST vs Accelerometer number of steps per day r = 0.59 p = 0.006; 1- STST vs Accelerometer uptime (min) r = 0.58 p = 0.007 (Nakazato et al., 2021); 1- STST vs WHO-FC r = −0.59 p < 0.001; 1- STST vs NT-proBNP r = −0.40 p < 0.001; 1- STST vs mPAP r = −0.28 p < 0.001; 1- STST vs BDS r = 0.70 p < 0.001 (Kronberger et al., 2023); 1- STST vs WHO-FC r=−0.50 p < 0.001; 1- STST vs NT-proBNP r= −0.26 p = 0.02; 1- STST vs emPHasis10 r= −0.43 p < 0.001; 1- STST vs age r= −0.39 p < 0.001(Keen et al., 2023).	indeterminate	**High**
Validity between groups 30- STST	**Doubtful** 30- STST (repetitions) and NYHA Class II 13.68 (3.34) versus NYHA Class III 10.25 (3.49); p = 0.004 (Kahraman et al., 2020).	sufficient	**Low**[1,2] duo to imprecision, risk of bias
Reliability (test-retest) 30- STST	**Inappropriate** ICC 0.95(0.90–0.97) (Kahraman et al., 2020).	sufficient	**Very low**[1,3] duo to imprecision, risk of bias
Responsiveness (Before and After) 5- STST	**Very good** 5- STST (performance time, s): 7.5 ± 1.4 to 6.0 ± 1.1 (p < 0.001) (González-Saiz et al., 2017).	sufficient	**Moderate**[1] duo to imprecision
Responsiveness (Other outcome measurement instrument) 30- STST	**Very good** 30- STST vs 6MWT r = 0.66 p < 0.001 (Kahraman et al., 2020).	sufficient	**Moderate**[1] duo to imprecision
1- STST	**Very good** 1- STST vs 6MWD r = 0.71, p < 0.001. (Kronberger et al., 2023); 1-STST	indeterminate	**High**

Table 4 (*continued*)

Properties	Summarized or grouped results	General evaluation	Certainty of evidence
	vs ISWT r = 0.70 p < 0.001 (Keen et al., 2023).		

STST, sit-to-stand test; 5- STST, 5-repetition sit-to-stand test; 30- STST, 30-second sit-to-stand test; 1- STST, one minute sit-to-stand test; NYHA, New York Heart Association; IPAQ-SF, international physical activity questionnaire; WHO-FC, WHO functional class; NT-proBNP, natriuretic peptide; emPHasis10, quality of life questionnaire in Phasis10; ICC, intraclass correlation coefficient; 6MWT, 6-minute walk test; ISWT, Incremental Shuttle Walking Test.

GRADE Working group of evidence.

High: Further research is very unlikely to change our confidence in the estimate of effect.

Moderate: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.

Low: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

Very low: We are very uncertain about the estimate.

[1] The evidence was downgraded due to its imprecision, as the studies included a limited number of participants.

[2] Considered very serious RoB, as there is only one study on the psychometric property evaluated, which is of dubious quality.

[3] Considered extremely serious RoB, as it is the only article with this psychometric property and presents inadequate quality.

Similarly, the study by Keen et al.[34] also showed an increase in systolic BP of 10.1 ± 10.5 mmHg, diastolic BP of 2.9 ± 7.8 mmHg, HR of 9.4 ± 8.0 beats/min, and a decrease in SpO2 of 3.8 ± 4.0 %. The PROMs were not reported in any included studies. We did not find sufficient results to justify conducting a meta-analysis.

Discussion

This systematic review aimed to determine the psychometric properties of different STST protocols in patients with PH. The primary findings indicate that the hypothesis test for construct validity, particularly in terms of convergent validity, is confirmed for the 30-STST and especially for the 1-STST. Moreover, the psychometric property of responsiveness is validated for the 5-STST, 30-STST, and 1-STST in patients with PH.

Bowman et al.[14] found a moderate correlation (rho = 0.49, p = 0.006) between the IPAQ-SF and the 30-STST protocol in the oncology population, confirming this review and suggesting good convergent validity between populations. Nakazato et al.[31] also found a moderate correlation between the 1-STST and accelerometer measurements in daily activity. Thus far, literature has not shown studies relating the STST with accelerometers, only correlations with the 6MWT have been observed. Cho et al.[35] demonstrated a moderate and significant correlation between step count and the 6MWT in individuals with sarcoidosis. Consequently, the use of test protocols in various populations remains limited, which restricts comparisons with our findings.

In the present study, moderate negative correlations were observed in Kronberger et al.[33] between 1-STST and WHO-FC, NT-proBNP, and weak correlations with mPAP. In Keen et al.,[34] negative correlations were noted, with moderate associations found with WHO-FC and emPHasis10, and weaker correlations observed with age and NT-proBNP. These data are frequently cited as prognostic factors for risk classification for PH.[1] Li et al.[36] showed that medication use improved patient classification based on WHO-FC; however, the effects on NT-proBNP were not observed. In the systematic review by Fu et al.[37] on new drugs for PH, improvements were observed in WHO-FC and mPAP. Regarding physical training, the review by An et al.[38] suggests that exercise improves mPAP. However, no study has demonstrated effects linking the STST with the variables identified in the present study.

Validity between groups was explored with a single PH.[25] The

30-STST protocol appears sufficiently sensitive to discern differences between NYHA classifications among patients with PH.[39] Similarly, Tarrant et al.[40] found significant differences using the 1-STST in post-operative and medical readmissions. Larger studies with robust statistical methods, including ROC curve analysis, are necessary for confirmation.

Reliability was assessed in one study[25] in this review, indicating excellent reliability but low-quality evidence. Figueiredo et al.[41] reported high reliability (ICC 0.93, 95 % CI: 0.86, 0.96) of the STST in hemodialysis patients. Bohannon et al.[16] systematic review also showed high reliability across diverse populations: ICC ranges from 0.80–0.98 in hemodialysis,[42] older adults,[43] and in cystic fibrosis.[44] Mong et al.[45] found the 5-STST highly reliable (ICC 0.98–0.99) in stroke patients. While suggesting excellent reliability in these groups, more studies are needed to validate reliability in PH patients.

In the present study, responsiveness was evaluated in two ways, the test's responsiveness before and after some type of training and responsiveness when compared to other instruments. The first analysis utilized data from a single article.[32] In this study, the authors assessed the responsiveness of the STST following an 8-week intervention involving aerobic, resistance, and inspiratory muscle training in patients with PH. They observed a reduction in test time, consistent with findings in the literature using a similar protocol. For instance, Augustín et al.[46] demonstrated responsiveness of the test in patients with stroke, highlighting how the severity and stage of recovery can impact functional outcomes. Zampogna et al.[47] demonstrated that both asthma and COPD populations experienced improvements in the 5-STST time following a pulmonary rehabilitation program. While the 5-STST shows promise as a tool for evaluating functional capacity response to intervention programs, further studies are needed to establish STST protocols in clinical practice for patients with PH and to explore additional properties of the test. Three studies were identified that assessed the responsiveness of the STST compared to other measurement instruments.[25,33,34] It appears that the 30-STST protocol, and particularly the 1-STST, shows good responsiveness to other measurement instruments. According to Kimberlin and Winterstein[39] it is crucial to evaluate the validity of an instrument by comparing it with a gold standard or other existing measures that assess the same construct.

This review compares three protocols measuring repetitions across specific populations. Gephine et al.[48] supported this review's reported 24 ± 5 repetitions for the 1-STST in COPD. In the 5-STST, PH patients[31] outperformed Zampogna et al.[47] findings in asthma and COPD (15.7 [IQR 12.7–17.3] seconds and 14.6 [IQR 12.1–16.6] seconds, respectively). Figueiredo et al.[41] found hemodialysis patients achieved 12.6 (11.8–13.4) repetitions in the 30-STST, comparable to PH patients.[25] Despite limited data, STST in PH suggests reliable benchmarks similar to other populations.

The secondary outcomes addressed in the included studies were: dyspnea and cardiovascular changes. On dyspnea, all three studies[31,33,34] reported final values "mild" measured by the Borg scale of 0–10. Corroborating with Ozalevli et al.,[49] in patients with COPD who reported "mild, moderate". Briand et al.,[50] in addition to agreeing with the aforementioned studies, also demonstrate that all three subgroups of patients studied had a significant difference in the most severe dyspnea, reported by the Borg scale, at the conclusion of the 1-STST in comparison to that observed during the 6MWT.

Limitations and clinical practice usefulness

A limitation of this systematic review study is the relatively small sample sizes in some of the included articles, and, the methodological quality of the included studies was found to be variable.

This study emphasizes using validated STST protocols in clinical practice for PH patients. Validity confirmation of 30-STST and 1-STST suggests accurate measurement of lower limb strength and function. Also, the responsiveness across the 5-STST, 30-STST, and 1-STST,

highlights their utility in detecting changes in physical function and endurance over time. This is crucial for monitoring disease progression, evaluating treatment efficacy, and guiding rehabilitation strategies.

Conclusion

Based on our systematic review, the 5-STST shows responsiveness to treatment. Moreover, both the 30-STST and 1-STST protocols exhibit valid convergent validity and excellent responsiveness when compared to other measurement instruments in the context of PH. Despite limited studies with appropriate methodologies, STST protocols are relevant for this population, and further comparisons are needed to enhance the evidence on measurement properties.

Declaration of competing interest

The authors declare no competing interest.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.bjpt.2025.101223.

References

1. Humbert M, Kovacs G, Hoeper MM, Badagliacca R, Berger RMF, Brida M, et al. ESC/ERS scientific document group, 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. *Eur Respir J*. 2023 Apr 17;44(15):1312. https://doi.org/10.1093/eurheartj/ehac237.
2. Hoeper MM, Humbert M, Souza R, Idrees M, Kawut SM, Sliwa-Hahnle K, et al. A global view of pulmonary hypertension. *Lancet Respir Med*. 2016 Apr;4(4):306–322.
3. Rose-Jones LJ, McLaughlin VV. Pulmonary hypertension: types and treatments. *Curr Cardiol Rev*. 2015;11:73–79.
4. Batt J, Ahmed SS, Correa J, Bain A, Granton J. Skeletal muscle dysfunction in idiopathic pulmonary arterial hypertension. *Am J Respir Cell Mol Biol*. 2014;v50:74–86.
5. Sun XG, Hansen JE, Oudiz RJ, Wasserman K. Exercise pathophysiology in patients with primary pulmonary hypertension. *Circulation*. 2001;104:429–435.
6. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee IM, et al. American college of sports medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Med Sci Sports Exerc*. 2011;43(7):1334–1359.
7. Jones SE, Kon SS, Canavan JL, Patel MS, Clark AL, Nolan CM, et al. The five-repetition sit-to-stand test as a functional outcome measure in COPD. *Thorax*. 2013;68(11):1015–1020.
8. Morita AA, Bisca GW, Machado FVC, Hernandes NA, Pitta F, Probst VS. Best protocol for the sit-to-stand test in subjects with COPD. *Respir Care*. 2018 Aug;63(8):1040–1049.
9. Bohannon RW. Test-retest reliability of the five-repetition sit-to-stand test: a systemic review of the literature involving adults. *J Strength Cond Res*. 2011;25(11):3205–3207.
10. Melo TA, Duarte ACM, Bezerra TS, França F, Soares NS, Brito D. Teste de Sentar-Levantar Cinco Vezes: segurança e confiabilidade em pacientes idosos na alta da unidade de terapia intensiva. *Rev Bras Ter Intensiva*. 2019;31(1):27–33.
11. Muñoz-Bermejo L, Adsuar JC, Mendoza-Muñoz M, Barrios-Fernández S, Garcia-Gordillo MA, Pérez-Gómez J, Carlos-Vivas J. Test-retest reliability of Five Times Sit to Stand Test (FTSST) in Adults: a systematic review and meta-analysis. *Biology (Basel)*. 2021 Jun 9;10(6):510. https://doi.org/10.3390/biology10060510.
12. Medina-Mirapeix F, Bernabeu-Mora R, Valera-Novella E, Gacto-Sánchez M, Bernabeu-Mora M, SánchezMartínez MP. The five-repetition sit-to-stand test is a predictive factor of severe exacerbations in COPD. *Ther Adv Chronic Dis*. 2021 Jan 22;12, 2040622320986718.
13. Bernabeu-Mora R, Medina-Mirapeix F, Llamazares-Herrán E, Oliveira-Sousa SL, Sánchez-Martinez MP, Escolar-Reina P. The accuracy with which the 5 times sit-to-stand test, versus gait speed, can identify poor exercise tolerance in patients with COPD: a cross-sectional study. *Medicine*. 2016 Aug;95(35):e4740.
14. Bowman A, Denehy L, Benjemaa A, Crowe J, Bruns E, Hall T, et al. Feasibility and safety of the 30s sit-to-stand test delivered via telehealth: an observational study. *PM R*. 2022 Feb 9. https://doi.org/10.1002/pmrj.12783.
15. Zhang Q, Li YX, Li XL, Yin Y, Li RL, Qiao X, et al. A comparative study of the five-repetition sit-to-stand test and the 30-second sit-to-stand test to assess exercise tolerance in COPD patients. *Int J Chron Obstruct Pulmon Dis*. 2018 Sep 10;13:2833–2839.
16. Bohannon RW, Crouch R. 1-minute sit-to-stand test systematic review. *J Cardiopulm Rehabil Prev*. 2018;00:1–7. https://doi.org/10.1097/HCR.0000000000000336.

17. Sevillano-Castaño A, Peroy-Badal R, Torres-Castro R, Gimeno-Santos E, Fernández PG, Vila CG, et al. Is there a learning effect on 1-min sit-to-stand test in post-COVID-19 patients? *ERJ Open Res*. Jul 2022;8(3):00189–02022.

18. Núñez-Cortés R, Rivera-Lillo G, Arias-Campoverde M, Soto-García D, García-Palomera R, Torres-Castro R. Use of sit-to-stand test to assess the physical capacity and exertional desaturation in patients post COVID-19. *Chron Respir Dis*. 2021 Jan-Dec;18, 1479973121999205.

19. Souza AC, Alexandre NMC, Guirardello EB. Propriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade. *Epidemiol Serv Saude. Brasília*. jul-set 2017;26(3):649–659.

20. Palese A, Navone E, Danielis M, Vryonides S, Sermeus W, Papastavrou E. Measurement tools used to assess unfinished nursing care: a systematic review of psychometric properties. *J Adv Nurs*. 2020;00:1–18.

21. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4(1):1.

22. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63(7):737–745.

23. Mokkink LB, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, Vet HCW, Terwee CB. COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). *User manual*; 2018. Available from: www.cosmin.nl.

24. Cardoso NL, de Sá JF, do Nascimento LFE, Mendes LA, Bruno S, Torres-Castro R, et al. Psychometric properties of the sit-to-stand test for patients with pulmonary hypertension: a systematic review protocol. *PLoS One*. 2022 Oct 5;17(10), e0275646.

25. Kahraman BO, Ozsoy I, Akdeniz B, Ozpelit E, Sevinc C, Acar S, et al. Test-retest Reliability and validity of the timed up and go test and 30-second sit to stand test in patients with pulmonary hypertension. *Int J Cardiol*. 2020 Apr 1;304:159–163.

26. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan — A web and mobile app for systematic reviews. *Syst Rev*. 2016;5:210. https://doi.org/10.1186/s13643-016-0384-4.

27. Mokkink LB, Boers M, van der Vleuten CPM, Bouter LM, Alonso J, Patrick DL, 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).

28. Mokkink LB, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, de Vet HCW, Terwee CB. COSMIN Study Design checklist for Patient-reported outcome measurement instruments. *Version*; July 2019. Available from: www.cosmin.nl.

29. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147–1157.

30. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol*. 2010 Mar 18;10:22.

31. Nakazato L., Mendes F., Paschoal I.A., Oliveira D.C., Moreira M.M., Pereira M.C. Association of daily physical activity with psychosocial aspects and functional capacity in patients with pulmonary arterial hypertension: a cross-sectional study. Pulm Circ. 2021 Mar 29;11(2):2045894021999955.

32. González-Saiz L, Fiuza-Luces C, Sanchis-Gomar F, Santos-Lozano A, Quezada-Loaiza CA, Flox-Camacho A, et al. Benefits of skeletal-muscle exercise training in pulmonary arterial hypertension: the WHOLEi+12 trial. *Int J Cardiol*. 2017 Mar 15; 231:277–283.

33. Kronberger C, Mousavi RA, Öztürk B, Willixhofer R, Dachs TM, Rettl R, et al. Functional capacity testing in patients with pulmonary hypertension (PH) using the one-minute sit-to-stand test (1-min STST). *PLoS One*. 2023 Mar 9;18(3), e0282697. https://doi.org/10.1371/journal.pone.0282697.

34. Keen C, Smith I, Hashmi-Greenwood M, Sage K, Kiely DG. Pulmonary Hypertension and Measurement of Exercise Capacity Remotely: evaluation of the 1-min Sit-to-Stand Test (PERSPIRE) - a cohort study. *ERJ Open Res*. 2023 Jan 23;9(1): 00295–02022.

35. Cho PSP, Vasudevan S, Maddocks M, Spinou A, Chamberlain Mitchell S, Wood C, et al. Physical inactivity in pulmonary sarcoidosis. *Lung*. 2019 Jun;197(3):285–293. Epub 2019 Mar 19.

36. Li J, Yang ZY, Wang S, Yuan P, Zhao QH, Gong SG, et al. Efficacy and safety of switching from bosentan or ambrisentan to macitentan in pulmonary arterial hypertension: a systematic review and meta-analysis. *Front Cardiovasc Med*. 2022 Dec 7;9, 977110.

37. Fu W, He W, Li Y, Chen Y, Liang J, Lei H, et al. Efficacy and safety of novel-targeted drugs in the treatment of pulmonary arterial hypertension: a Bayesian network meta-analysis. *Drug Deliv*. 2021 Dec;28(1):1007–1019.

38. An QY, Wang L, Yuan P, Zhao QH, Gong SG, Zhang R, et al. Effectiveness and safety of exercise training and rehabilitation in chronic thromboembolic pulmonary hypertension: a systematic review and meta-analysis. *Ann Palliat Med*. 2021 Jul;10 (7):8134–8146.

39. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm*. 2008 Dec;65(23):2276–2284.

40. Tarrant BJ, Robinson R, Le Maitre C, Poulsen M, Corbett M, Snell G, et al. The utility of the sit-to-stand test for inpatients in the acute hospital setting after lung transplantation. *Phys Ther*. 2020 Jul 19;100(7):1217–1228.

41. Figueiredo PHS, Veloso LRS, Lima MMO, Vieira CFD, Alves FL, Lacerda ACR, et al. The reliability and validity of the 30-seconds sit-to-stand test and its capacity for assessment of the functional status of hemodialysis patients. *J Bodyw Mov Ther*. 2021 Jul;27:157–164.

42. Segura-Ortí E, Martinez-Olmos FJ. Test-retest reliability and minimal detectable change scores sit-to-stand to-sit tests, the six-minute walk test, the one-leg heel-rise test, and ongoing hemodialysis. *Phys Ther*. 2011;91:1244–1252. https://doi.org/10.2522/ptj.20100141.

43. Ritchie C, Trost SG, Brown W, Armit C. Reliability and validity of physical fitness tests for adults aged 55 to 70 years. *J Sci Med Sport*. 2005;8:61–70.

44. Vaidya T, de Bisschop C, Beaumont M, Ouksel H, Jean V, Dessables F, Chambellan A. Is the 1-minute sit-to-stand test a good tool for the evaluation of impact of pulmonary rehabilitation? Determination of the minimal important difference in COPD. *Int J Chron Obstruct Pulmon Dis*. 2016;11:2609–2616.

45. Mong Y, Teo TW, Ng SS. 5-repetition sit-to-stand test in subjects with chronic stroke: reliability and validity. *Arch Phys Med Rehabil*. 2010 Mar;91(3):407–413.

46. Augustín RMS, Crisostomo MJ, Sánchez-Martínez MP, Medina-Mirapeix. Responsiveness and minimal clinically important difference of the five times sit-to-stand test in patients with stroke. *Int. J. Environ. Res. Public Health*. 2021;18(5):2314. https://doi.org/10.3390/ijerph18052314.

47. Zampogna E, Pignatti P, Ambrosino N, Cherubino F, Maria Fadda A, Zappa M, et al. The 5-repetition sit-to-stand test as an outcome measure for pulmonary rehabilitation in subjects with asthma. *Respir Care*. 2021 May;66(5):769–776.

48. Gephine S, Bergeron S, Tremblay Labrecque PF, Mucci P, Saey D, Maltais F. Cardiorespiratory response during the 1-min sit-to-stand test in chronic obstructive pulmonary disease. *Med Sci Sports Exerc*. 2020 Jul;52(7):1441–1448.

49. Ozalevli S, Ozden A, Itil O, Akkoclu A. Comparison of the sit-to-stand test with 6 min walk test in patients with chronic obstructive pulmonary disease. *Respir Med*. 2007 Feb;101(2):286–293.

50. Briand J, Behal H, Chenivesse C, Wémeau-Stervinou L, Wallaert B. The 1-minute sit-to-stand test to detect exercise-induced oxygen desaturation in patients with interstitial lung disease. *Ther Adv Respir Dis*. 2018 JanDec;12, 1753466618793028. https://doi.org/10.1177/1753466618793028.