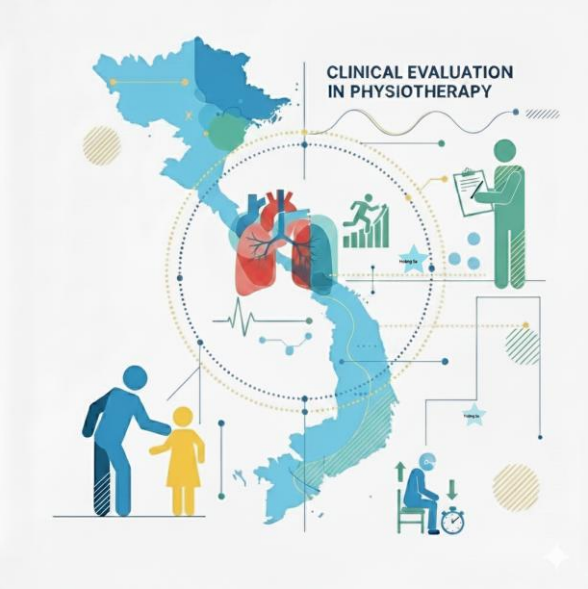
Adapting Cardiopulmonary and Pediatric Functional Assessments for Physiotherapy in Vietnam

NGUYEN NGOC MINH

August, 2025

Thèse présentée en vue de l'obtention du grade de docteur en sciences de la motricité

Secteur des sciences de la santé



Member of the Jury

Promoter of the thesis

Prof. Dr. Grégory Reychler

Institut de Recherche Experimental et Clinique (IREC), Pneumology, ENT and Dermatology Pole (LUNS), Université Catholique de Louvain (UCLouvain), Brussels, Belgium

President

Prof. Dr. William Poncin

Institut de Recherche Experimental et Clinique (IREC), Pneumology, ENT and Dermatology Pole (LUNS), Université Catholique de Louvain (UCLouvain), Brussels, Belgium

Members:

Prof. Dr. Giuseppe Liistro

Institut de Recherche Experimental et Clinique (IREC), Pneumology, ENT and Dermatology Pole (LUNS), Université Catholique de Louvain (UCLouvain), Brussels, Belgium

Dr. Nicolas Audag

Institut de Recherche Experimental et Clinique (IREC), Pneumology, ENT and Dermatology Pole (LUNS), Université Catholique de Louvain (UCLouvain), Brussels, Belgium

Invited experts :

Prof. Dr. Gael Deboeck

Research Unit in Rehabilitation Sciences, Faculty of Human Movement Sciences, Université Libre de Bruxelles (ULB), Brussels, Belgium

Prof. Dr. Olivier Contal

University of Applied Sciences & Arts Western Switzerland - School of Health Sciences Geneva

Acknowledgement :

Too many people to thank during this journey. I will try my best. I cannot mention all of you, but I appreciate all the people I met. Without any of you, my life would be very different.

Simply say: I won the lottery many times

I won the lottery of family: thanks to my wonderful wife and my amazing son for always being my rocks and always encouraging me. To my wife, thank you for your incredible sacrifice; without your unwavering support and boundless love, this thesis would have taken me forever. Thanks to my 2-year-old son for his…spirited…attempts to delay my graduation. I have to say, good try, boy.

Thanks to my parents for everything, who had to stop their education very early due to the harsh realities of life at a young age, but you still encouraged your son to finish a PhD. You don't even know what a PhD is; you only care whether I am happy or not, that's the purest love that exists. I won the "parent lottery".

Thanks to my professional "father", Prof. Gregory Reychler, for allowing me to pursue my ideas, no matter how silly they seemed. You made my PhD journey feel like a scientific exploration and pure enjoyment. Standing on your shoulder is my privilege; a vantage point offering a broader perspective on the scientific world. YOU RAISE ME. I won the Promotor lottery.

Thanks to Prof. Yevs Castile and Patrick Willems, Ms. Nguyen Anh Chi, and Ms. Le Thanh Van. Without you, there would be no ARES project. Thanks for your advocacy, dedication, kindness, and willingness. Thanks to Anne, your care makes my stay in Belgium much less difficult. I won the Life Mentor lottery.

Thanks to Bernard and Dominique, it was difficult to finish a PhD while homeless. Thank you for giving me not only a house but a home in Belgium. Your kindness and generosity will always be remembered. I won the adopted parent lottery.

Thanks to Marjorie and other administrative officers, without your help with the administration work and travelling, this PhD journey must be much harsher.

Thanks to my committee and the jury for orienting my journey in the right direction.

Thanks to my teachers and my students for supporting and being contributors to my academic growth.

I appreciate Belgium, a beautiful country known for its incredible educational philosophy. This country serves as a foundation not only for enhanced understanding but also for fostering kindness. Thanks to the Académie de recherche et d'enseignement supérieur (ARES), without the development funding for the project in Vietnam, nothing would have happened.

Finally, thanks to everyone I met in my life. “Every person you meet is the person you should meet.”

Louvain-la-Neuve, August 2025

Nguyen Ngoc Minh

INTRODUCTION

Measurement plays a crucial role in physiotherapy practice by optimizing the effectiveness of interventions [1]. Clinical measurements also aid in prognosis, enhancing intervention planning, and leading to more tailored and effective treatment strategies. However, when applying a measurement in a new context, culture, or population, it is essential to investigate its validity and reliability to ensure accurate and meaningful results [2]. Furthermore, the need for standardized equipment can hinder the accessibility and implementation of clinical measurements, particularly in middle- and low-income countries where resources may be limited [3]. The lack of, or infrequent measurement, reduces the effectiveness of interventions, which increases healthcare costs and diminishes societal well-being, creating a cycle of disease burden and risk factors.

In Vietnam, clinical evaluation and measurement in physiotherapy are not yet routinely performed due to the lack of appropriate assessment tools. Several contributing factors to the unavailability of physiotherapy assessments are the limited research capacity of physiotherapists and the restricted accessibility of equipment. This presents a significant barrier to advancing physiotherapy practice and improving patient outcomes [1]. Therefore, accessible and validated measurement tools are important. Thus, this project was conducted to validate assessment tools, specifically in the context of a middle-income country, which is crucial to improving clinical evaluation and measurement.

However, the extreme diversity in medical assessment properties and their definitions has led to inconsistencies and incomparable research findings across different tools, consequently impairing the conclusions of appropriate and well-designed existing studies [4]. For instance, inconsistent terminology for concepts like reliability, which

can also be referred to as reproducibility, stability, repeatability, and precision, complicates the interpretation and synthesis of research results. Therefore, to acquire reliable results while investigating the assessment instruments' properties, the Consensus-based Standards for the selection of health Measurement Instruments initiative (COSMIN) has to be followed [2]. It provides methodological guidance for the development, selection, and evaluation of health measurement instruments, which helps to minimize the risk of bias. The three important properties of a medical measurement are validity, reliability, and responsiveness [2]. Validity is the property that indicates the capacity of a measurement to accurately measure what it is intended to measure. Reliability is the capacity to demonstrate the same result under the same conditions at two different times, allowing clinicians to recognize changes in patients' conditions. Changes in disease or health status over time are important, that is responsiveness characteristic of a measurement. Adhering to COSMIN standards ensures that the validation process is comprehensive and rigorous.

Furthermore, not all assessments are direct and straightforward. For instance, assessing quality of life presents its own set of challenges due to its subjective nature. Similarly, dyspnea or hyperventilation syndrome can be ambiguous and vague to measure, presenting a considerable challenge for precise quantification in clinical settings. In such situations, questionnaires are typically used to capture the subjective perspective. Utilizing the COSMIN taxonomy provides a systematic approach to assess and adapt these subjective instruments for the Vietnamese context, thereby enhancing both their validity and cross-cultural relevance.

To make assessments more accessible and available for Vietnamese physiotherapists and to expand the scientific understanding of clinical evaluation, this thesis conducted studies in three main areas: respiratory muscle function, functional exercise capacity, and the translation and validation of questionnaires commonly used in

physiotherapy assessment. Each assessment area was selected to address specific needs and gaps in the current physiotherapy practice in Vietnam. The studies of this thesis adhered to the COSMIN initiative. The thesis comprises four chapters:

Chapter 1: Respiratory Muscle Assessment and Training
Chapter 2: Functional Exercise Capacity Assessment for persons with heart failure
Chapter 3: Translation, Validation, and Reliability Investigation of Questionnaires in Vietnamese.
Chapter 4: General discussion and future studies

Chapter 1: Respiratory Muscle Assessment and Training

1. Respiratory muscle training: the supportive evidence

The respiratory system requires harmony between the lungs, the mechanical pump, and the central control to maintain optimal function. Any impairment in any of these systems can deteriorate respiratory function and reduce health status and well-being. Specifically, respiratory muscle weakness is common in individuals with chronic respiratory diseases and can significantly impact their quality of life. Recently, respiratory muscle training, particularly inspiratory muscle training (IMT), has emerged as a promising intervention in many conditions to enhance respiratory muscle strength and endurance, resulting in improved functional capacity and reduced dyspnea. To contribute to the implementation of respiratory muscle training, an investigation of the effectiveness of IMT for different populations is necessary. The thesis conducted two systematic reviews to evaluate the effectiveness of IMT for specific populations, including persons with lung cancer after surgery and persons with heart failure. Another aim of this project was to determine the optimal IMT program, identifying the most appropriate intensity and duration for each population studied.

The effect of inspiratory muscle training in patients with lung cancer after surgery: A systematic review.

Ngoc-Minh Nguyen; Fabien Latiers; Frank Aboubakar Nana; Valérie Lacroix; Gregory Reychler
Adapted from: *Rehabilitation Oncology,p 202 – 212, October 2023*
DOI: 10.1097/01.REO.0000000000000352

Abstract

Objective: The standard of care treatment for potentially resectable lung cancer (LC) is surgery. However, postoperative pulmonary complications (PPCs) and impairments in physical capacity are common. Recently, the effect of inspiratory muscle training (IMT) in postoperative patients with LC (PWLC) was investigated in these outcomes in different studies. The purpose of this systematic review was to synthesize the effect of postoperative IMT (P-IMT) on PPCs and physical capacity in PWLC.

Database: PubMed, EMBASE, Cochrane, and CINAHL were searched.

Study Selection: Randomized controlled trials, including control/sham group, IMT as the intervention group, and key measures including PPCs, 6-minute walk test (6MWT), V_{O2peak}, maximal inspiratory pressure (MIP), maximal expiratory pressure (MEP), quality of life (QoL), physical activity level, hospital length of stay, spirometry.

Data Synthesis: The quality of the studies was assessed using the Physiotherapy Evidence Database (PEDro) scale. The outcome findings were compared and interpreted.

Results: Five studies with 249 patients were analyzed. The PEDro scores of studies ranged from 6 to 8. There is no significant difference in PPCs between the groups. The effects of P-IMT on the 6MWT, V_{O2Peak}, MIP, MEP, and QoL were ambiguous. No effect of P-IMT on spirometry was reported.

Conclusion: No effect of P-IMT in PPCs was reported. The effect of P-IMT on physical capacity and respiratory muscle strength was not observed. The level of the effect of P-IMT on hospital length of stay and postoperative physical activity was low. No optimal setting of P-IMT for PWLC with surgery was found. More studies are needed.

Key words: Inspiratory muscle training, Lung cancer, postoperative pulmonary complications.

Introduction

Lung cancer (LC) is one of the leading causes of death in the world, with more than 1.8 million deaths [5]. Recently, there were improvements in LC management and minimally invasive surgical approaches, with a reduction in mortality [6-8]. Unfortunately, the postoperative pulmonary complications (PPCs) rate is between 29.9% and 36.6% in the elderly population, depending on the type of operation (segmentectomy, lobectomy, or pneumonectomy) [9,10]. The decrease in physical function in survivor PWLC was also reported up to 2 years after surgery[11]. Investigating the potential effect of IMT on physical capacity is important due to the inverse correlation between physical capacity and the risk of death in PWLC[12,13]. One focal point of PWLC surgery was respiratory muscle weakness [14,15]. Indeed, the inspiratory muscle weakness induces impairment of cough, chest compliance, and physical capacity, and is associated with PPCs [16-20]. To address these postoperative issues, inspiratory muscle training (IMT) has emerged [21].

Several systematic reviews with meta-analysis showed the effectiveness of preoperative IMT in PWLC, but the evidence of postoperative IMT (P-IMT) was scattered [22,23]. Moreover, the guidelines for dosing of P-IMT, such as intensity, frequency, and time, are not yet understood.

This systematic review was conducted, first, to summarize the effects of P-IMT on PPCs, physical capacity, maximal inspiratory pressure (MIP), maximal expiratory pressure (MEP), Quality of life (QoL),

physical activity level, hospital length of stay, and lung function after LC surgery. Secondly, the settings of the programs of P-IMT were also summarized in PWLC after surgery.

Method

The preferred reporting items for systematic review and meta-analysis guidelines (PRISMA) were used to conduct this review [24].

Literature search

Four electronic databases were used for searching: PubMed, EMBASE, Cochrane Library, and CINAHL. The equation of searching strategy in PubMed was based on PICO criteria and as follows: ("inspiratory muscle training"Title/Abstract) OR (IMT(Title/Abstract) OR ("respiratory muscle training"(Title/Abstract)) AND (lung* AND (carcinogen* OR sarcom* OR tumor* OR tumour* OR carcinoma* OR cancer* OR neoplasm*)) AND (“postoperative” OR “post-surgery”). The equation was adapted to the other databases. The searching and screening were conducted independently by two reviewers (N-M. N. and G. R.). There was no publication date limitation for the search.

Study selection

Studies were selected based on the following inclusion criteria: 1) the main purpose of studies was to investigate the effect of IMT in the postoperative lung cancer population; 2) randomized controlled trial including sham/control group; 3) outcomes including at least one of following measures: PPCs, 6MWT, MIP, MEP, QoL questionnaires, physical activity level, hospital length of stay, and lung function.

Exclusion criteria were: 1) non-English language, 2) studies including preoperative IMT intervention, and 3) inclusion of diseases other than lung cancer.

After removing duplicated studies with EndNote [25], two independent reviewers (N-M. N., G. R.) screened all potential studies by title and abstract. Then, a full test reading was performed with the selected

studies. The final list of included studies was discussed and confirmed by consensus.

Quality assessment

The quality of studies was checked by the online Physiotherapy Evidence Database (PEDro) website and then assessed again by one reviewer (N-M N). The PEDro scale includes 11 items with a score of zero or one [26]. Item 1 (eligible criteria) is not included in the cumulative score; thus, the maximal score was 10. The higher the score, the better the internal validity. These items satisfied the core set of items for RCTs' quality assessment [27]. The levels of quality were divided into three groups: good (>6), fair (4-5), and poor (≤3) [28].

Data extraction

The extracted data from all included studies were: 1) characteristics of demographic data (sample size, age, BMI, smoking, MIP); 2) type of surgery; 3) intervention program of study and control group (method, duration, starting point); 4) outcomes (PPCs, 6MWT, VO_{2peak}, MIP, MEP, QoL, hospital length of stay, spirometry). Data extraction was conducted independently by two reviewers. The final table of data was cross-checked.

Result

Study selection

After a comprehensive literature search, there was a total of 201 studies met the inclusion criteria (Figure 1). After Title/abstract screening, 27 studies were assessed by full-text reading. Five articles were included in the final systematic review (Figure 1).

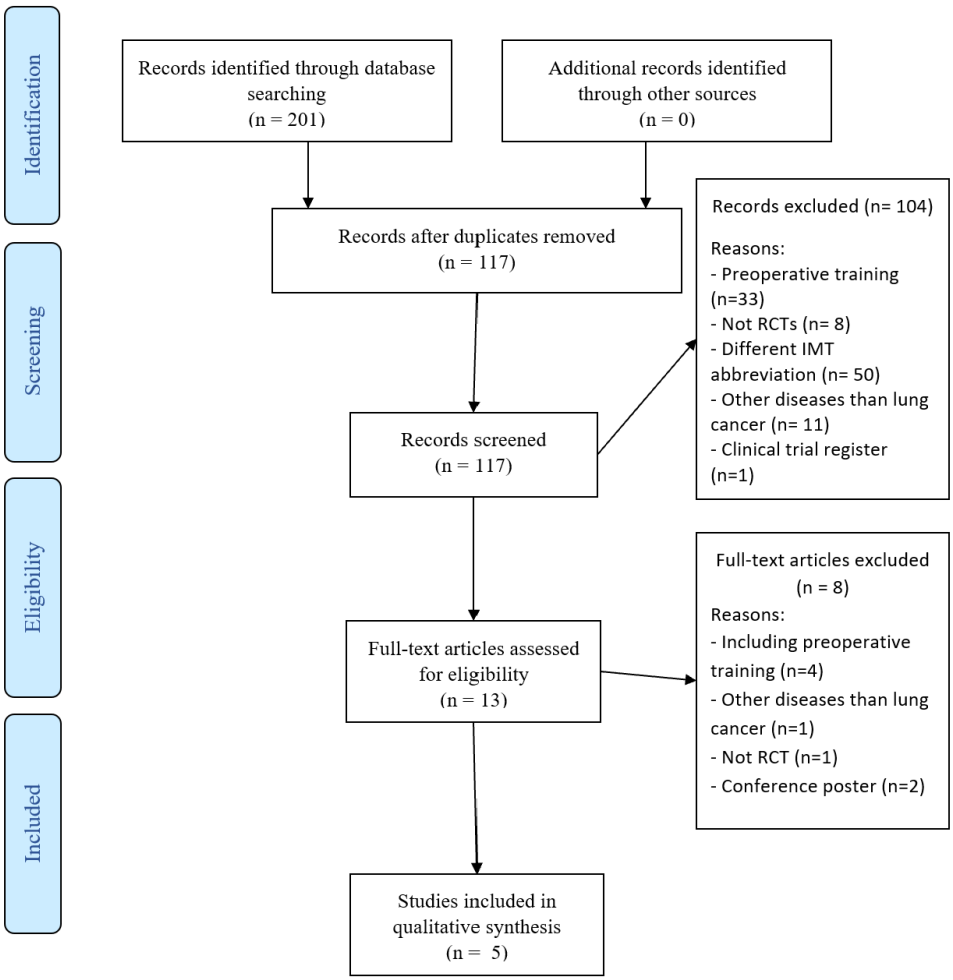


Figure 1. A Flow chart of the search strategy

Study description and quality

The five RCTs included 276 patients with a confirmed diagnosis of LC. All patients underwent surgery. The types of surgery included open surgery (134 participants) and video-assisted thoracic surgery (VATS) (142 participants). There was a dropout of 27 participants in three studies, thus, 249 participants were analyzed after training. One study included only PWLC with VATS [29], the others combined both types of surgery, with no significant difference in the open-VATS surgery ratio between the control group (CG) and intervention group (IG) [30-33]. The median age of the participants was 64.2 (70.5; 53.3). Body mass index (BMI) was in the normal or overweight range [34]. The smoking rate was above 60% in the four studies reporting this rate [31]. There was no significant difference in demographic characteristics between IG and CG in all studies (P>0.05) (Table 1).

The median PEDro score was 7. All the scores were between 8 and 6.

Intervention description

All the patients from IG received P-IMT combined with a rehabilitation program. The rehabilitation programs were composed of aerobic training, breathing exercise, positive expiratory pressure, early mobilization, airway clearance techniques, strength limb exercise, physical activity advice, and respiratory muscle training (RMT). The duration of training was from 2 weeks to 8 weeks [30-32]. In one study, the training program was finished when patients were discharged without a fixed duration [33]. One study started the training 4 days after surgery, and it continued until 6 weeks after discharge [29]. Three studies started the training 1-to-2-day after surgery without mentioning the tube removal [30,31,33], and one study started at 4-6 weeks post-surgery [32].

The control groups received the same rehabilitation program in three studies [29-31,33]. One CG received only physical activity advice. Two studies used Powerbreath, two others used Threshold, and one

study combined inspiratory and expiratory muscle training (IEMT) with an Oxygen dual valve. The intensity of P-IMT varied from 15% to 50% of the maximal inspiratory pressure, and the training frequency ranged from 100 to 120 repetitions a day and 3 to 5 days a week (Table 2).

Table 1. Characteristics of studies

First author (year)	Design	Sample size	Mean age (Years)	Gender (Female/Male)	BMI	Smoking rate	Pre-op MIP (cmH_2O)	Surgery		PEDro score
								VATS	Thoracotomy	
Brocki et al. (2015) [30]	IG: IMT + Rehab	34	69.7 ± 7.9	14/20	25.4 ± 4.3	33/34	82.8 ± 26.8	59%	41%	8
	CG: rehab	34	70.5 ± 7.5	15/19	28.0 ± 5.7	30/34	78.5 ± 29.5	44%	56%	
Brocki et al. (2018)[31]	IG: IMT + Rehab	34	70 ± 8	14/20	26 ± 4	-	83 ± 27	59%	41%	7
	CG: rehab	32	70 ± 8	13/19	28 ± 6	-	78 ± 29	44%	56%	
Mesaggi-Sartor et al. (2018)[32]	IG: IEMT + aerobic	16	64.2 ± 8.1	8/8	28.1 ± 4.6	13/16	69.4 ± 25.0	12.5%	87.5%	6
	CG: Health advice	21	64.6 ± 8.9	3/18	26.8 ± 3.1	20/21	75.4 ± 30.4	4.8%	95.2%	
Taşkin et al. (2020) [33]	IG: IMT + CT + early mobilization	20	53.3 ± 10.4	5/15	26.5 ± 4.0	17/20	65.1 ± 15.5	15%	85%	7
	CG: CT + early mobilization	20	57.1 ± 8.7	7/13	28.4 ± 6.1	13/20	59.2 ± 13.7	15%	85%	
Liu et al., (2021) [29]	IG: IMT + aerobic + rehab	28	66.3 ± 7.9	18/10	73% > 23	18/28	53.5 ± 15.6	100%		7
	CG: rehab	26	64.2 ± 5.9	14/12	57% > 23	23/26	49.4 ± 15.1	100%		

Abbreviations: BMI, body mass index; CG, control group; CT, chest therapy; IEMT, inspiratory-expiratory muscle training; IG, intervention group; IMT, inspiratory muscle training; pre-op MIP, preoperative maximal inspiratory pressure; rehab, rehabilitation program; VATS, video-assisted thoracoscopic surgery.

Table 2. Characteristics of the interventions and results

First author (year)	Sample size	Intervention program	Time (beginning and duration)	IMT program			Outcome findings
				Intensity	Frequency	Device	
Brocki et al. (2015)	SG: 34	IMT + Rehab: mobilization, deep breathing, PEP, cough.	after surgery, 2w	15% MIP, increase 2 cmH_2O/d	30x2 Twice a day Daily	Powerbreath	PPCs: • Hypoxemia: IG (5/34) vs CG (12/34), P = .049 • Pneumonia: same presentation than for hypoxemia, P = 0.14 • Atelectasis: idem, P = 0.11 6MWT: • IG: D = - 48.1 ± 71.9 m, P < 0.001 • CG: D = - 31.7 ± 79.1m, P > 0.05 • D = -16.4 m, 95% CI, -52.4 - 21.3, P = 0.21 MIP: • IG: MD = .21 ± 17.9 cmH_2O, P > 0.05 • CG: MD = - 4.92 ± 15.1 cmH_2O, P > 0.05 • D = 4.5 cmH_2O (95% CI, - 3.85 - 12.85), P = 0.22 MEP: • IG: MD = - 4.7 ± 16.3 cmH_2O, P > 0.05 • CG: MD = - 0.5 ± 16.2 cmH_2O, P > 0.05 • D = - 4.12 cmH_2O (95% CI, -12.27 - 4.04), P = 0.26 Spirometry:
	CG: 34	Rehab: mobilization, deep breathing, PEP, cough	after surgery, 2w	-	-	-	

							• FVC (% Pred): no difference, P = 0.57 • FEV1 (% Pred): idem, P = 0.14
Brocki et al. (2018)	SG: 34	IMT + Rehab: mobilization, positive expiration pressure, cough.	after surgery, 2w	15% MIP, increase 2 cmH_2O/d	30 x2, Twice a day, continued for 2w	Powerbreath	QoL: IG vs CG: P = 0.8 Physical activity scale 2.1: • Sedentary: IG (6%) vs CG (22%) P = 0.006 • Low activity: IG (56%) vs CG (66%) P = 0.006 • Moderate activity: IG (38%) vs CG (12%) P = 0.006
	CG: 32/34	Rehab: mobilization, positive expiration pressure, cough	after surgery, 2w	-	-	-	
Mesaggi-Sartor et al. (2018)	SG: 11/16	IEMT + Aerobic	4-6 w post-surgery, 8w	50% MIP/MEP, increase 10 cmH_2O/w	10 x5 twice a day, 3d/w	Orygen dual valve	6 months after surgery VO2peak: MD = 2.13 ml/Kg/min (95% CI, 0.06 - 4.20), P = 0.044 MIP: • IG (% Pred): Pre (68.5 ± 27.5) – Post (83 ± 21.2) • CG (% Pred): Pre (71.6 ± 32.9) – Post (70 ± 29.1) • MD between 2 groups = 13.42 cmH_2O (95% CI, 2.7 - 24.1), P = 0.015 MEP: • IG (% Pred): Pre (64.5 ± 19.5) – Post (78.4 ± 17.8) • CG (% Pred): Pre (62.2 ± 21.6) – Post (68.8 ± 20.1) • MD between 2 groups = 18.58 cmH_2O (95% CI, 4.0 - 33.1), P = 0.023 QoL: improve EORTC QoL-C30 score after intervention, but no significant difference between IG
	CG:13/21	PA advice, WHO-recommended exercise encouragement	4-6 w post-surgery	-	-	-	

							and CG (P > 0.05)
Taşkin et al. (2020)	SG: 20	CT + early mobilization + IMT	1-2 days after surgery, around 2w	15% MIP, Increase 2 cmH_2O/d up to 45%	10 x6 Once/d, 5ds/w	Threshold IMT	6MWT: 6 months after surgery • IG: MD = - 29.5 ± 122.3 m, P > 0.05 • CG: MD = - 115.2 ± 128.9m, P < 0.01 • D = 85.72m (95% CI, 166.15 - 5.28), P = 0.037 MIP: • IG: MD = 3.1 ± 17.3 cmH_2O, P > 0.05 • CG: MD = - 8 ± 16.4 cmH_2O, P > 0.05 • D = 11.05 (95% CI 21.84 to 0.25), P = 0.045 MEP: • IG: MD = 1.1 ± 17.7 cmH_2O, P > 0.05 • CG: MD = -24.1 ± 34.6 cmH_2O, P < 0.01 • D = 25.23 (95% CI, 42.83 - 7.62), P = 0.006 Hospital stays: IG (9.1 ± 3) vs CG (12.9 ± 4.2) P = 0.002
	CG: 20	CT + early mobilization	1-2 days after surgery, around 2w	-	-	-	
Liu et al. (2021)	SG: 26/32	Aerobic +IMT+ rehab (breathing control, limbs exercises, airway clearance technique, incentive spirometry, lung expansion)	4 days after surgery), 6w	30% MIP, increase to 45%MIP	10-15 3-4 /d, daily	Threshold IMT	PPCs: • Pneumonia: idem, P = 0.165 • Atelectasis: idem, P = 0.295 • ARDS: idem, P = 0.519 • Respiratory failure: idem, P = 0.519 • Subcutaneous emphysema:

	CG: 28/31	Rehab (breathing control, limbs exercises, airway clearance technique, incentive spirometry, lung expansion)	(4 days after surgery), 6w	-	-	-	idem, P = 0.893 • Air leak: idem, P = 0.177 6MWT at week 2: • CG (332.9 ± 78.4) vs IG (412.57 ± 74.2) P = 0.002 6MWT at week 6: • CG (306.8 ± 70.6) vs IG (419.02 ± 60.7) P = 0.007 6MWT at week 12: • CG (306.9 ± 58.2) vs IG (402.4 ± 65.2) P = 0.036 MIP at week 2: • CG (77.4 ± 34.9) vs IG (91.9 ± 25) P = 0.124 MIP at week 6: • CG (71.7 ± 27.9) vs IG (94.3 ± 32.8) P = 0.018 MIP at week 12: • CG (80.0 ± 30.9) vs IG (95.7 ± 34.7) P = 0.120 MEP at week 2: • CG (70.9 ± 24.3) vs IG (90.9 ± 28.1) P = 0.015 MEP at week 6: • CG (79.1 ± 24.5) vs IG (94.8 ± 30.4) P = 0.066 MEP at week 12: • CG (76.1 ± 20.2) vs IG (98.6 ± 35.3) P = 0.012

Abbreviations: ARDS, acute respiratory distress syndrome; CG, control group; CT, chest therapy; FEV1, first second of forced expiration; FVC, forced vital capacity; IEMT, inspiratory-expiratory muscle training; IG, intervention group; IMT, inspiratory muscle training; MD, mean difference; MEP, maximal expiratory pressure; MIP, maximal inspiratory pressure; PA, physical activity; PEP, positive expiratory pressure; PPC, postoperative pulmonary complication; QoL, quality of life; rehab, rehabilitation program; 6MWT, 6-minute walking test; WHO, World Health Organization.

Pulmonary Postoperative Complications

Postoperative pulmonary complications (PPCs) were assessed in two studies. Only one study reported a significantly lower rate of hypoxemia in the IG than in the CG (15% vs 35%, P = 0.049), but no difference in pneumonia (P = 0.14) or atelectasis (P = 0.11) between the two groups without associated with the type of surgery [30]. Another study reported the same result of PPCs in two groups, with VATS used in both groups [29].

Physical capacity

Three studies assessed 6MWT [29,30,33]. Only two of them demonstrated a significant difference in the walking distance between the two groups. One of them showed a significant difference after two weeks, six weeks, and 12 weeks [29]. The other reported a significant decrease in 6MWT in CG, while there was no significant change in IG; the difference between the two groups was statistically significant [33].

One study assessed $VO_{2peak,}$ and the authors reported a significant improvement in the IG compared to the CG [32].

Maximal inspiratory pressure and Maximal expiratory pressure

Maximal inspiratory pressure (MIP) was assessed in four studies. Three studies showed a significant difference in MIP between the two groups. One reported a significant difference at 6 months after surgery [33], and one after 10 to 12 weeks since the end of the intervention [32]. One of them reported a significant difference at week 6, but not at weeks 2 and 12 [29]. Another study demonstrated no significant difference between IG and CG (P=0.22) [30].

Three of the four studies that investigated MEP demonstrated a significant difference in MEP between the two groups [32,33] but one of them only reported a significant difference at weeks 2 and 12, but not at week 6 [29].

Quality of Life

Two studies that assessed QoL reported no significant difference between IGs and CSs. One study used EQ-5D-5L (P=0.8) [31]. The other assessed QoL using the European Organization for Research and Treatment of Cancer Quality of Life (EORTC QL-30) (P>0.05) [32].

Physical activity level

One study assessed PA by using the physical activity scale 2.1 reported less sedentary and more moderate activity levels in the IG than in the CG (P = 0.006) [31]. Another study assessed the physical functioning scale with EORTC QL-30, finding no significant difference between the two groups (P > 0.05) [32].

Hospital length of stay

Only one study assessed the hospital length of stay as an outcome. It was shorter in the IG (9.1 ± 3 days) than in the CG (12.9±4.2 days) (P=0.002) [33].

Spirometry

Only one study assessed the effect of P-IMT on lung function, without a significant difference between IG and CG [30].

Discussion

This systematic review focused on the effect of P-IMT after surgery in PWLC. There was insufficient evidence to promote the use of P-IMT. The included studies were heterogeneous in terms of intervention programs, which made it challenging to analyze the results. Due to the few studies, the findings should be cautiously interpreted.

The effect of P-IMT on PPCs was difficult to analyze because of the methodology of the studies. According to the result, the earlier starting point with a higher %MIP did not show more benefit in PPCs compared to the four-day delayed lower intensity P-IMT. The longer duration of the 6-week P-IMT did not show any benefit on PPCs compared with the 2-week program. However, due to the natural

recovery process of the operative lung being 6 weeks, the assessment conducted earlier than 6 weeks after surgery may not be good for evaluating the effect of P-IMT in PPCs [35]. In addition, the high rate of smoking and high BMI in these studies, which also increases the risk of PPCs [36], should be considered while interpreting the results. The benefit of P-IMT on PPCs with different surgery types was not found. However, with the same percentage of subcutaneous emphysema and air leak in both groups, P-IMT did not increase these risks, contributing to the feasibility of P-IMT in postoperative PWLC. Therefore, P-IMT can be used to improve diaphragm function after surgery [37-40]. More studies are needed to clarify this point, as well as the ideal characteristics of the program.

Studies in this systematic review showed that the short-term effect of P-IMT on 6MWT was unclear. Noticeably, the two studies, which showed a benefit in 6MWT [29,32], added aerobic exercise in the IG but not in the CG; therefore, aerobic exercise can be the reason for the improvement of 6MWT [41-43]. In addition, it was impossible to recommend when to start P-IMT due to the different starting points in the included studies. The result of this review did not suggest any ideal setting of P-IMT for the improvement of 6MWT. Despite the similar duration, intensity, and frequency of P-IMT in the two studies, only one of them reported a significant difference in 6MWT between the two groups [30,33]. Due to this discrepancy, concluding the short-term effect of P-IMT on 6MWT was challenging. The ideal intensity and frequency of P-IMT for short-term 6MWT could not be determined. With only one study, it was difficult to assess the effect of P-IMT on VO_{2peak}. Like 6MWT, the improvement of VO_{2peak} may be induced by aerobic exercise applied in IG but not CG [41-43]. The same caution should be applied while interpreting the effect of mid-to-long-term P-IMT on physical capacity. More studies with longer follow-ups are needed to understand the specific role as well as the ideal characteristics of P-IMT on physical capacity.

As noted earlier, understanding the effect of P-IMT on inspiratory muscles is important because of the negative correlation previously

observed between MIP and PPCs [44]. With similar intensity and frequency, the short-term effect of P-IMT on MIP was contradictory. Two studies reported a significant short-term reduction in post-operative MIP prior to recovering to the pre-operative state without significant difference between the two groups [30,33]. Thus, the decline may be due to inspiratory muscle dysfunction caused by postoperative pain in the early days [45]. And the recovery of MIP can be a natural process after 2 weeks of surgery [46]. There was insufficient evidence to suggest that P-IMT had different efficacy in patients with different surgeries. More studies are needed to understand the effect of P-IMT on MIP.

According to our results, the mid-to-long-term effects of P-IMT on MIP were weak. Two studies showed a significant difference in MIP between IG and CG with 6-to-8-week training [29,32]. However, both studies included aerobic exercise in IG but not in CG; thus, the improvement of aerobic-induced general health can be the reason for the enhancement of MIP [47]. Noticeably, the significant difference in MIP between the 2 groups disappeared after 12 weeks post-surgery [29]. Thus, retaining training may be a good recommendation to maintain MIP, especially in elderly people [48].

For the MEP, the evidence was conflicting. Two out of the three studies showed short-term benefits of MEP [29,33]. As no changes were observed in IG contrary to the CG, it is suggested that P-IMT can maintain or promote the recovery of MEP [33]. However, this study applied PEP for control and intervention groups. Although PEP cannot be exactly defined as expiratory muscle training, it has to be considered while interpreting the results [49]. Another study reported a significant reduction in MEP after surgery, but recovered to pre-surgery status after 2 weeks in both groups without a significant difference between the two groups [30]. Thus, recovery may be the natural process after surgery [46]. Although there was some supportive evidence, it was challenging to define the effect of P-IMT on short-term MEP.

The improvement of the long-term MEP was observed [29,32]. However, the effect should be cautiously interpreted since aerobic exercises, which were applied in IGs but not in the CGs, may be the reason for the improvement [41]. Besides, the MEP reevaluation time of these studies was less than 12 weeks, which was believed to be not long enough to reverse the strength induced by resistance training [50-53].

Due to the importance of the change in QoL, PA level, hospital length of stay, and spirometry in the LC population, investigating the effects of P-IMT in these aspects was worthy [54-57]. However, with only one study assessing each of the outcomes stated above, it was difficult to conclude. In addition, the time of the QoL assessment in the included studies was too early to catch the true score. Indeed, QoL was usually assessed 6 months after surgery [11,58]. Our result reported a higher PA level with P-IMT [31]. The reason can be the improvement of the whole-body exercise capacity contributed by inspiratory muscle strengthening, which facilitates the PA level [59]. One study reported a shorter hospital length of stay with early IMT [33]. The earlier discharge was only observed in IG despite early mobilization, which was believed to reduce the hospital length of stay, was applied in both groups [60,61]. Therefore, P-IMT can promote the earlier discharge after LC surgery. No effect of P-IMT on spirometry tests was found in this systematic review, with only one study assessing spirometry. In addition, the test was conducted at week 2 postoperatively, when pain can still impair breathing maneuvers, leading to unreliable results [62].

The main limitation of this review was the small number of studies. With only five studies included, data extraction and analysis were challenging. The included studies had a heterogeneous population with differences in intensity and training programs. Thus, it made the comparison difficult. The wide range of duration and starting point of intervention was also another limitation. In addition, the inconsistent outcomes measurement also challenged the interpretation.

Conclusion

Overall, this systematic review reported conflicting results of P-IMT in PWLC with surgery. No effect of P-IMT in PPCs was reported. The effect of P-IMT on physical capacity and respiratory muscle strength was not clear. P-IMT can be used as an additional therapy to improve hospital length of stay and postoperative physical activity levels with a low level of evidence. In addition, we could not find the optimal setting of P-IMT for PWLC with surgery. Further studies with clear settings of the intervention program and specific outcomes should be conducted to confirm the role of P-IMT on PWLC with surgery.

The impact of isolated inspiratory muscle training on exercise capacity, dyspnea, and quality of life in heart failure patients: a systematic review and meta-analysis

Ngoc-Minh Nguyen; David Vancraeynest; Anne-Catherine Pouleur; Marc Beaumont; Philippe Fait; Gregory Reychler

Adapted from: Physical Therapy Reviews, P 93-105, March 2025

DOI: 10.1080/10833196.2025.2475643

Abstract

Introduction: Recently, inspiratory muscle training (IMT) has emerged as an alternative therapeutic approach for patients with heart failure (HF), characterized by ventilation inefficiency and reduced functional capacity. However, the evidence regarding the effects of IMT on patients with HF remains ambiguous. This meta-analysis aims to investigate the impact of isolated IMT on the 6-minute walk test (6MWT), VO2peak, dyspnea, resting heart rate (RHR), and Quality of Life (QoL) in patients with HF.

Methods: A systematic literature search was conducted across four databases: Science Direct, PubMed, Cochrane, and EBSCO. Two reviewers independently performed literature searches and screenings. The quality and risk of bias were assessed using the Physiotherapy Evidence Database (PEDro) scale and the Cochrane Risk-of-Bias tool for randomized trials (RoB2). Meta-analysis was conducted using RevMan 5.4.1 and Open Meta-analyst.

Results: Seven clinical trials involving 211 patients with HF were included. Meta-analysis revealed that IMT with intensity above 25% Maximum Inspiratory Pressure (MIP) significantly improved the 6MWT (61.81 m; 95% CI: 10.99 - 112.63 m), VO2peak (3.82 ml/kg/min; 95% CI: 3.25 – 4.39 ml/kg/min), dyspnea (-0.44; 95% CI: -0.76 to -0.12), and QoL (-12.23; 95% CI: -21.28 to -3.19), but not in RHR (-4.59; 95% CI: -10.91 to 1.72).

Conclusion: IMT with intensity above 25% MIP demonstrated improvements in the 6MWT, VO2peak, dyspnea, and QoL in patients with HF.

Keywords: inspiratory muscle training, heart failure, clinical trial, functional exercise capacity, quality of life.

Introduction

Heart failure (HF), afflicting over 26 million patients worldwide, poses a significant societal and healthcare burden [63]. Projections indicate a 127% increase in the estimated economic cost of HF treatment from 2015 to 2035, reaching $69.8 billion in the United States [64]. Despite advancements in medical interventions, patients with HF continue to experience elevated 1-year mortality rates and frequent hospitalizations [65]. Furthermore, a considerable percentage (16.7%-28%) of patients with HF grapple with moderate to severe disabilities, encompassing limitations in daily life activities (8.9%-30.3%), mobility (2%-36%), and societal participation (53.3%) [66]. Dyspnea and ventilation inefficiency contribute to compromised sleep quality and diminished exercise tolerance [67]. Notably, patients with HF exhibit a reduced quality of life (QoL), diminished functional capacity, and inadequate self-care [68,69].

Rehabilitation has emerged as a pivotal therapeutic approach for managing HF [70]. While traditional aerobic and resistance training have long been established as effective and are widely endorsed in numerous HF management guidelines [71-73], barriers such as time constraints, physical limitations, perceptions, and supervision requirements hinder patients' participation in conventional rehabilitation [74,75]. Consequently, identifying exercises that enhance rehabilitation engagement poses a challenge. Recently, alternative approaches like respiratory muscle training have gained attention and are being explored as potential interventions.

Inspiratory muscle training (IMT) has garnered increasing attention due to its simplicity, cost-effectiveness, and patient-administered

nature. Furthermore, inspiratory muscle weakness has been linked to dyspnea during physical tasks, ventilation inefficiency, and diminished functional capacity in patients with heart failure (HF) [76-78], providing the rationale for strengthening these muscles in this patient population. While there is current evidence supporting the benefits of inspiratory muscle training, the existing evidence on the impact of isolated IMT in patients with HF remains inconclusive, as most studies have incorporated other interventions such as aerobic or resistance training alongside IMT. This systematic review with meta-analysis seeks to elucidate the effects of isolated IMT on key parameters such as the 6-minute walk test (6MWT), peak oxygen consumption (VO2peak), dyspnea, resting heart rate (RHR), and Quality of Life (QoL) in patients with HF. Additionally, the study aims to compare the effectiveness of various IMT approaches.

Method

Search strategy

This systematic review, adhering to the PRISMA checklist [79] was conducted from inception to January 2025 across four electronic databases: ScienceDirect, PubMed, Cochrane, and EBSCO. Two reviewers (N-M. N. and G. R.) independently identified and reviewed the studies. The search employed the equation: ("inspiratory muscle training" (Title/Abstract) OR (IMT (Title/Abstract)) OR ("respiratory muscle training" (Title/Abstract)) AND ("heart failure" (Title/Abstract) OR "cardiac failure" (Title/Abstract) OR "CHF" (Title/Abstract) OR "chronic heart failure" (Title/Abstract) OR "congestive heart failure" (Title/Abstract)) AND (6MWT OR functional capacity OR VO2max OR Dyspnea OR resting heart rate OR quality of life).

Eligibility Criteria

Studies were selected based on the following eligibility criteria: 1) the primary focus was to investigate the effect of IMT in patients with HF, 2) randomized or controlled clinical trials with a controlled/Sham group (CG), and 3) main outcomes included at least one of the following measures: 6MWT, VO2peak, dyspnea, RHR, and QoL.

Exclusion Criteria

Exclusion criteria encompassed: 1) studies incorporating combined training methods other than IMT in the intervention group (IG), 2) comorbidities such as COPD or diabetes mellitus, 3) studies not in English.

Data Extraction

The data from selected studies were independently extracted by two reviewers (N-M. N., G. R.), encompassing: 1) participant characteristics and demographic data (sample size, age, gender, VO2peak, maximal inspiratory pressure (MIP), New York Heart Association classification of heart failure (NYHA) level, body mass index (BMI), left ventricle ejection fraction (LVEF)), 2) details of the intervention program (frequency, intensity, type, duration) for both the intervention group (IG) and control group (CG), and 3) outcome measures and key findings (6MWT, VO2peak, dyspnea, RHR, and QoL).

Statistical Analysis

Meta-analysis was conducted using RevMan 5.4.1 [80]. Mean changes and standard deviations from the studies were processed with Open Meta-analyst [81]. Pooled-effect estimates were calculated using the least square mean percentage change from baseline and presented as the weighted mean difference (WMD) or standardized mean difference (SMD) as appropriate. Mean and standard deviation were estimated from the median and range using the method of Hozo et al. [82]. For continuous data, the meta-analysis employed the inverse variance, random-effect model, and a 95% confidence interval (95% CI) [83]. Heterogeneity in the meta-analysis was assessed using the I^2 statistic [84], with classification into low (I^2 = 25%), moderate (I^2 = 50%), and high (I^2 = 75%) levels [85].

Methodological Quality

The methodological quality of the studies was evaluated using the Physiotherapy Evidence Database (PEDro) scale, with the assigned

scores verified by a single reviewer (N-M. N.) [26]. The PEDro scale comprises 11 items, with item 1 (eligibility criteria) left unscored, resulting in a maximum score of 10. A higher score on the PEDro scale indicates better study quality. The risk of bias in the included studies was assessed by two reviewers using version 2 of the Cochrane Risk-of-Bias Tool for Randomized Trials (RoB 2) [86].

Results

Initially, 450 articles were identified through the electronic database search (Web of Science = 10, PubMed = 111, Cochrane = 212, and EBSCO = 117). After removing duplicates, 368 studies remained. The abstracts and titles of these studies were screened for inclusion criteria, resulting in 12 articles selected for full-text reading. Five articles were subsequently excluded, and seven studies were systematically reviewed. Figure 1 illustrates the flow of the selection process.

Description of included studies

A total of 211 patients participated in the seven included studies, with sample sizes ranging from 23 to 38 patients. The gender distribution was approximately one-third female and two-thirds male. The median age of the patients across studies was 57.3 years. Four studies recruited patients classified with the New York Heart Association (NYHA) classification of heart failure as NYHA II/III [87-90], while one study included patients with NYHA III/IV [91]. The NYHA classification of patients was not reported in two studies [92,93]. The mean BMI of all patients was categorized as overweight (Table 1) [94]. Six studies encompassed patients with reduced LVEF [87-90,92,93], while one study included patients with preserved LVEF [91]. The mean baseline maximal inspiratory pressure (MIP) of patients in each study exceeded 50 cmH2O, surpassing the threshold for inspiratory muscle weakness based on the mean age of participants [95]. Furthermore, there were no statistically significant differences in baseline values for the 6MWT, VO2peak, RHR, and QoL between groups (all P >

0.05). Data in all studies were reported as mean and standard deviation, except for Palau et al. (2014). (Table 1)

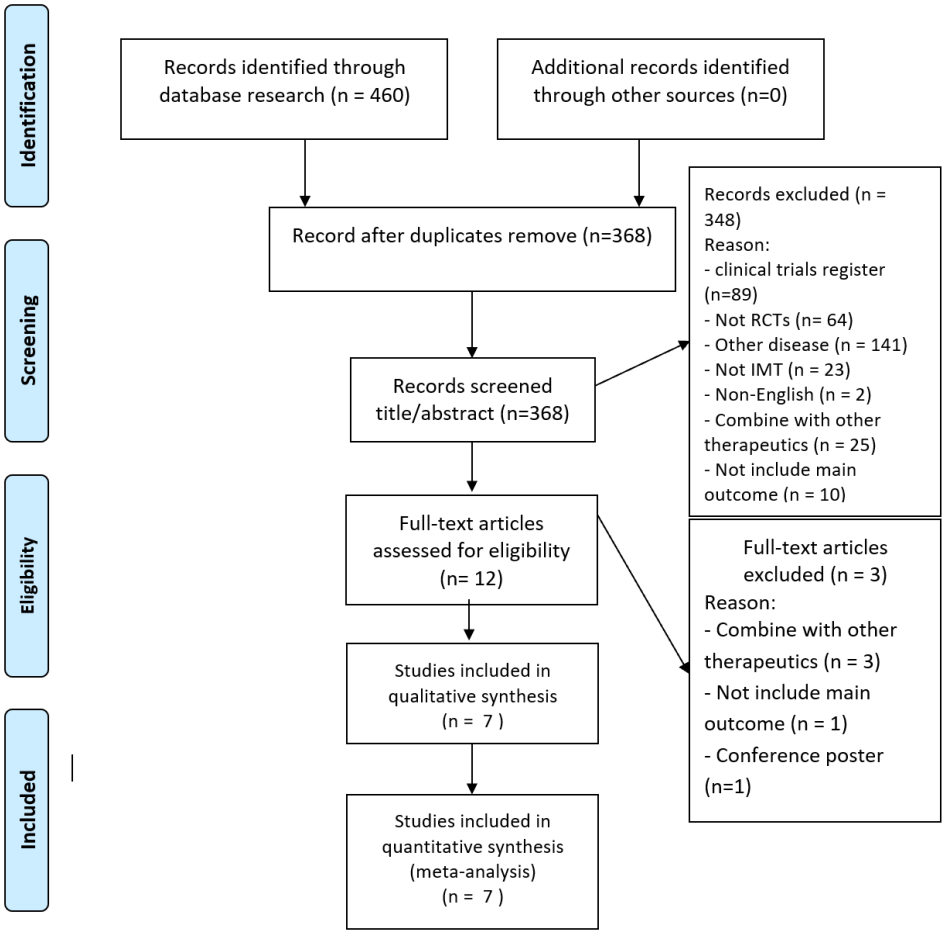


Figure 1. Flow chart of the search strategy

Table 1. Characteristics of included studies

	Design	Sample size	Mean age (SD)	Gender (F/M)	BMI (SD)	NYHA (II/III)	LVEF (%)	MIP (cmH_2O)	VO_2peak (ml/kg/min)	PEDro score
Laoutaris et al. (2004)	IG	20	57.6 (2.3)	2/18	28.0 (1.1)	12/8	23.4 (1.5)	82.8 (25.5)	15.4 (4.0)	5
	CG	15	60.0 (2.6)	3/18	27.0 (1.0)	7/7	25.3 (2.2)	78.4 (26.7)	14.7 (3.8)	
Dall' Ago et al. (2006)	IG	16	54.0 (33)	5/11	27.0 (5.0)	-	39.0 (3.0)	60.2 (9.2)	17.2 (0.5)	5
	CG	16	58.0 (2.0)	6/10	27.0 (4.0)	-	38.0 (3.0)	58.1 (1.0)	17.0 (0.7)	
Laoutaris et al. (2007)	IG	15	53.0 (2.0)	3/12	29.2 (2.1)	10/5	28.0 (3.0)	79.8 (18.2)	17.3 (3.5)	6
	CG	23	59.0 (2.0)	3/20	28.6 (0.8)	12/11	27.0 (2.0)	80.2 (24.0)	15.7 (3.8)	
Laoutaris et al. (2008)	IG	14	53.4 (2.1)	3/11	29.3 (1.3)	9/5	28.9 (2.4)	79.1 (18.7)	17.1 (2.6)	6
	CG	9	57.3 (4.0)	0/9	30.6 (1.3)	6/3	28.6 (1.9)	84.2 (26.1)	17.7 (3.9)	
Bosnak-Guclu et al. (2011)	IG	16	69.5 (7.9)	4/12	26.8 (4.3)	11/5	33.4 (7.4)	62.0 (33.6)	-	6
	CG	14	65.7 (10.5)	2/12	25.1 (3.2)	9/5	36.1 (7.6)	78.6 (36.0)	-	
Mello et al. (2012)	IG	15	54.3 (2.0)	6/9	28.8 (1.6)	15/0	33.6 (2.3)	59.2 (19.0)	14.4 (2.7)	6

	CG	12	53.3 (2.0)	7/5	27.4 (0.9)	12/0	37.6 (1.6)	63.2 (18.4)	16.2 (1.7)	
Palau et al. (2014) *	IG	14	68.0 (4.6)	7/7	34.3	9/5 (III/IV)	69.5 (4.2)	69.9 (8.2)	10.3 (1.5)	6
	CG	12	74.0 (1.5)	6/6	30 (P =0.03)	9/3 (III/IV)	77.8 (4.4)	71.3 (10.4)	9.4 (1.6)	

*: the mean and SD of the study were estimated by the method of Hozo et al. (2005), IG: Intervention group, CG: Control group; F/M: female/male; BMI: Body mass index; NYHA: New York heart association classification; LVEF: left ventricle ejection fraction; MIP: maximal inspiratory pressure; PEDro: the physiotherapy evidence database.

Intervention Program Description

All included studies exclusively focused on IMT without incorporating concurrent nonpharmacological treatments (Table 2). Training frequency varied, with sessions lasting 30 or 40 minutes per day, conducted 3 to 7 days per week. The duration of training programs ranged from 6 to 12 weeks. Among the studies, four provided detailed descriptions of IMT training procedures [87-90]. In three studies, the IMT group was compared with a controlled/sham group (no-load or usual care) using resistance intensities ranging from 25% to 30% of MIP [91-93]. Additionally, four studies compared two different intensities of inspiratory muscle training. One study compared 40% MIP-IMT vs. 15% MIP-IMT [87], while three others compared 60% MIP-IMT vs. 15% MIP-IMT [88-90]. To evaluate the optimal IMT intensity, changes were assessed within moderate-to-high MIP-IMT (>25% MIP-IMT) groups, low MIP-IMT (15% MIP-IMT) groups, and sham/usual care (S-UC) groups.

Quality Assessment and Risk of Bias

The PEDro scores for the included studies ranged from 5 to 6 (Table 1). An overview of the overall risk of bias in the studies is presented in Figure 2, where the risk of bias in five domains of each study was deemed high in 42% of cases in the randomized process (n=3 studies) and high in 14% in terms of missing outcome data (n=1 study) (Figure 3).

The Effect of Inspiratory Muscle Training

A meta-analysis was conducted to compare the impact of >25% MIP-IMT with the control group (15% MIP-IMT and S-UC). Key outcome measurements are detailed in Table 2.

Table 2. Characteristics of intervention and outcome findings

	Design (Device)	Sample size	Intervention program				Outcome findings 6MWT, VO_2peak, RHR, ans QoL (SD)
			Intensity	Frequency	Time	Type	
Laoutaris et al. (2004)	60% MIP (Electrical device: TRAINAIR)	20	60% MIP	6 efforts/level, more than 6 levels/session, 3 days/w	10 weeks	Trained until fatigue with reduced interval resting time from 60s to 5s	• 6MWT: 60%: 367.1 (99.7) to 433.4 (75.6m) (P=0.001) 15%: 343.7 (96) to 352.1 (85.6m) (P>0.05) Difference between 2 groups: No information • VO_2peak: 60%: 15.4 (4) to 17.8 (5.4) ml/kg/min (P=0.002) 15%: 14.7 (3.8) to 14.7 (3.8) ml/kg/min (P>0.05) Difference between 2 groups: No information • Dyspnea: 6-20 Borg scales after 6MWT 60%: 10.5 (3.1) to 9 (2.2) (P = 0.001) 15%: 12.7 (3.1) to 12.6 (3.1) (P > 0.05) Difference between 2 groups: No information • RHR: 60%: 80.2 (13.4) to 76.8 (14.7) bpm (P=0.04)
	15% MIP (Electrical device: TRAINAIR)	15	15% MIP	6 efforts/level, 6 levels/session, 3 days/w	10 weeks	6 levels with reduced interval resting time from 60s to 5s	

							15%: 78.6 (20.5) to 76.1 (18.2) bpm (P>0.05) Difference between 2 groups: No information • QoL (Minnesota living with heart question): 60%: 25.2 (17.8) to 21.1 (15.6) (P=0.004) 15%: 22.9 (10) to 22.6 (9.7) (P>0.05) Difference between 2 groups: No information
Dall'Ago et al. (2006)	30% MIP (Threshold IMT)	16	30% MIP Readjust the load weekly	30mins/ 7days/w	12 weeks	No information	• 6MWT: 30%: 449 (17) to 550 (17) m (P<0.002) Sham: 432 (41) to 411 (60) m (P>0.05) Difference between 2 groups: P < 0.002 • VO_2peak: 30%: 17 (0.6) to 21 (0.7) ml/kg/min (P<0.001) Sham: 17 (0.6) to 17 (0.8) ml/kg/min (P>0.05) Difference between 2 groups: P < 0.001 • Dyspnea: 0-10 Borg scales after 6MWT 30%: 3.7 (2) to 1,5 (1.4) Sham: 3.1 (1.3) to 3 (1.4)
	Sham IMT (Threshold IMT)	16	Sham IMT No inspiratory load	Idem	Idem	Idem	

Laoutaris et al. (2007)	60% MIP (Electrical device: TRAINAIR)	15	60% MIP Readjust the load after every session	6 efforts/level, more than 6 levels/session, 3 days/w	10 weeks	Trained until fatigue with reduced interval resting time from 60s to 5s	Difference between 2 groups: P < 0.002 • QoL (MLHFQ): 30%: 27 (4) to 6 (2) Sham: 30 (13) to 30 (13) Difference between 2 groups: P < 0.001 • 6MWT: 60%: 378.2 (40.3) to 404.3 (46) m (P<0.01) 15%: 358 (47.9) to 366 (79.1) m (P>0.05) Difference between 2 groups: Nonsignificant • VO_2peak: 60%: 17.3 (3.5) to 19.4 (4.6) ml/kg/min (P<0.01) 15%: 15.7 (3.8) to 14.8 (3.8) ml/kg/min (P>0.05) Difference between 2 groups: P<0.01 • Dyspnea: 6-20 Borg scales after 6MWT 60%: 9.2 (1.5) to 8 (1.5) (P = 0.001) 15%: 11.8 (2.87) to 11.5 (2.87) (P > 0.05) Difference between 2 groups: P <0.001 • Resting HR:
	15% MIP (Electrical device: TRAINAIR)	23	15% MIP Readjust the load after every session	6 efforts/level, 6 levels/session, 3 days/w	10 weeks	6 levels with reduced interval resting time from 60s to 5s	

							60%: 82 (19.4) to 81 (11.6) bpm (P>0.05) 15%: 82 (19.2) to 81 (19.2) bpm (P>0.05) Difference between 2 groups: Nonsignificant
Laoutaris et al. (2008)	60% MIP (Electrical device: TRAINAIR)	14	60% MIP Readjust the load after every session	6 efforts/level, more than 6 levels/session, 3 days/w	10 weeks	Trained until fatigue with reduced interval resting time from 60s to 5s	• VO_2peak: 60%: 17.1 (2.6) to 19 (4.5) ml/kg/min (P=0.01) 15%: 17.7 (3.9) to 17.3 (4.5) ml/kg/min (P=0.9) Difference between 2 groups: P = 0.7 • Dyspnea: 6-20 Borg scales after ergospirometry 60%: 18.1 (0.37) to 17.6 (0.74) (P = 0.02) 15%: 17.6 (0.6) to 17.9 (0.9) (P = 0.08) Difference between 2 groups: P = 0.05 • Resting HR: 60%: 83 (22.4) to 80 (11.2) bpm (P=0.6) 15%: 86 (15) to 89 (15) bpm (P=0.4) Difference between 2 groups: P= 0.3
	15% MIP (Electrical device: TRAINAIR)	9	15% MIP Readjust the load after every session	6 efforts/level, 6 levels/session, 3 days/w	10 weeks	6 levels with reduced interval resting time from 60s to 5s	
Bosnak-Guclu et	40% MIP (Threshold	16	40% MIP Readjust	30mins/ 7days/w	6 weeks	Diaphragm breathing	• 6MWT:

al. (2011)	IMT)		the load weekly			with device: 25-30 efforts, with 5-10s resting time between breath	40%: 418.59 (123.32) to 478.56 (131.58) m (P<0.001) 15%: 462.31 (113.59) to 475.99 (135.79) m (P=0.091) Difference between 2 groups: P < 0.001
	15% MIP (Threshold IMT)	14	Fixed 15% MIP without adjustment during protocol	Idem	Idem	Idem	• Dyspnea: 0-10 Borg scales after 6MWT 40%: 2.42 (1.73) to 1.42 (1.31) (P = 0.025) 15%: 2.64 (1.21) to 1.77 (1.13) (P = 0.059) Difference between 2 groups: P = 0.82 • QoL (SF-36): + Physical health: 40%: 46.01 (28.03) to 67.38 (24.16) (P<0.001) 15%: 51.96 (23.41) to 69.19 (22.06) (P<0.001) Difference between 2 groups: P = 0.81 + Mental health: 40%: 57.67 (23.61) to 70.06 (20.91) (P=0.004) 15%: 55.39 (24.39) to 71.78 (21.60) (P=0.001) Difference between 2

							groups: P = 0.68
Mello et al. (2012)	30% MIP (Threshold IMT)	15	30% MIP Adjust the load weekly	30 mins/session 7 days/w	12 weeks	No information	• VO_2peak: 30%: 14.4 (2.7) to 18.9 (3.1) ml/kg/min (P < 0.05) Sham: 16.2 (1.7) to 16.3 (2.1) ml/kg/min (P > 0.05) Difference between 2 groups: P < 0.05 • QoL (MLHFQ): 30%: 26.6 (14.7) to 9.2 (9.2) (P < 0.05) Sham: 30.8 (21.1) to 32.7 (19.4) (P < 0.05) Difference between 2 groups: P < 0.05
	Sham (Threshold IMT)	12	Sham IMT No load	30 mins/session 7 days/w	12 weeks	No information	
Palau et al. (2014)*	25-30% MIP (Threshold IMT)	14	25-30% MIP Readjust the load weekly	20mins, 2 times/ day, 7days/w	12 weeks	No information	• 6MWT: 25-30%: 320 (78.3) to 387 (22.3) m (P<0.001) Usual care: 273.8 (65.08) to 260 (76.35) m (P = 0.167) Difference between 2 groups: P < 0.001 • VO_2peak: 25-30%: 10.25 (1.47) to 12.88 (1.28) ml/kg/min (P < 0.001) Usual care: 9.4 (1.62) to 8.83 (1.07) ml/kg/min (P =
	Usual care	12	Usual care without IMT	N/A	12 weeks	N/A	

					0.206) Difference between 2 groups: P < 0.001 • Resting HR: 25-30%: 72.25 (5.51) to 64.75 (6.66) bpm (P<0.001) Usual care: 72.25 (10.6) to 71.25 (6.03) bpm (P=0.146) Difference between 2 groups: P = 0.012 • QoL (LMHFQ): 25-30%: 41 (4.04) to 30 (2.89) (P = 0.002) Usual care: 45.5 (11.35) to 41 (10.69) (P = 0274) Difference between 2 groups: P = 0.037

*: the mean and SD of the study were estimated by the method of Hozo et al. (2005), F/M: female/male; NYHA: New York Heart Association classification; LVEF: left ventricular ejection fraction; MIP: maximal inspiratory pressure; PEDro: the physiotherapy evidence database; MLHFQ: Minnesota Living with Heart question; Resting HR: resting heart rate

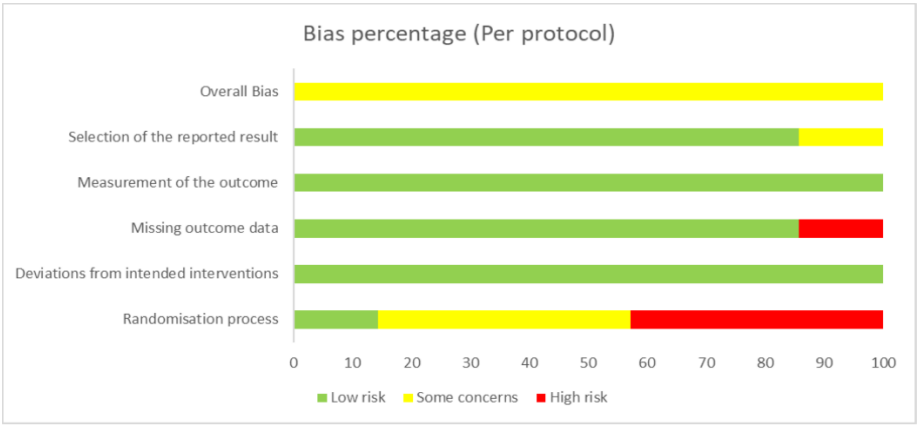


Figure 2. Overall risk of bias. The overall Bias has some concerns

Figure 3. Risk of bias for individual study

6 Minutes Walking Test

Five studies assessed 6MWT [88,90-93]. The meta-analysis indicated a pooled effect of 61.81 m (95% confidence interval (CI): 10.99, 112.63 m) for the distance covered in the 6MWT in the >25% MIP-IMT vs control group. The studies exhibited moderate to high heterogeneity (I^2 = 67%) (Figure 4). No significant change in 6MWT was observed within the 15% MIP-IMT and S-UC groups (Table 2).

VO_2peak

Six studies provided data on peak oxygen consumption (VO2peak) [88-93]. The meta-analysis favored the >25% MIP-IMT group, revealing a pooled effect of 3.82 ml/kg/min (95% CI: 3.25, 4.39 ml/kg/min). Heterogeneity was low, with I^2 = 0% (see Figure 5). No significant improvement in VO2peak was observed within the 15% MIP-IMT and S-UC groups (Table 2).

Dyspnea

Dyspnea levels were assessed in five studies using either a 6-20 scale or a modified Borg scale after the 6MWT or ergospirometry [87,89,90,93,96]. Due to differences in measurement scales, the standardized mean difference was employed to demonstrate the pooled effect. Figure 6 illustrates a significant reduction in the Borg score (-0.44; 95% CI: -0.76, -0.12), with an I^2 of 0, indicating a low heterogeneity level.

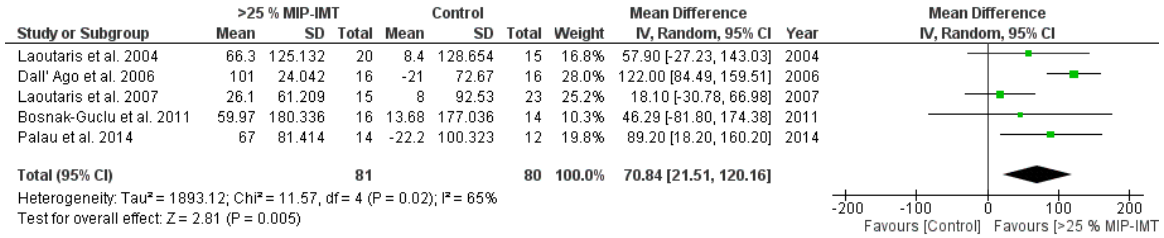


Figure 4. The pooled effect on the distance of 6MWT

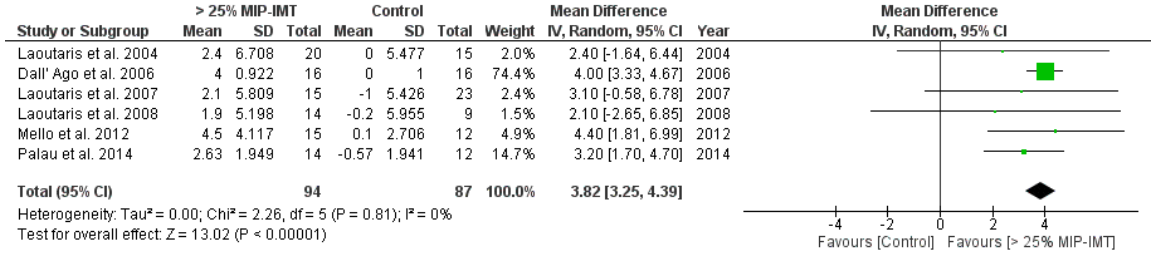


Figure 5. The pooled effect on the change of VO_{2peak}

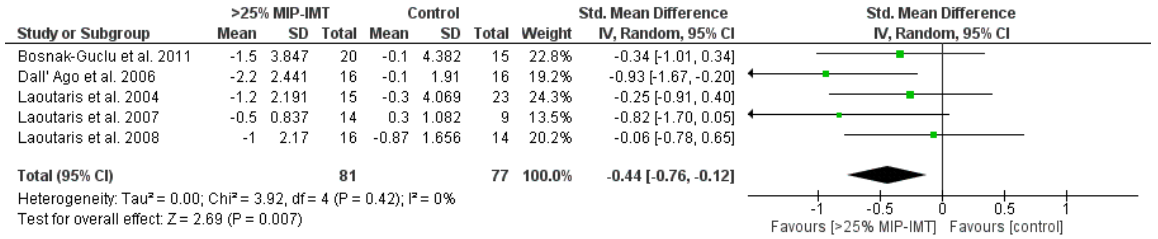


Figure 6. The pooled effect on the change of the Borg score

Resting Heart Rate

Resting heart rate was assessed in four studies [88-91]. The meta-analysis revealed no significant difference between groups, with a pooled effect of -4.59 (95% CI: -10.91, 1.72). The I^2 statistic indicated a low heterogeneity level (Figure 7).

Quality of life

Although five studies measured Quality of Life (QoL), the meta-analysis incorporated only four studies that utilized the Minnesota Living with Heart Failure Questionnaire (MLHFQ) [88,91-93]. One study that measured QoL with the SF-36 was excluded from the meta-analysis [87]. The total pooled effect on QoL was -12.23 (95% CI: -21.28, -3.19). The heterogeneity level was high, with $I^2 = 58\%$ (Figure 8).

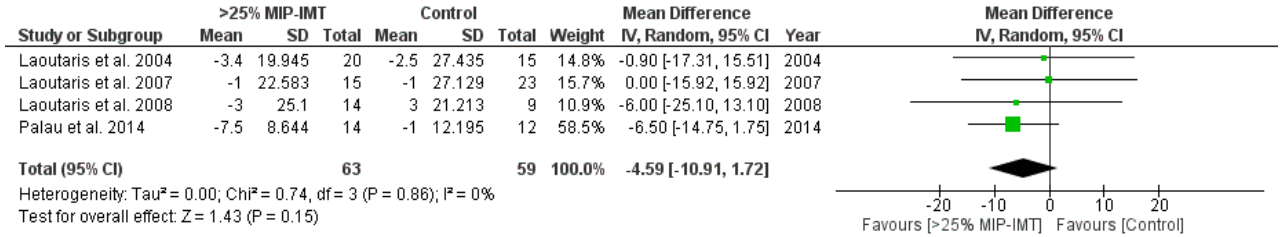


Figure 7. The pooled effect on the change in the resting heart rate

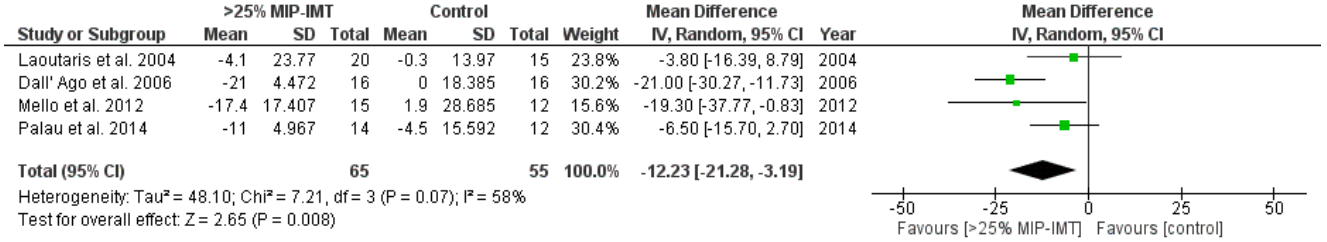


Figure 8. The pooled effect on the change in the Quality of Life

Discussion

This meta-analysis underscores the positive impact of IMT on key outcomes such as 6MWT, VO2peak, dyspnea, and QoL in patients with HF. However, certain considerations and cautions in interpreting the results should be acknowledged.

Firstly, the aggregated effect favors >25% MIP in 6MWT, potentially attributable to improved inspiratory muscle function linked to enhanced functional exercise capacity in patients with HF [78]. Additionally, IMT has demonstrated positive effects on peripheral chemoreflex, exercise ventilation oscillation, and blood flow to peripheral limbs, collectively enhancing physical capacity in patients with HF [97,98]. The improvement in 6MWT with >25% MIP-IMT training surpasses the minimal clinically important difference of 30-50m, indicating a meaningful clinical impact on the clinical symptoms, morbidity, and mortality in patients with HF [99-103]. However, caution is advised due to the high heterogeneity level among studies. Moreover, we encountered unexpected data in a particular study, indicating a reported standard deviation (SD) of 17 meters in the IMT group [93]. In our assessment, when compared to findings in other studies, this reported SD appears unusually low. Upon excluding this study, the I^2 statistic decreased to 0%, and the 95% pooled effect reduced to 44 meters, which is not exceeding the range of MCID of 6MWT mentioned above. The presence of anomalies in reported data in this specific study underscores the importance of conducting additional high-quality studies to validate the effects of IMT on 6MWT.

The improvement in VO2peak by 3.82 ml/min/kg with >25% MIP-IMT is clinically significant, as higher VO2peak is associated with reduced mortality risk in patients with HF [104,105]. Furthermore, an increase of roughly 1 MET (1 MET = 3.5 ml/kg/min) can enhance the survival rate of patients with HF by 20% [106] to 28% [107]. The increase in diaphragm thickness and lung volume after IMT training may contribute to this enhancement [108,109]. High-intensity IMT could be a valuable intervention, especially for patients with HF unable to

engage in traditional resistance and aerobic training due to musculoskeletal issues [110-112]. However, all studies were not longer than 3 months, unable to observe the long-term benefit of IMT. Sensitivity analysis did not reveal a significant change in the pooled effect.

Given dyspnea's prominence in patients with HF, better dyspnea management can help to improve QoL [113]. The study provides evidence supporting the alleviation of dyspnea with >25% MIP-IMT. Improved respiratory muscle perfusion and deoxygenation are potential contributors to lower dyspnea levels [114,115]. While respiratory muscle weakness is considered a factor in dyspnea [116,117], this cannot be solely attributed to respiratory muscle training due to the multifaceted nature of dyspnea [118,119]. IMT is suggested as a supplemental intervention, emphasizing the need for a multifactorial approach for comprehensive dyspnea management [118].

The results of this study recommend a load of 25-30% MIP-IMT as an initial load for patients with HF, deviating from the potentially insufficient effects observed with a 15% MIP load, which is similar to the proposed intensity of IMT for COPD patients [120]. As resistance training, current evidence suggests that with an intensity of less than 20% of the maximal voluntary contraction, fiber recruitment is not effective in inducing muscle hypertrophy [121,122].

No significant difference in Resting Heart Rate (RHR) was observed between the groups. Notably, none of the studies defined how and under what conditions RHR was measured, underscoring the necessity for a standardized approach to RHR measurement for a more comprehensive understanding of IMT effects on this parameter. For instance, if RHR is measured before engaging in exercise, the true baseline RHR may not be accurately captured due to the anticipated rise in heart rate in response to physical activity [123]. Therefore, the lack of a standardized measurement for RHR, which is notably absent in the existing literature [124,125], warrants attention. Additionally, the lack of information regarding Beta-blocker

medication means its potential confounding effect on RHR variability across studies cannot be definitively excluded.

The pooled effect on QoL, as measured by the Minnesota Living with Heart Failure Questionnaire (MLHFQ), favored >25% MIP-IMT, with a reduction score of 12.23, indicative of potential mortality reduction[126-128]. One study measured QoL with SF-36 and reported no significant difference between groups, but significant improvement after the intervention time within the groups [87]. The evidence supporting QoL improvement was more robust in studies with an intervention duration of 12 weeks or longer, reinforcing the importance of prolonged IMT for QoL benefits. This result is similar to a systematic review of Azambuja, de Oliveira and Sbruzzi [129].

This meta-analysis focused on isolated IMT, thereby mitigating biases associated with other exercise interventions. However, the small number of trials and subjects and the low quality of included studies pose limitations. The I^2 statistic may be overestimated with the current number of studies [130], and the overall study quality is a concern. Additionally, the difference in training protocol may be a potential factor contributing to the high heterogeneity. To establish a more reliable conclusion regarding IMT's effect on patients with HF, additional high-quality randomized controlled trials are imperative. Investigating key parameters such as optimal durations, frequency, long-term effects, and potential variations in response based on HF severity is crucial. Additionally, exploring the synergistic effects of combining IMT with other rehabilitation modalities is essential for optimizing and diversifying heart failure rehabilitation strategies.

In conclusion, this study supports the benefits of IMT with intensity above 25% MIP in enhancing 6MWT, VO2peak, and alleviating dyspnea in patients with HF. Furthermore, prolonged IMT has exhibited a positive impact on quality of life. IMT stands as a valuable adjunct to cardiac rehabilitation programs for patients with HF, though further high-quality studies are required to fine-tune the optimal IMT program.

2. A reliability aspect of maximal inspiratory pressure measurement

To evaluate the trajectory of inspiratory muscle training or monitor changes in inspiratory muscle strength, it is important to have good reliability of MIP measurement. Although it is well established that MIP measurement has moderate reliability in general, there is a gap in evidence regarding the reliability of MIP in persons with Chronic Obstructive Pulmonary Disease (COPD). Current evidence on MIP reliability in persons with COPD shows inconsistencies across different equipment and protocols. In addition, due to the new updated guidance from the European Respiratory Society on respiratory muscle pressure measurement, published in 2019, some studies conducted before may not adhere to the latest guidance, thereby reducing confidence in their results. To address this gap and ensure the reliable use of MIP measurement in clinical practice, a study was conducted to investigate the between-session reliability of maximal inspiratory pressure measurement in persons with COPD. This is particularly important due to the high prevalence rate of COPD in Vietnam (6.7%).

The Reliability of Maximal Inspiratory Pressure Measurement and Its Correlation with Submaximal Exercise Test in Persons With Chronic Obstructive Pulmonary Disease

Ngoc-Minh Nguyen; Bich-Ly Thi Ngo; Hien Thi Le; William Poncin; Gregory Reychler

Status: under review – Cardiopulmonary Physical Therapy Journal

Abstract

Background: Maximal inspiratory pressure (MIP) is a simple measure that reflects inspiratory muscle strength, which is often impaired in individuals with Chronic obstructive pulmonary disease (COPD). This study aimed to assess the test-retest reliability, minimal detectable change (MDC) of MIP in persons with COPD, as well as to revisit its relationship with the 1-minute sit-to-stand test (1STST).

Methods: Individuals with stable COPD were recruited. The study protocol included two visits. During the first visit, participants performed the MIP measurement and the 1STST with a 30-minute interval. The second visit was repeated within 10 days to assess between-day test-retest reliability.

Results: Fifty-five participants were recruited, with 44 completing the full protocol. The MIP measurement demonstrated excellent test-retest reliability (ICC = 0.94, 95% IC: 0.89–0.97, p = 0.001). Bland-Altman plots revealed a mean bias in the MIP measurement of 3.3 cmH_2O, with limits of agreement ranging from -17.7 to 24.3 cmH_2O. The MDC for MIP was estimated to be 7.3 cmH_2O. A weak positive correlation (r=0.29, p=0.03) was observed between MIP and 1STST performance.

Discussion: This study demonstrated excellent between-day test-retest reliability of MIP measurement in persons with mild to moderate COPD. Additionally, a weak positive correlation was found between 1STST and MIP in this population.

Keywords: maximal inspiratory pressure, test-retest reliability, COPD, minimally important difference, 1-minute sit-to-stand test.

Introduction

Chronic Obstructive Pulmonary Disease (COPD) is a major cause of mortality worldwide, with rates continuing to rise according to the World Health Organization[131]. COPD has systemic effects that extend beyond the lungs, leading to multiple impairments that severely impact patients' quality of life, exercise capacity, and functional capacity[132]. Among these impairments, inspiratory muscle weakness is especially prevalent in persons with COPD, contributing to both breathlessness and diminished exercise capacity[133]. Particularly, the decrease in inspiratory muscle strength becomes substantial in advanced stages, potentially leading to a further reduction in exercise capacity, an important factor associated with mortality and hospitalization [134].

Evaluating inspiratory muscle strength through maximal inspiratory pressure (MIP) has become essential in COPD management, as it is associated with disease prognosis and mortality rates[135]. For instance, MIP correlates with the 6-minute walking distance or 1-minute sit-to-stand test (1STST)[133], which can predict mortality risk and disease progression[136,137]. Thus, identifying a true change in MIP can help determine the appropriateness of therapeutic interventions, such as inspiratory muscle training, which has been shown to improve functional capacity in many populations [138,139]. Therefore, establishing the minimally detectable change (MDC) for MIP in COPD patients is essential for informing clinical decision-making and guiding rehabilitation strategies, particularly concerning functional capacity. Although MIP measurement validity in COPD is well established, its reliability has not yet been thoroughly investigated, as highlighted by a recent systematic review [140]. For instance, three studies included in this systematic review on persons with COPD reported test-retest reliability using correlation coefficients (rho), a method no longer recommended, rather than Intraclass Correlation

Coefficients (ICC) [2]. Additionally, prior studies assessing MIP reliability in persons with COPD show inconsistencies across equipment and protocols: some employed devices that did not adequately address buccal pressure effects due to limited technological advancement [141,142], while others deviated from American Thoracic Society/European Respiratory Society guidelines, likely impacting MIP measurement reliability [143,144]. Consequently, there remains a need to enhance the understanding of MIP measurement reliability in persons with COPD.

Therefore, this study aimed to examine the test-retest reliability of MIP in persons with COPD. Other objectives were to determine the minimally detectable change (MDC), along with the relationship between MIP and exercise capacity. This study also revisited the correlation between 1STST and MIP.

Methods

Ethics statement

This cross-sectional study was approved by the Ethics Committee of University of Medicine and Pharmacy at Ho Chi Minh City with the approval No. 3813/HĐĐĐ-ĐHYD. All participants were informed about the protocol and provided written consent. The data of this study was securely stored and accessible only with proper authorization. All stages of the study were conducted following the Declaration of Helsinki and Clinical Good Practice guidelines[145].

Participants

Persons with COPD receiving outpatient care at the Gia Dinh People's Hospital respiratory clinic (Ho Chi Minh City, Vietnam) were recruited into this study from November 2023 to June 2024. The study included outpatients diagnosed with COPD based on the Global Initiative for Chronic Obstructive Lung Disease (GOLD)[146], who were clinically stable for at least 4 weeks and could provide informed consent. Individuals were excluded if they had recent (< 4 weeks) infectious exacerbations, cognitive impairments, musculoskeletal disorders, or contraindications to performing a

functional exercise test. Additionally, those with mobility limitations or oral defects that prevented them from closing their mouth were also excluded.

Protocol and measurements

The study procedure was conducted over two visits. During the first visit, demographic data and comorbidities were collected. Spirometry results from tests conducted within 6 months were retrieved from medical records. Modified Medical Research Council dyspnea (mMRC) score, COPD assessment test (CAT) score, history of exacerbations, and hospitalizations over the past year were obtained from the medical records. The spirometry was conducted at the hospital following ERS/ATS guidelines by trained health staff [147]. During this visit, participants performed the MIP measurement and the 1STST, with a 30-minute interval between tests. The second visit was repeated within 10 days to assess the between-day test-retest reliability of the MIP. The 1STST measurement was also repeated during this second visit to determine whether any change in exercise capacity occurred and ensure consistent conditions between the two testing sessions, following the guidance of COSMIN[2]. Participants were excluded from the test-retest reliability assessment of MIP if there was a change of at least 3 repetitions in the 1STST between the two visits, as this represents the Minimal important difference (MID) for exercise capacity change[148]. Additionally, participants were excluded from the test-retest reliability assessment if they experienced any changes in their medications, forced expiratory volume in 1 second (FEV1), severity stage classification, or had an acute exacerbation during the study period[146].

The maximal inspiratory pressure

The measurement of MIP followed the ATS/ERS guidelines [149] using the RP Check device (MD Diagnostic Ltd, UK) with a filter and plastic flanged mouthpiece. The pressure meter had a 2 mm internal diameter air leak and was 20-30 mm in length, which allowed the opening of the glottis to be maintained and prevented extra pressure

from the buccal muscles [149]. The measurements were performed at the residual volume. All assessments were conducted with participants sitting upright, with both feet on the floor. The mouthpiece was properly fitted, and participants were guided through initial breaths before being instructed to perform a maximal inspiratory effort starting from the residual volume [149]. A nose clip was applied, and participants were verbally encouraged to exhale fully before inhaling maximally and holding for at least 1.5 seconds [149]. Each measurement was repeated at least 5 times, with a 1-minute rest between trials, until the best three values varied within 10% [149]. The highest value was recorded. All participants were coached to prevent air leaks during the test. The test was performed after the administration of routine medication. The Borg scale was used to ensure participants performed the maneuver without fatigue or dyspnea, aiming to achieve the best possible results. The assessor was trained before data collection. A single assessor conducted all measurements for the participants. To ensure consistency between tests, a checklist following the guidelines was used.

The 1-minute sit-to-stand test

A standard chair measuring 46 cm in height with no armrests was used for the 1STST following the standard guidance [150]. Participants first familiarized themselves with the movement by crossing their arms over their chests and performing one full cycle of standing up and sitting down. The participants were then instructed to stand up completely and sit down as many times as possible within one minute, without using their arms for support. The total number of completed cycles was recorded. Vital signs, including blood pressure, heart rate, oxygen saturation, and Borg's scale, were measured both before and after the test to monitor participant status [150]

Data analysis

The data analysis was conducted using SPSS version 29.0. Descriptive statistics were used to summarize the participant characteristics, and the Shapiro-Wilk test was conducted to assess normality[151]. The sample size was determined based on an anticipated ICC of 0.9, with an acceptable ICC of 0.75, 80% power, an alpha of 0.05, and a 20% dropout rate, resulting in a required sample of 42 participants [2].

The test-retest reliability of the MIP was evaluated using a two-way mixed, absolute agreement, average-measurement ICC model. ICC values over 0.5 were considered satisfactory, above 0.8 indicated good reliability, and over 0.9 was considered excellent [152]. The level of agreement between MIP measurements at the two visits was analyzed using a Bland-Altman plot, and linear regression was used to check for proportional bias [2]. The MDC was calculated using the formula MDC = 1.96 x SEM x $\sqrt{2}$, where SEM = SD x $\sqrt{(1\text{-ICC})}$ [2,153]

The relationship between the MIP and the 1STST was assessed by Pearson's or Spearman's correlation coefficients, with rho coefficients < 0.5, between 0.5 and 0.69, and > 0.7 classified as low, moderate, or high, respectively [154].

Results

Demographic

The study involved 55 persons with COPD who met the inclusion criteria. All 55 participants completed the 1STST and MIP tests in the first assessment. By the second visit, 44 participants continued and completed the full protocol. Eleven of the original 55 participants were unable to complete the full protocol due to reasons such as failing to meet the test-retest criteria (n=7), missing scheduled visits (n=2), and declining further participation (n=2). Most participants were male, and the majority had a normal body mass index. The disease severity was distributed equally across the GOLD ABE categories, with 60% of the participants having moderate airflow

obstruction. Almost all the participants had at least one additional medical condition, primarily hypertension, lipid disorders, and osteoporosis (Table 1).

Table 1. Characteristic of participants (N = 55)

Demographics	Value	Minimum - Maximum
Age, year (mean ± SD)	67.0 ± 6.3	50-81
Male sex, n (%)	53 (96)	
BMI, kg/m2 (mean ± SD)	20.9 ± 2.6	16.2-27.2
BMI classification, n (%)		
Underweight	12 (21.8)	
Normal weight	32 (58.2)	
Overweight	8 (14.5)	
Obesity	3 (5.5)	
FEV1, L (mean ± SD)	1.43 (0.43)	0.64-2.45
FEV1, % predicted (mean ± SD)	60 (17.9)	21-100
FVC, L (mean ± SD)	2.61 (0.54)	1.61-4.12
FEV1/FVC, ratio (mean ± SD)	0.57 (0.14)	0.28-0.69
GOLD classification, n (%)		
GOLD 1 (mild)	9 (16.4)	
GOLD 2 (moderate)	33 (60.0)	
GOLD 3 (severe)	11(20.0)	
GOLD 4 (very severe)	2 (3.6)	
ABE classification, n (%)		
Group A	19 (34.5)	
Group B	18 (32.7)	

Group E	18 (32.7)
Comorbidities n(%)	*52 (95)*
Hypertension	24 (76.4)
Lipid disorder	22 (40)
Osteoporosis	18 (32.7)
Benign Prostatic Hyperplasia	17 (30.9)
Diabetes	7 (12.7)
Cardiac disease	16 (29.1)

n: number of subjects; BMI: body mass index; GOLD: Global Initiative for Chronic Obstructive Lung Disease. The variables are presented as mean ± standard deviation, or numbers (%).

Test-retest reliability

The mean MIP was 81.2 ± 25.3 cmH_2O on the first visit and 84.4 ± 22.8 cmH_2O on the second visit. The ICC was 0.94 (95% CI, 0.89–0.97, p = 0.001). The Bland-Altman plot of between-day MIP measurements demonstrated a mean bias of 3.3 cmH_2O and limits of agreement from -17.7 to 24.3 cmH_2O (Figure 2). The linear regression model showed no evidence of proportional bias (p = 0.1). Based on the distribution-based method, the MDC was 7.3 cmH_2O.The correlation coefficient between the MIP and the 1STST was rho = 0.29, p = 0.03 (Figure 2).

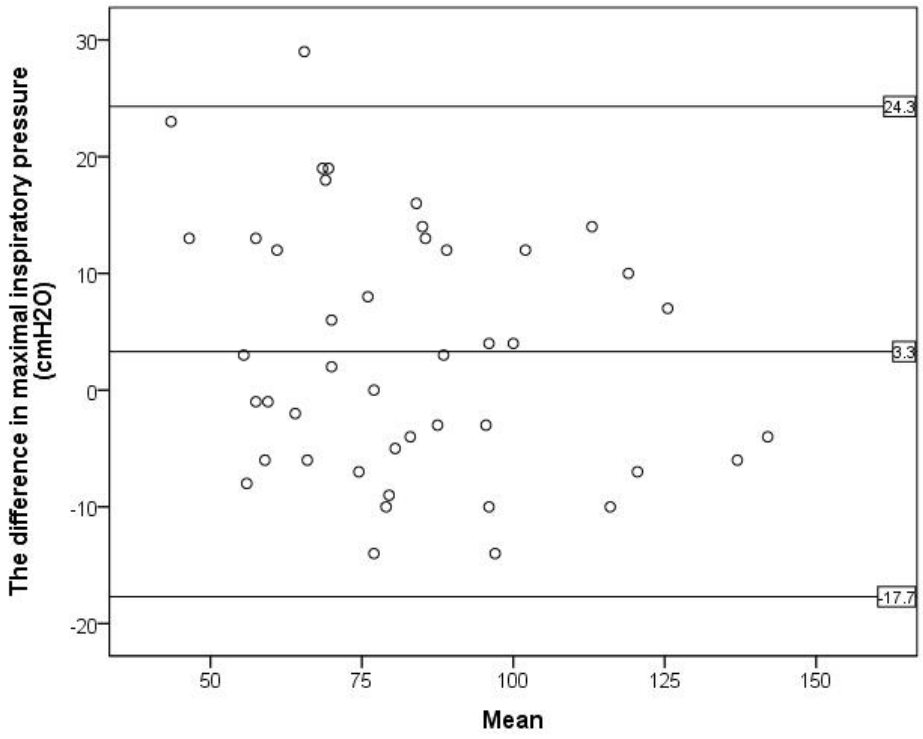


Figure 1. The Bland-Altman plot for the two maximal inspiratory pressure measurements

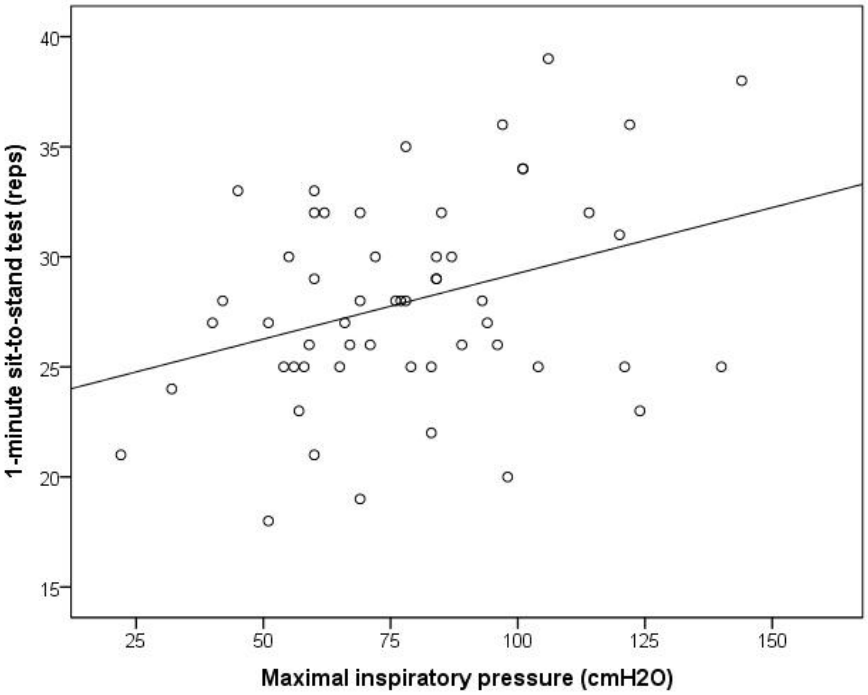


Figure 2. The correlation between maximal inspiratory pressure and the 1-minute sit-to-stand test

Discussion

This study demonstrated excellent test-retest reliability for MIP measurement in persons with COPD, indicating that the measurement of MIP is highly reliable and reproducible in this population. The Bland-Altman plot revealed good agreement between the two MIP assessments, with a systematic error of only 3.3 cmH_2O. Additionally, this error is less than 10% of the overall mean MIP value reported in the study, satisfying the reproducibility criteria of the MIP measurement by international guidelines [149]. Besides, without proportional bias, the reliability of MIP measurement was not affected by the magnitude of the values in persons with COPD [155]. However, the range between the lower and upper limits of agreement, indicating the random error, was relatively wide, even with the strict implementation of the measurement protocol following

updated international guidelines to minimize random error and optimize reliability [149]. This finding can be attributed to the variability in individual effort, motivation, and encouragement during the maneuvers, which are similarly observed in other muscle assessments [156,157]. Given the volitional nature of the MIP test, even with a rigorous protocol, multiple attempts may be needed to help individuals understand the movement better, refine their technique, and reduce the learning curve, leading to more consistent and reliable measurements [158].

The minimal detectable change for the MIP was estimated at 7.3 cmH_2O, which is very close to the range of MID from 7.9 to 13.5 cmH_2O suggested in the study of Beaumont et al. [159]. This finding supports the use of the MID from the mentioned study, as it meets the criterion of being higher than the MDC from this study [2]. However, this study also revealed a substantially wider range of LoA than the MDC of this study and even than the MID from the mentioned study. Thus, the reported LoA challenges the capacity of the MDC of 7.9 cmH_2O to identify true physiological changes. While the MDC is a distribution-based analysis, derived from group-level statistics to demonstrate a threshold for meaningful change, the limits of agreement represent individual-level variability, indicating the random error inherent in the measurement [2]. Clinically, this distinction highlights that the MDC should be carefully interpreted alongside other clinical and physiological parameters, such as dyspnea or accessory respiratory muscle overactivity, to differentiate true clinical improvement from mere measurement variability. Additionally, this finding suggests that a high ICC alone may not sufficiently capture a measurement's consistency and variability; therefore, LoA may also be essential for a more comprehensive test interpretation. This is particularly relevant for the test requiring the best individual effort and encouragement, such as the MIP, as discussed above.

This study found a weak positive correlation between MIP and the 1STST, which was lower than the existing evidence (rho = 0.29 versus rho = 0.51), despite having comparable participants [160]. The reason for this disparity is unclear. Compared to the observed correlations between MIP and exercise capacity measured by the 6-minute walk test (6MWT) (rho = 0.53-0.55), our study also reported a lower correlation [161,162]. Although both the 6MWT and 1STST demonstrate a good correlation with exercise capacity, they may assess different aspects of exercise capacity in persons with COPD[163]. While the 1STST focuses primarily on lower limb performance, the 6MWT requires a more integrated contribution of physiological functions, including respiratory muscle function. Therefore, maximal inspiratory pressure may more influence on the 6MWT results than the 1STST. Given the observed low correlation, our study suggests that MIP may only partially contribute to overall exercise capacity in persons with COPD. However, regularly assessing MIP in conjunction with other objective and subjective measures of respiratory function, muscle status, and exercise capacity could offer a more comprehensive understanding of a patient's functional status and response to interventions.

Strengths and limitations

The strengths of this study include the robust methodology for assessing the test-retest reliability, which used controlled criteria to avoid changes in health conditions. This approach was highlighted by COSMIN as a way to improve confidence in the reliability analysis[164]. Furthermore, the study employed well-defined respiratory muscle testing procedures, following the European Respiratory Society guidelines[149]. The MIP was measured using a handheld digital manometer that met the guidance requirements, ensuring the quality of the data collection.

This study's primary limitation was using a distribution-based approach, which is generally considered less robust than an anchor-based approach, to estimate the MDC for MIP. Inadequate recording

of medication timing before the test prevented the determination of its effect on MIP measurement reliability. Additionally, due to the small sample size, we were unable to investigate the reliability of MIP across each severity stage of the disease, where the MIP values and their associated reliability metrics might vary. This limitation underscores the necessity for larger-scale studies encompassing diverse cohorts to comprehensively evaluate the nuances of MIP reliability across the spectrum of COPD severity.

Conclusion

This study demonstrated excellent test-retest reliability of MIP measurement in persons with COPD, with the MDC of 7.3 cmH2O. However, the MDC should be used with caution in clinical practice. There was a weak positive correlation between 1STST and MIP in this population.

3. To implement inspiratory muscle training in Vietnam: How to identify individuals with inspiratory muscle weakness?

To utilize IMT, the physiotherapist must identify patients who will benefit from the intervention. Those exhibiting respiratory muscle weakness are the most likely to benefit from this intervention. The most common clinical measurements to evaluate respiratory muscle strength are respiratory maximal pressures, including maximal inspiratory pressure (MIP), maximal expiratory pressure (MEP), and sniff nasal inspiratory pressure (SNIP). However, these parameters are influenced by several factors, including age, sex, height, and, in particular, ethnicity. As there are no reference values for Southeast Asians in general, and Vietnamese in particular, a study was conducted to establish the predictive values of MIP, MEP, and SNIP in the Vietnamese population. With the established reference values, clinicians can compare and classify respiratory muscle strength more appropriately, enabling more targeted interventions and improved outcomes.

The Predictive Value of Maximal Inspiratory Pressure, Maximal Expiratory Pressure, and Sniff Nasal Inspiratory Pressure for Southeast Asian Adults

Ngoc-Minh Nguyen; Hanh Thi-Bich Tran; Thi-Quynh-Nhu Do; Nicolas Audag; Giuseppe Liistro, Philippe Fait; Gregory Reychler

Status: accepted – Respiratory Medicine and Research

Abstract

Background: Accurate assessment of respiratory muscle strength is crucial for diagnosing and managing respiratory diseases. However, existing reference values may not be generalizable across diverse populations. This study aimed to establish predicted values for maximal inspiratory pressure (MIP), maximal expiratory pressure (MEP), and sniff nasal inspiratory pressure (SNIP) in Southeast Asian adults.

Methods: MIP, MEP, and SNIP were measured in 301 healthy Vietnamese adults. Two-way ANOVA and post-hoc tests were used to examine differences in these measures among age groups and between genders. Stepwise multiple linear regression was used to develop predictive equations for MIP, MEP, and SNIP, with potential predictors including age, gender, body mass index, and lung function. The lower limit of the normal range was determined using the fifth percentile of the negative residuals.

Results: MIP, MEP, and SNIP were higher in males than in females. MIP and MEP declined with age, while SNIP remained relatively stable. Predictive equations were established: MIP = 95.2 – 32.1 x gender (male = 0, female = 1) – 0.41 x age + 1.2 x BMI (adjusted R^2: 41%, LLN = predicted MIP – 34), MEP = 135.1 – 46.75 x gender (male = 0, female = 1) - 0.6 x age + 1.34 x BMI (adjusted R^2: 41.8%, LLN = predicted MEP – 50), SNIP = 63.8 – 18.16 x gender (male = 0, female = 1) (adjusted R^2: 14.8%, LLN = predicted SNIP – 30).

Conclusions: This study provides ethnic-specific predictive equations for MIP, MEP, and SNIP, which may serve as a preliminary step toward developing reference values for the Southeast Asian region.

Keywords: predicted values, maximal inspiratory pressure, maximal expiratory pressure, sniff nasal inspiratory pressure, Southeast Asian adults

Introduction

Respiratory muscles play a crucial role in ventilation and gas exchange [165]. Impaired respiratory muscle function is related to the survival rate of patients with heart failure [166]. Respiratory muscle weakness is associated with adverse outcomes in various health conditions, including chronic obstructive pulmonary disease, neuromuscular diseases, heart failure syndrome, and critical care scenarios [167-170]. Thus, respiratory muscle strength assessment is important for better intervention strategies [149]. It is typically measured by the mouth pressure generated during maximal inspiratory (MIP) or expiratory (MEP) efforts, as defined in clinical guidelines [149]. Also, sniff nasal inspiratory pressure (SNIP) is a non-invasive and convenient alternative to MIP with a low risk of adverse effects [171]. Due to their usefulness in clinical practice, these measurements are widely recommended [149].

Maximal inspiratory pressure is a valuable prognostic marker for survival in heart failure, neurological disorders, and COPD [167,172,173]. Additionally, MIP serves as an index for predicting the success of ventilator weaning [174], and the risk of pneumonia in patients with myocardial infarction [175].

Similarly, MEP is recommended for assessing the efficacy of cough in patients with neuromuscular disorders [176] and airway clearance capacity in patients with head and neck cancer [177].

Moreover, the ratio between MIP and MEP has been proposed as a screening tool for diaphragm dysfunction. This approach offers a

more convenient assessment than traditional methods that require measuring vital capacity in the supine position, which may be uncomfortable for some patients [178]. Meanwhile, SNIP has been demonstrated as a sensitive index for detecting the health status decline in boys with Duchenne muscular dystrophy [179] and for alerting to the risk of intubation and mortality in patients with amyotrophic lateral sclerosis, particularly during the palliative care stage [180,181].

Although MIP, MEP, and SNIP are widely recommended, their clinical application depends on their predicted value [149]. Despite the importance of ethnic differences in respiratory pressures [149], the predicted value of these measurements has not yet been well established in the Southeast Asian population. Therefore, this study aims to investigate the predicted value of these measurements in Southeast Asian adults.

Methods

Ethic approval

The study was approved by the Ethics Committee at the University of Medicine and Pharmacy at Ho Chi Minh City (approval No. 44 HĐĐĐ-ĐHYD, approval date the 9th of January, 2024). All participants were fully informed about the study protocol and provided informed consent. Participant information will be stored securely, accessible only with a password. The study was conducted following the Declaration of Helsinki and Good Clinical Practice guidelines [145].

Participants

The study recruited healthy Vietnamese volunteers from three regions of Vietnam. Recruitment information was posted on social media, and potential subjects were contacted for in-person appointments. The snowball sampling method was used to increase the number of participants [182]. The sample size was calculated following the formula n = 100 + 50*i, with “i” is the expected

independent variables [183]. Following previous studies investigating reference norms for MIP, MEP, and SNIP [149], there are four expected predictive variables: age, gender, weight, and height. Thus, data collection was carried out between January and June of 2024, continuing until the target of 300 participants was reached. All subjects were naive to the respiratory pressure measurement procedures.

Inclusion criteria

Healthy subjects were 18 years or older without any known cardiopulmonary or neuromuscular conditions reported. They demonstrated normal lung function, with forced expiratory volume in the first second (FEV1), forced vital capacity (FVC), and FEV1/FVC ratio above the lower limit of normality [147,184], and were non-smokers without any medical treatment or thoracic deformities. All subjects had the necessary level of consciousness to perform the respiratory maneuvers.

Exclusion criteria

Exclusion criteria were any respiratory infection within the past month [185]. Any scars or abnormal signs on the thoracoabdominal area were observed and investigated to exclude subjects who had undergone thoracoabdominal surgery. Participants with a history of preterm birth were also excluded. Persons with likely respiratory symptoms (coughing, dyspnea, sore throat, fever, or swollen lymph node) confirmed by a trained medical professional were not included. Additionally, to ensure the safety of the participants, subjects with abnormal vital signs, measured before and during the test, were excluded. Abnormal signs were defined as heart rate outside the range of 60-100 bpm, a respiratory rate of less than 8 or more than 25 breaths per minute, an oxygen saturation of less than 95%, or a temperature outside the range of 36.5 - 37.5 degrees Celsius.

Measurements

Demographic features were collected. Height and weight were measured by the principal investigator. The lung function test was conducted with a handheld spirometer (Spirobank II Advanced, MIR, Italy) [186] following the requirement of the American Thoracic Society/ European Respiratory Society (ATS/ERS) Guidance [187].

A portable respiratory pressure meter, with an operating pressure of ± 300 cm H_2O and burst pressure of ± 2000 cm H_2O (RP Check, MD diagnostic Ltd, UK), was used to measure MIP, MEP, and SNIP. The MIP and MEP were measured with a filter and plastic flanged mouthpiece, while a nostril sensor completely occluding one nostril while keeping the other open was used to measure SNIP, with both nostrils tested for SNIP to select the highest generated pressure. The pressure meter device has a 2 mm internal diameter air leak and is 20-30 mm in length. The ERS recommends maintaining the opening glottis and avoiding extra pressure generated by the buccal muscles [149]. The measurements were performed at or close to the residual volume (RV), total lung capacity (TLC), or functional residual capacity (FRC) for MIP, MEP, and SNIP, respectively. The maximal pressure needs to be maintained for at least 1.5 seconds to record the maximum pressure sustained for 1 second ideally. All these measurements were performed in an upright sitting position with two feet on the floor, following the guidance of ATS/ERS for spirometry and respiratory muscle pressures [149,187]. A standardized respiratory warm-up was conducted before testing. Then, each measurement was repeated at least 5 times with a 1-minute rest between trials until the best three times varied within 10% [171,188]. The highest values were kept. All subjects were coached to prevent air leaks during the test. All assessors were trained and coached before collecting data.

Data analysis

Data were analyzed using the Statistical Package for the Social Sciences (SPSS 27.0, IBM Software, Bethesda). The data distribution was assessed using the Kolmogorov-Smirnov or Shapiro-Wilk statistic, depending on the number of subjects [189]. The results of

MIP, MEP, and SNIP were reported by age group. Two-way ANOVA and post-hoc tests were conducted to examine the differences in MIP, MEP, and SNIP among age groups and between genders. The effect size of gender and age groups was evaluated using partial Eta squared (pEs) [190]. Effect sizes were considered small (0.1 to 0.3), medium (0.3 to 0.5), or large (above 0.5) [191]. The Bonferroni method was applied to control for type 1 errors.

Predicted values were constructed using stepwise multiple linear regression, with MIP, MEP, and SNIP as the dependent variables [192]. The predictive candidate variables included gender, age, height, weight, BMI, FEV1, FVC, and FEV1/FVC ratio. The relationships between MIP, MEP, SNIP, and the predictive factors were assessed using Pearson or Spearman correlation coefficients, depending on the normality of the data distribution. Significant variables were then entered into the linear regression models [193]. The goodness of fit of a model was assessed by the coefficient of determination by adjusted R^2. The lower limit of the normal range for MIP, MEP, and SNIP was determined using the fifth percentile of the negative residuals, equivalent to 1.64 × the residual standard deviation [194,195]. A value of $p \leq 0.05$ was considered significant.

Result

Demographic and maximal respiratory pressure comparisons

A total of 347 subjects were recruited. Out of them, 46 were excluded due to abnormal spirometry (28), premature birth history (1), thoracic and abdominal scars (2), refusal to perform spirometry (2), having flu within 7 days (3), and smoking (10). The final analysis was conducted on 301 subjects (62% from the South, 11% from the Central, and 27% from the North). Among the subjects, 12 refused to perform the SNIP measurement due to discomfort with the nostril sensor. The demographic features and results of the measurements according to age group are shown in Table 1. In males, the mean of MIP, MEP, and SNIP were 107.0 ± 25.0, 143.7 ± 39.6, and 63.8 ± 24.5 cmH2O, respectively. The obtained values for females were

72.6 ± 20.8 cmH2O for MIP, 93 ± 25 cmH2O for MEP, and 45.6 ± 18.6 cmH2O for SNIP. The two-way ANOVA test demonstrated that age group and gender contributed to the between-subject effects for MIP, MEP, and SNIP. All the measurements were significantly higher in males than in females (+33, +45, and +17 cmH2O, in MIP, MEP, and SNIP, respectively). Among contributors, the pEs of the gender were 0.318, 0.327, and 0.122 for MIP, MEP, and SNIP, respectively ($p < 0.05$). The age groups had a smaller effect than gender with pEs of 0.078, 0.111, and 0.041 for MIP, MEP, and SNIP, respectively ($P < 0.05$). No significant effect of the interaction between age groups and gender was found for MIP and SNIP but for MEP ($p = 0.043$) with a very low pEs of 0.033. The difference in MIP, MEP, and SNIP between age groups is demonstrated in Table 2.

Table 1. The demographic and respiratory pressures among age groups

Age group	N (Female/Male)	Female										Male									
		Age	Height	Weight	BMI	MI	ME	SNI	FEV1	FVC	FEV1/F	Age	Height	Weight	BMI	MI	ME	SNI	FEV1	FVC	FEV1/FV
18-30	44/41 (40/39 for SNIP)	25 (20 -30)*	157 (5)	50 (38 -70)*	23.1 (17 -32)*	78 (19	97 (22	44 (17	2.5 (1.8 -3.2)*	3 (0.4)	87 (6)	24 (18 -30)*	170 (6)	64 (50 -100)*	21 (2.5)	11 (21	15 (39	64 (26	3.9 (0.5)	4.6 (0.7)	85 (6)
31-40	49/36 (46/35 for SNIP)	36 (31 -40)*	155 (5)	54 (6)	21.5 (19 -27)*	73 (23	97 (29	45 (7 -)*	2.5 (0.4)	3.1 (0.5)	82 (52 -94)*	36 (31 -40)*	167 (5)	68 (9)	23.8 (18 -31)*	11 (24	16 (35	6 0 (35 -122)*	3.4 (2.6 -5.2)*	4.2 (3.3 -5.9)*	82 (5)
41-50	29/24 (28/23 for SNIP)	45 (41 -50)*	153 (5)	52 (44 -79)*	22.6 (2.6)	77 (18	93 (23	50 (19	2.1 (0.3)	2.7 (0.5)	79 (8)	45 (3)	166 (7)	66 (9)	23.8 (2.3)	10 (26	13 (32	67 (22	3.1 (0.4)	3.8 (0.5)	82 (4)
51-60	31/16	56 (51 -60)*	155 (5)	57 (7)	23.8 (2.7)	72 (19	96 (22	48 (19	2 (0.4)	2.4 (1.6 -4.3)*	80 (6)	57 (51 -60)*	165 (160 -178)*	67 (9)	24.2 (3)	10 (33	12 (48	64 (28	2.9 (0.5)	3.7 (0.8)	79 (7)
≥ 61	20/11	65 (61 -80)*	152 (4)	54 (8)	23.3 (3.2)	57 (22	71 (18	36 (13	1.8 (0.4)	2.2 (0.4)	79 (8)	67 (3)	167 (5)	61 (7)	22.1 (2.6)	86 (18	11 (22	49 (22	2.4 (0.4)	3.3 (2.3 -3.7)*	74 (4)

N: number of subjects, MIP: Maximal inspiratory pressure, MEP: Maximal expiratory pressure, SNIP: Sniff nasal inspiratory pressure, BMI: Body mass index
MIP, MEP, SNIP, FEV1, FVC, FEV1/FVC were expressed in Mean (SD)
Age(years); Height (cm); Weight (kg); MIP/MEP/SNIP (cmH_2O); FEV1 (L); FVC (L); FEV1/FVC (%)
*: Values were described in median (min-max)

Table 2. The difference in MIP, MEP, and SNIP among age groups in both genders

Age groups		Female						Male					
a	b	The mean difference in MIP: a-b (SE)	p	The mean difference in MEP: a-b (SE)	p	The mean difference in SNIP: a-b (SE)	p	The mean difference in MIP: a-b (SE)	p	The mean difference in MEP: a-b (SE)	p	The mean difference in SNIP: a-b (SE)	p
18-30	31-40	5.3 (4.2)	0.7	-0.2 (5.0)	1.0	-2.0 (4.0)	0.9	0.6 (5.5)	1.0	-10.1 (8.4)	0.7	-1.7 (5.7)	0.9
	41-50	1.4 (4.9)	0.9	4.5 (5.7)	0.9	-5.9 (4.5)	0.6	7.4 (6.2)	0.7	18.5 (9.5)	0.3	-3.5 (6.4)	0.9
	51-60	6.0 (4.8)	0.7	1.2 (5.6)	0.9	-3.9 (4.4)	0.9	6.0 (7.1)	0.9	23.0 (10.9)	0.2	-0.5 (7.3)	1.0
	≥ 61	21.0 (5.0)*	0.002	26.0 (6.5)*	0.001	8.4 (5.0)	0.4	26.0 (8.2)*	0.01	34.4 (12.6)*	0.05	14.8 (8.4)	0.4
31-40	18-30	-5.3 (4.2)	0.7	0.2 (5.0)	1.0	2.0 (4.0)	0.9	-0.6 (5.5)	1.0	10.1 (8.4)	0.7	1.7 (5.7)	0.9
	41-50	-3.9 (4.8)	0.9	4.7 (5.6)	0.9	3.9 (4.4)	0.9	6.7 (6.4)	0.8	28.5 (9.8)*	0.03	-1.8 (6.6)	0.9
	51-60	0.7 (4.7)	1.0	1.4 (5.5)	0.9	-1.9 (4.3)	0.9	5.4 (7.3)	0.9	33.0 (11.1)*	0.03	1.2 (7.4)	1.0
	≥ 61	15.3 (5.4)*	0.04	26.1 (6.4)*	0.001	10.4 (4.9)	0.2	25.4 (8.3)*	0.02	44.5 (12.7)*	0.006	16.5 (8.5)	0.3
41-50	18-30	-1.4 (4.9)	0.9	-4.5 (5.7)	0.9	5.9 (4.5)	0.7	-7.4 (6.2)	0.7	-18.5 (9.5)	0.3	3.5 (6.4)	0.9
	31-40	3.9 (4.8)	0.9	-4.7 (5.6)	0.9	3.9 (4.4)	0.9	-6.7 (6.4)	0.8	-28.5 (9.8)*	0.03	1.8 (6.6)	0.9
	51-60	4.5 (5.2)	0.9	-3.2 (6.2)	0.9	1.9 (4.8)	0.9	-1.3 (7.8)	1.0	4.5 (11.9)	0.9	3.0 (8.0)	0.9
	≥ 61	19.1 (5.9)*	0.01	21.5 (7.0)*	0.02	14.3 (5.4)	0.06	18.6 (8.8)	0.2	16.0 (13.5)	0.7	18.3 (9.0)	0.3
51-60	18-30	-6.0 (4.8)	0.7	-1.2 (5.6)	0.9	3.9 (4.4)	0.9	-6.0 (7.1)	0.9	-23.0 (10.9)	0.2	0.5 (7.3)	1.0

	31-40	-0.7 (4.7)	1.0	-1.4 (5.5)	0.9	1.9 (4.3)	0.9	-5.4 (7.3)	0.9	-33.0 (11.1)*	0.03	-1.2 (7.4)	1.0
	41-50	-4.5 (5.2)	0.9	3.2 (6.2)	0.9	-1.9 (4.8)	0.9	1.3 (7.8)	1.0	-4.5 (11.9)	0.9	-3.0 (8.0)	0.9
	≥ 61	14.6 (5.8)	0.09	24.7 (6.9)*	0.004	12.3 (5.3)	0.14	19.9 (9.5)	0.2	11.4 (14.5)	0.9	15.3 (9.6)	0.5
≥ 61	18-30	-21.0 (5.0)*	0.002	-26.0 (6.5)*	0.001	-8.4 (5.0)	0.4	-26.0 (8.2)*	0.01	-34.4 (12.6)*	0.05	-14.8 (8.4)	0.4
	31-40	-15.3 (5.4)*	0.04	-26.1 (6.4)*	0.001	-10.4 (4.9)	0.2	-25.4 (8.3)*	0.02	-44.5 (12.7)*	0.006	-16.5 (8.5)	0.3
	41-50	-19.1 (5.9)*	0.01	-21.5 (7.0)*	0.02	-14.3 (5.4)	0.06	-18.6 (8.8)	0.2	-16.0 (13.5)	0.7	-18.3 (9.0)	0.3
	51-60	-14.6 (5.8)	0.09	-24.7 (6.9)*	0.004	-12.3 (5.3)	0.14	-19.9 (9.5)	0.2	- 11.4 (14.5)	0.9	-15.3 (9.6)	0.5

MIP: Maximal inspiratory pressure, MEP: Maximal expiratory pressure, SNIP: Sniff nasal inspiratory pressure, SE: standard error.
*: $p < 0.05$

Correlation

The study identified significant negative correlations between age and both MIP and MEP across the entire study population. However, no such correlation was found for SNIP. Additionally, significant positive correlations were observed between MIP, MEP, SNIP, and predictive variables such as gender, height, weight, BMI, FEV1, and FVC (Table 3). However, fewer significant correlation results were obtained when examining the relationships between MIP, MEP, SNIP, and predictive factors within each gender separately. The correlation plots between age and MIP, MEP, and SNIP for both genders are presented in Figure 1.

Predictive Equations for Maximal Respiratory Pressures

The regression models developed for MIP and MEP demonstrated relatively robust predictive power, with adjusted R^2 values exceeding 35%. In contrast, the predictive models for SNIP exhibited lower adjusted R^2 values, less than 20%, indicating weaker predictive ability. Nevertheless, a moderate correlation was observed between MIP and SNIP ($r = 0.6$, $p = 0.001$). Furthermore, when analyzed separately by gender, the predictive equations for MIP, MEP, and SNIP demonstrated relatively weak predictive power, with adjusted R^2 values below 15%. Table 4 outlines the predictive equations and the corresponding lower limit of normal for these respiratory pressure measurements.

Table 3. The correlations between maximal respiratory pressures and predictive variables

Predictive variables	The whole study population			Female			Male		
	MIP	MEP	SNIP	MIP	MEP	SNIP	MIP	MEP	SNIP
Gender	-0.60*	-0.62*	-0.38*	NA	NA	NA	NA	NA	NA
Age	-0.24*	-0.25*	-0.11	-0.23*	-0.23*	-0.07	-0.23*	-0.27*	-0.09
Height	0.45*	0.47*	0.3*	0.09	0.11	0.12	0.08	-0.05	-0.06
Weight	0.42*	0.42*	0.31*	0.007	-0.004	0.04	0.16	0.11	0.16
BMI	0.21*	0.19*	0.18*	-0.04	-0.07	-0.3	0.25*	0.18*	0.24*
FEV1	0.58*	0.60*	0.38*	0.25*	0.27*	0.14	0.32*	0.33	0.19*
FVC	0.58*	0.59*	0.39*	0.24*	0.21*	0.14	0.27*	0.29*	0.18*
FEV1/FVC	0.07	0.09	0.04	0.1	0.2*	0.07	0.17*	0.15	0.05

MIP: maximal inspiratory pressure, MEP: maximal expiratory pressure, SNIP: sniff nasal inspiratory pressure, BMI: Body mass index, FEV1: Forced expiratory volume in the first second, FVC: Forced vital capacity.
NA: not applicable
*: $p < 0.01$

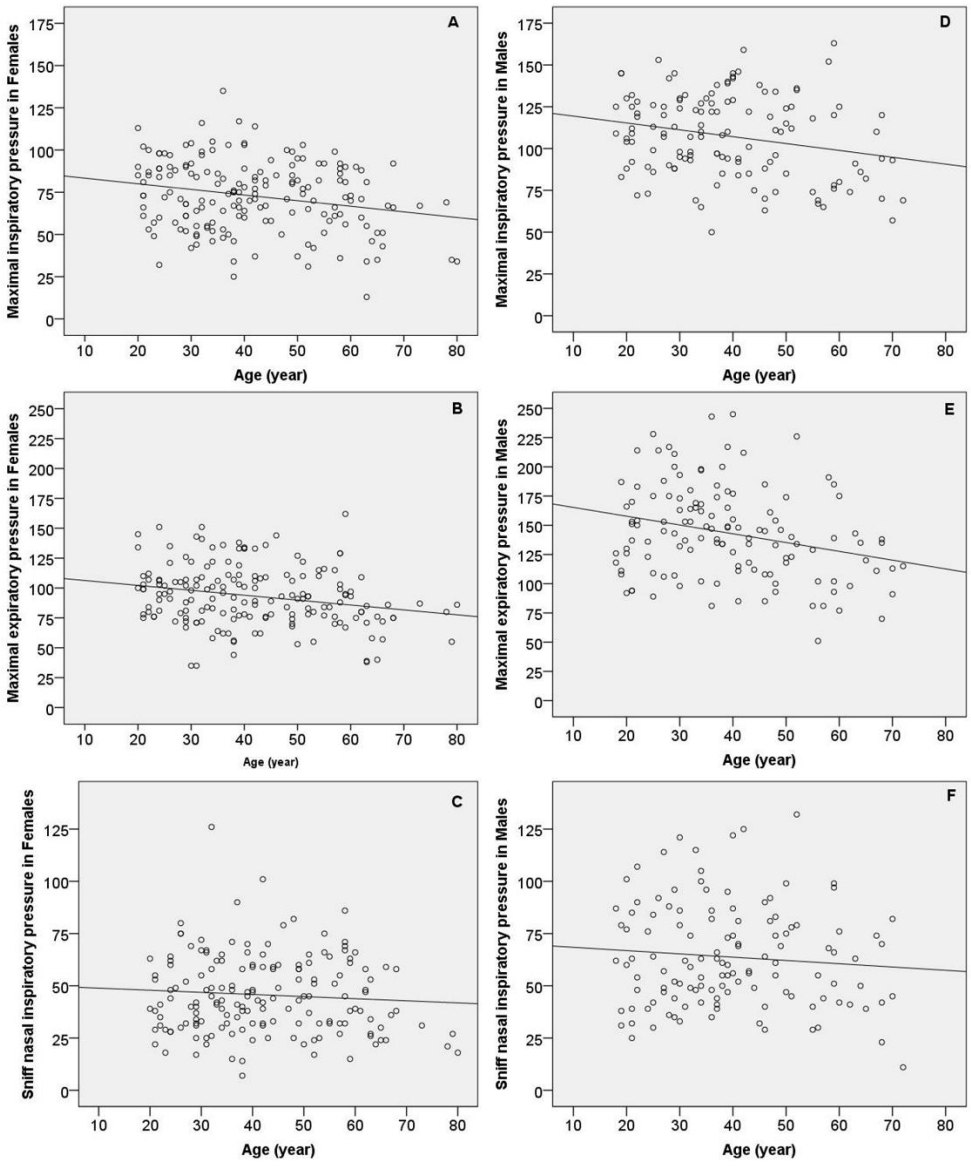


Figure 1. The plots of maximal inspiratory pressure, maximal expiratory pressure, and sniff nasal inspiratory pressure versus age in females (A, B, C) and males (D, E, F).

Table 4. Predictive equations

Model	Equations	LLN*	Adjusted R^2 (%)	SEE	p	No.
MIP:						
1	MIP = 163.4 – 27.2 x gender + 14.3 x FEV1 – 62.2 x height	-35	43	21.6	0.001	301
2	MIP = 68.6 – 22.3 x gender + 11.5 x FEV1	-35	41.4	21.8	0.001	
3	MIP = 95.2 – 32.1 x gender – 0.41 x age + 1.2 x BMI	-34	41	21.9	0.001	
4	MIP = 121.7 – 33.7 x gender -0.37 x age	-34	39.8	22.1	0.001	
5	MIP = 107.4 – 34.7 x gender	-34	36.3	22.7	0.001	
MEP:						
1	MEP = 211.7 – 38.2 x gender + 20.8 x FEV1 -82.7 x height	-49	44.6	30.2	0.001	301
2	MEP = 85.6 – 31.7 x gender +17.1 x FEV1	-49	43.5	30.4	0.001	
3	MEP = 135.1 – 46.75 x gender -0.6 x age + 1.34 x BMI	-50	41.8	30.8	0.001	
4	MEP = 165 – 48.6 x gender – 0.55 x age	-50	41.1	30.9	0.001	
5	MEP = 143.3 – 50.1 x gender	-50	37.5	31.9	0.001	
SNIP:						
1	SNIP = 40.95 – 10.8 x gender +5.5 x FVC	-30	17.0	21.1	0.001	289
2	SNIP = 63.8 – 18.16 x gender	-30	14.8	21.4	0.001	

MIP: maximal inspiratory pressure, MEP: maximal expiratory pressure, SNIP: sniff nasal inspiratory pressure, BMI: Body mass index, FEV1: Forced expiratory volume in the first second, FVC: Forced vital capacity, SEE: standard error of the estimate, No.: number of subjects.

Gender: Female = 1; Male = 0; Height in meters; weight in kg; FVC in L; FEV1 in L.

*LLN: the lower limit of the normal range was defined by subtracting the fifth percentile negative residuals from the predictive value

Discussion

This study investigated the gender and age-related effects on MIP, MEP, and SNIP in a Southeast Asian adult population. The predictive equations for MIP, MEP, and SNIP were established. The findings were consistent with the well-established evidence of greater respiratory muscle strength in males than in females [196].

This study contributes data on MIP and MEP in a Southeast Asian population. Compared to previous studies, the MIP and MEP values in this study are higher than those reported for Malaysian, Chinese, and Indian populations in Singapore [197]. The MIP and MEP values reported in this study were lower than those observed in Caucasian populations but higher than those reported in Indian populations [149,198], suggesting that respiratory muscle strength may not be uniform across different ethnic backgrounds and justifying the need for ethnic-specific values. Although our study reported higher pressures in MIP (males: 107 cmH_2O versus 75 cmH_2O; females: 72.6 cmH_2O versus 48.8 cmH_2O) and MEP (males: 143 cmH_2O versus 93.4 cmH_2O; females: 93 cmH_2O versus 60.7 cmH_2O) compared to a previous study on Indians, this does not imply that the Southeast Asian population has superior respiratory muscle strength [199]. The Indian study's low repetition count (3) may have prevented participants from achieving their peak performance, given the typical recommendation of 5 repetitions [149,171].

For SNIP, the participants exhibited lower values compared to Caucasians (males: 63 cmH_2O versus 104 cmH_2O; females: 45.6 cmH_2O versus 84.8 cmH_2O) [200] and a Brazilian cohort (males: 63 cmH_2O versus 114 cmH_2O; females: 45.6 cmH_2O versus 92.6 cmH_2O) [201]. When compared to other Asian populations, all the SNIP measurements in this study also tended to be lower than those previously reported in Japan [202]. However, the study on Japanese (males: 63 cmH_2O versus 76 cmH_2O; females: 45.6 cmH_2O versus 60 cmH_2O) did not include subjects ≥ 70 years old, potentially driving the values up. Similarly, a study on Taiwanese subjects reported

higher SNIP [203], but the data are not directly comparable due to differences in measurement protocols. The disparate SNIP assessment methodologies employed across the referenced studies, involving the occlusion or non-occlusion of the contralateral nostril, contributed to the variability observed in the reported SNIP values. The latest ATS/ERS recommendations required leaving the contralateral nostril open, as in this study, which potentially lowered the pressure values.

The observed variability in respiratory muscle strength across different ethnic groups highlights the need for population-specific reference values and diagnostic criteria for respiratory muscle weakness. The cut-off points for respiratory muscle weakness recommended by the ATS/ERS may not accurately reflect the normative data for non-Caucasian populations, such as the population examined in this study. One hypothesis is that Caucasians have a higher lung volume than the Asian population [204], leading to a greater demand for muscle recruitment, thus higher MIP and MEP values. This suggests that applying these generic guidelines to diverse ethnic groups could lead to inaccurate diagnoses and suboptimal clinical management. Further research is needed to establish appropriate reference standards that consider ethnic differences in respiratory muscle function, ensuring reliable assessment and care for people from diverse backgrounds.

Our study found that gender had a more pronounced effect on MIP and MEP than age groups. While the effect size of gender was moderate for MIP and MEP, and small for SNIP, the effect size of the age group was small across all three outcomes. The reason for this can be the weak negative correlation observed between these three values and age (Table 3). Additionally, the limited effect of the age group may be attributed to the approach used to subdivide the age groups. Similar to previous investigations, this study grouped participants based on decimal age rather than the true chronological age regarding respiratory muscle strength changes. This subgrouping method may have obscured the true relationship

between age and respiratory muscle strength, resulting in less reliable statistical findings due to the disregard of age-specific patterns of change [205,206]. For example, the non-linear nature of MIP/MEP reduction, with a slower decline in younger age groups and a more pronounced decrease above 50 years old, may reduce the sensitivity of the ANOVA test in assessing the effect size of age. This study did not detect a significant age-gender interaction for MIP, MEP, and SNIP, which is consistent with the results of a previous study [207]. Further studies are required to elucidate the complex interplay of biological and chronological influences on the maintenance of respiratory muscle function across the lifespan. These results highlighted a declining trend in MIP and MEP with age, whereas SNIP tended to plateau in both genders (Figure 1), as illustrated by the linear regression (Table 4). The tendency of an inverse relationship between MIP and MEP, and age, is consistent with existing evidence on age-related changes in respiratory muscle strength [208]. Interestingly, the female participants exhibited a later reduction in MIP and MEP, with significant decreases observed in the above-60 age group, while changes already occurred in younger age groups for males (Table 2). However, a longitudinal study would be better suited to verify the influence of age on these variables [209].

In contrast to studies on Caucasian and Latino populations [200,201], this study found no significant decrease in SNIP with aging for both genders. In the Asian population, there are discrepancies about the age effect, with one study on Taiwanese individuals showing no change in SNIP with aging [203], while another on Japanese subjects reported a decrease in elderly males but not females [202]. The reason for the difference between SNIP and MIP/MEP trends with aging is unclear. One potential hypothesis is that individuals with weaker baseline respiratory muscle function may experience a lower absolute loss in muscle strength than those with stronger baseline function, despite similar relative losses. This is because weaker muscles may have less absolute capacity to lose, even if the relative loss is comparable across individuals [210]. Additionally, unlike MIP, the

SNIP is more influenced by diaphragm performance, which remains stable with aging [211].

Adding FEV1 in predictive equations increased the explained variance in MIP and MEP, which could be reasonable in healthy individuals due to the effort dependence of FEV1. However, in patients with airflow obstruction or restriction, FEV1 may be more influenced by the airways' mechanical properties rather than the forced maneuver [212]. Moreover, it requires the measurement of lung function, increasing the cost and time. Therefore, the equation with only gender, age, and BMI may be more appropriate for practical and clinical use, as it does not have a significant difference in adjusted R^2 compared to the equation with FEV1. Using the gender, age, and BMI equation, MIP and MEP are predicted to annually decrease by 0.41 and 0.60 cmH2O, respectively, as long as BMI remains unchanged. Compared to the predicted values, the descriptive values of MIP and MEP decrease less before the age of 50 and more rapidly after 50. This suggests that the same absolute change may be more significant and clinically relevant before the age of 50. Compared to previous established equations, our study established equations for MIP and MEP with relatively good adjusted R^2 values (Table 5). When compared to the equation for Malaysians, we have more subjects and a higher explained variance [197]. Additionally, the study on Malaysians was conducted almost 20 years ago, which raises the possibility of variations in measurement techniques and equipment that could impact the comparability of results. In contrast, the predictive equation for sniff nasal inspiratory pressure had a low adjusted R^2. Interestingly, our study showed a similar R^2 value compared to previous equations, except for a study by Urdry and Fitting (Table 5). Most studies show the adjusted R2 for SNIP below 20%. This may be attributed to the larger inter-individual variability in muscle activation patterns during sniffing, which reduced the explained variance [213]. Given this limited predictive power, clinicians should be cautious in using SNIP equations as diagnostic tools. Instead, these equations may be more appropriate as exploratory

indicators, particularly in specific clinical contexts where SNIP may complement other assessments. Among respiratory muscle pressures, more discrepancies in SNIP results, as well as measurement protocols, were observed than in MEP and MIP [149,202,203,208]. Like other studies, we found a moderate correlation between SNIP and MIP [149]. This study also provided the LLN, which can be used to identify respiratory muscle weakness. Identifying values below the lower limit of normal can help guide treatments such as respiratory muscle training. Awareness of respiratory muscle weakness can aid in managing risks, especially for individuals prone to complications, such as persons with HF, COPD, or difficulty weaning off a ventilator [139,167,214]. Considering both the LLN and the expected range of respiratory muscle strength for a given age can refine clinical applications, such as tailoring respiratory muscle training or rehabilitation programs to address specific deficits while accounting for age-related physiological changes. While it may be an overstatement to definitively claim that our predictive equations can be applied to Southeast Asian countries, the use of anthropometric indices as predictive factors, which are suggested to be similar across Southeast Asia [215]. Given the limited data available on respiratory muscle strength in this region, the results of this study, which employed the most up-to-date measurement protocols recommended by the ATS/ERS, may be useful to other Southeast Asian populations.

This study has several limitations. First, the optimal method for assessing sniff nasal inspiratory pressure remains unclear due to varying measurement approaches. Additionally, due to the non-smoking criteria, the small number of male participants aged 61 and above may have limited the ability to accurately characterize age-related trends in respiratory muscle strength for this demographic.

Table 5. Predictive equations comparison

Reference	Equations	LLN*	Adjusted R^2 (%)	SEE	No.	Population
	MIP					
This study	95.2 – 32.1 x gender – 0.41 x age + 1.2 x BMI	-34	41	21.9	301	Vietnamese
	121.7 – 33.7 x gender -0.37 x age	-34	39.8	22.1	301	Vietnamese
Johan et al., 1997	151.32 - 0.33A - 0.55 x height + 0.38 x weight	NR	21.9 †	NR	42	Malaysian (Female)
	52.48 + 0.18A - 0.09 x height + 0.12 x weight	NR	14.4 †	NR	69	Malaysian (Male)
Gopalakrishna et al., 2011	278.53 – 1.23 × height + 1.60 × weight – 3.80 × BMI – 0.27 × age	NR	NR	NR	125	Indian (Male)
	379.62 – 2.34 × H + 4.67 × weight – 9.83 × BMI – 0.00 × age	NR	NR	NR	125	Indian (Female)
Bruschi et al., 1992	4.02 – 0.26 x gender – 0.004 x age + 0.47 x body surface area	NR	46.0	21.3	669	Caucasian
Neder et al., 1999	155 – 0.8 x age	NR	42.4	17.3	50	Latino (Male)
	110 – 0.49 x age	NR	46.5	9.1	50	Latino (Female)
Black and Hyatt., 1969	143 – 0.55 x age	NR	NR	NR	60	Caucasian (Male)
	104 – 0.51 x age	NR	NR	NR	60	Caucasian (Female)
Evans et al., 2009	120 – 0.41 x age	62 – (0.15 x 2)	NR	NR	NR	Caucasian (Male)
	108 – 0.61 x age	62 – (0.5 x 2)	NR	NR	NR	Caucasian (Female)

	MEP					
This study	135.1 – 46.75 x gender -0.6 x age + 1.34 x BMI	-50	41.8	30.8	301	Vietnamese
	165 – 48.6 x gender – 0.55 x age	-50	41.1	30.9	301	Vietnamese
Johan et al., 1997	109.82 + 0.05A - 0.22 x height + 0.30 x weight	NR	14.6 †	NR	42	Malaysian (Female)
	181.87 - 0.16A - 0.90x height + 0.43 x weight	NR	24.4 †	NR	69	Malaysian (Male)
Gopalakrishna et al., 2011	566.98 – 2.85 × height + 3.29 × weight – 7.13 × BMI – 1.04 × age	NR	NR	NR	125	Indian (Male)
	178.49 – 0.88 × height + 2.19 × weight – 3.65 × BMI – 0.42 × age	NR	NR	NR	125	Indian (Female)
Bruschi et al., 1992	4.54 – 0.35 x gender – 0.003 x age + 0.24 x body surface area	NR	27	33.3	669	Caucasian
Neder et al., 1999	165 – 0.81 x age	NR	47.7	15.6	50	Latino (Male)
	115 – 0.61 x age	NR	47.9	11.2	50	Latino (Female)
Black and Hyatt., 1969	268 – 1.03 x age	NR	NR	NR	60	Caucasian (Male)
	170 – 0.53 x age	NR	NR	NR	60	Caucasian (Female)
Evans et al., 2009	174 – 0.83 age	117 – (0.83 age)	NR	NR	NR	Caucasian (Male)
	131 – 0.86 age	95 – (0.57 age)	NR	NR	NR	Caucasian (Female)
	SNIP					
This study	SNIP = 63.8 – 18.16 x gender	-30	14.8	21.4	289	Vietnamese

Araujo et al., 2012	-0.47 x age + 135.6	-69.2	9.0	NR	111	Brazilian (Male)
	-0.36 x age + 110.1	-61.9	10	NR	132	Brazilian (Female)
Kamide et al., 2009	-0.67 x age + 104.65	-32.9	14.0	NR	112	Japanese (Male)
	2.31 x BMI + 10.26	-28.8	9.0	NR	111	Japanese (Female)
Uldry and Fitting, 1995	-0.42 x age + 126.8	(-0.42 x age) + 126.8 – 39.0	0.30 †	NR	80	Caucasian (Male)
	-0.22 x age + 94.9	(-0.22 x age) + 94.9 – 28.0	0.23 †	NR	80	Caucasian (Female)
Huang et al., 2014	21.10 + 1.24 x weight	60.3	16.0	NR	58	Taiwanese (Male)
	19.44 + 5.65 x BMI – 2.06 x Body Fat%	52.0	18.0	NR	61	Taiwanese (Female)

MIP: maximal inspiratory pressure, MEP: maximal expiratory pressure, SNIP: sniff nasal inspiratory pressure, BMI: Body mass index, SEE: standard error of the estimate, No.: number of subjects; NR: Not reported

*LLN: the lower limit of the normal

†: The study reported R instead of adjusted R-squared

Conclusion

This study provides ethnic-specific predictive equations for MIP, MEP, and SNIP, which may serve as a preliminary step toward developing reference values for the broader Southeast Asian region. Subtracting 34 cmH_2O for MIP and 50 cmH_2O for MEP from the suggested predictive values will establish the LLN. The declining trend with aging was observed in MIP and MEP but not in SNIP. However, our study reported lower mean values for these outcomes than for other ethnicities. This issue should be addressed carefully to prevent dysfunctions and risks of muscle deterioration.

4. The Alternation method for inspiratory muscle training and measurement.

To effectively measure and train the respiratory muscles, specialized measurement and training equipment, along with trained personnel, are essential. However, access to these devices is not always readily available in Low- and Middle-Income Countries, including Vietnam. This scarcity of resources can hinder the implementation of effective respiratory muscle training programs. Therefore, a simple, alternative method for performing inspiratory muscle training is needed to overcome these limitations. Among the available monitoring tools for training, the Borg CR10 scale is a potential solution, serving as a subjective measure of perceived exertion that does not require specialized equipment and can be used to determine the intensity of aerobic and resistance training. Consequently, a study was conducted to investigate the validity and reliability of the Borg CR10 scale in monitoring the intensity of IMT to provide a practical and accessible alternative for monitoring training intensity in resource-limited settings.

Can We Use the Borg Scale to Identify the Intensity of Inspiratory Muscle Training and Estimate the Maximal Inspiratory Pressure?

Ngoc-Minh Nguyen; Thi-Quynh-Nhu Do; Diem-Quynh Nguyen Tran; Ngoc-Tram Thi Vo; Chung Huynh; Gregory Reychler

Status: Under review – Journal of Physiotherapy

Abstract

Objective: This study aimed to investigate the validity and reliability of the Borg CR10 scale to identify the intensity of inspiratory muscle training (IMT). The study also aimed to determine the Borg score corresponding to 30% maximal inspiratory pressure (MIP) and establish a predictive model for %MIP using the Borg score.

Methods: The corresponding %MIP at different Borg CR10 ratings was determined through two assessments. First, MIP was measured, then the inspiratory pressures at each Borg rating were measured and converted to %MIP. A second assessment was conducted one week later to evaluate the reliability. Validity was analyzed using Spearman correlation, reliability with Intraclass correlation, and measurement errors with Bland-Altman. Receiver Operating Characteristic (ROC) analysis identified the Borg score for 30% MIP, and curve fit analysis established predictive models for %MIP using the Borg score.

Results: The results showed significant positive correlations between %MIP and Borg score ($r = 0.79$). The Borg CR10 exhibited excellent reliability (ICC = 0.88), although reliability varied across the scale. There was a small systematic error (<5% MIP) in most Borg levels. ROC analysis indicated a Borg 3-4 corresponds to 30% MIP, and linear, cubic, and quadratic models reliably predict %MIP from the Borg score.

Conclusion: The Borg CR10 scale can estimate MIP and identify IMT intensity, providing a practical approach for clinicians to assess and adjust respiratory muscle training.

Keywords: Borg CR10, maximal inspiratory pressure assessment, inspiratory muscle training.

Introduction

Inspiratory muscle training (IMT) has become an increasingly popular intervention for various health conditions, such as chronic obstructive lung disease [216], heart failure [139], mechanical ventilation [217], and neuromuscular diseases [218]. Similar to all muscle training, the training intensity has to be determined [219]. The current guidelines recommend the intensity of the training to be between 30% and 80% of the maximal inspiratory pressure (MIP) [139,216,220]. To conduct the IMT, the MIP has to be measured using a respiratory pressure meter [149]. Then, the acquired MIP is used to determine the intensity of the IMT with a training device [221]. Furthermore, this process may need to be repeated weekly during the training to adjust the resistance and optimize the benefit [222]. However, the assessment and training device may not be readily accessible, with high costs and healthcare service dependence accompanying it. Therefore, simple and practical methods to identify the training intensity for IMT are necessary to enable patients to train optimally. Among intensity anchor methods, the Borg CR10 rating of perceived exertion scale, recommended by the American Thoracic Society for exercise prescription [223], is a simple and cost-free tool for monitoring intensity during aerobic [224] or resistance training [225]. However, it has never been investigated to identify the IMT intensity. Thus, this study aimed to investigate the validity and reliability of the Borg CR10 scale to identify the intensity of the IMT. The second goal was to determine the Borg score level that corresponds to the minimum recommended intensity of 30% MIP and establish a predictive model for the percentage of MIP (%MIP) using the Borg score.

Methods

Ethics statement

This study was approved by the Ethics Committee in Biomedical Research at the University of Medicine and Pharmacy in Ho Chi Minh City, with approval number 725/ĐHYD-HĐĐĐ dated 12/02/2025. All participants provided informed consent to take part in the study. The research was conducted following the Declaration of Helsinki for medical research involving human subjects [145].

Participants

The study recruited volunteer adults ≥ 18 years old. Individuals were excluded if they had spontaneous pneumothorax, eardrum injury, or unstable asthma [226]. Following the COSMIN guidance, the minimum sample size for validity and reliability is 50 participants [2].

Protocol

The study was conducted at the Rehabilitation Department, University of Medicine and Pharmacy at Ho Chi Minh City. The study comprised two visits scheduled one week apart.

During the first visit, demographic data were collected. The participants were explained the Borg CR10 scale [227]. The scale ranged from zero to 10, with zero classified as "no perceived effort" and 10 as "maximal perceived effort". Then, participants were instructed to inhale through the mouth three times at maximum effort, with one-minute rest periods between trials, to warm up and become accustomed to the maneuver. Participants were asked to anchor their perception of maximal inspiratory effort with a Borg score of 10. Subsequently, the MIP measurement was performed using a portable respiratory pressure meter (RP Check, MD Diagnostic Ltd, UK) with a filter and plastic flanged mouthpiece, following the European Respiratory Society (ERS) guidelines [149]. The measurement was performed at the residual volume (RV). The participants were instructed to inhale through the mouth with maximal effort. The maximal pressure needs to be maintained for at least 1.5

seconds. The measurement was performed in an upright sitting position with two feet on the floor, following the guidance of ERS [149]. Each measurement was repeated at least 5 times, with a 1-minute rest between trials, until the best three times varied within 10% [149]. The highest value was recorded as MIP. Finally, the corresponding %MIP at different Borg scores was measured. The participants were asked to breathe in with different effort levels corresponding to the Borg scores from 10 to 1, with a 1-minute rest between trials. The corresponding inspiratory pressures were measured using the same respiratory pressure meter for each Borg CR10 scale rating. These inspiratory pressures were then converted to percentages of recorded MIP to establish the corresponding %MIP at each Borg score (from 1 to 10). There was a 15-minute resting period between measurements. All assessors were trained and coached before collecting data. During both measurements, the participants were blinded to the outcomes.

In the second visit, the procedural protocol mirrored that of the first visit precisely. However, a key modification was implemented: the measurement of the corresponding %MIP at different Borg scores was conducted prior to the MIP measurement. This deliberate reversion in the order of procedures was strategically designed to mitigate any potential carry-over effect that the MIP measurement might have on the participants' subjective anchoring of the %MIP values relative to the Borg CR10 scale ratings.

Data analysis

The data analysis was performed using SPSS version 29.0. Descriptive statistics were used to summarize the participants' characteristics with mean and standard deviation or medians and interquartile ranges.

The validity of the Borg CR10 scale to quantify the level of effort in IMT was confirmed with Spearman’s correlation coefficients between the Borg scores and the corresponding %MIP of both visits [2]. The correlation was classified as low when $r < 0.5$, moderate when

between 0.5 and 0.69, and high when ≥ 0.7 [228]. The distribution of corresponding %MIP to each Borg score was visualized using box plots. The paired t-test was used to investigate the difference in corresponding %MIP at each Borg score between the two visits. The one-way ANOVA test was used to examine the differences in %MIP between different Borg scores in each visit [229].

The test-retest reliability of the corresponding %MIP at different Borg scores was tested using the ICC with a two-way mixed, absolute agreement, single-measurement model [2]. The ICC values greater than 0.5 are acceptable, values over 0.75 suggest the test has good reliability, and those above 0.9 are considered excellent [152]. The level of agreement was determined with the Bland-Altman plot to evaluate the measurement errors [230]. The linear regression was performed to check for proportional bias [230].

To identify the Borg score threshold corresponding to the minimal recommended IMT intensity of 30% MIP [139,216], a Receiver Operating Characteristic (ROC) curve analysis was performed [231]. This analysis identified the optimal cutoff Borg score using a 30% MIP threshold, with the acceptable sensitivity and specificity set to exceed 70% [232]. Values above 30% MIP were considered positive, and values less than 30% MIP were considered negative. The acceptable area under the curve (AUC) is above 0.8 [233]. Additionally, the curve estimation analysis was conducted to investigate the explanatory model for predicting %MIP based on the Borg score. The statistical significance was set at $p < 0.05$.

Result

Demographic

A total of 50 participants (female/male: 27/23) were recruited for the study. The demographic characteristics are shown in Table 1.

Table 1. Characteristic of participants (N = 50)

	Mean (SD)	Range
Age (year)	31.1 (12.4)	20.0 – 63.0
Weight (Kg)	57.1 (10.6)	40.0 – 90.0
Height (m)	1.61 (0.1)	1.47 – 1.93
BMI	21.8 (2.5)	16.4 – 26.4
MIP (cmH_2O)	81.4 (29.3)	35.0 – 161.0

Validity and the distribution of the percentage of maximal inspiratory pressure at different Borg scores

There were significant positive correlations between corresponding %MIP and Borg score for the first ($r = 0.79$, $p = 0.001$) and second visit ($r = 0.80$, $p = 0.001$). The corresponding %MIP values at each Borg score are demonstrated in Figure 1 for both visits. The descriptive data for the corresponding %MIP at each Borg score are shown in Table 2. There was no significant difference in %MIP at each Borg score between the two visits (all $p > 0.05$). The ANOVA test revealed that, except for the Borg scores of 9 and 10, there were no statistically significant differences between the two continuous Borg scores in %MIP in both visits (Table 3).

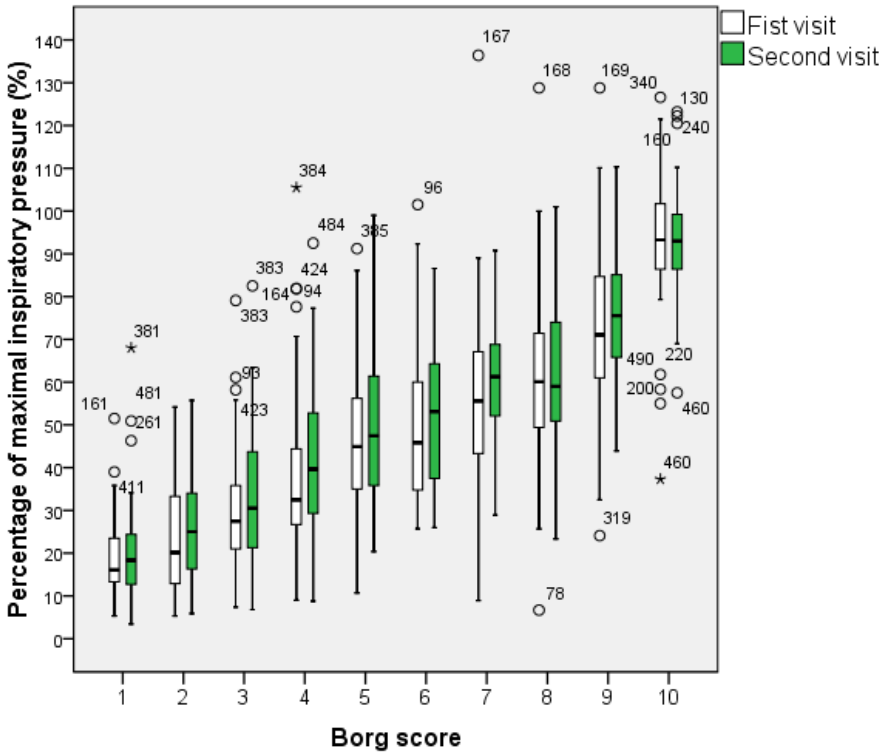


Figure 1. The percentage of maximal inspiratory pressure at each Borg score across two visits

Table 2. The distribution of the %MIP at different Borg scores

	First visit (N = 50)		Second visit (N = 50)		Difference in %MIP between the two visits	
	Mean (SD)	95% CI	Mean (SD)	95% CI	Mean (SD)	95% CI
%MIP-Borg 1	18.9 (9.0)	16.4 - 21.5	20.3 (12.1)	16.9 - 23.8	1.4 (12.0)	-2.0 – 4.8
%MIP-Borg 2	23.4 (13.0)	19.7 - 27.0	26.3 (12.2)	22.9 - 29.8	3.0 (12.6)	-0.6 – 6.6
%MIP-Borg 3	29.9 (14.4)	25.8 - 34.0	33.2 (15.9)	28.7 - 37.8	3.3 (13.5)	-0.5 – 7.1
%MIP-Borg 4	38.0 (19.5)	32.5 - 43.5	42.4 (17.9)	37.3 - 47.5	4.4 (18.8)	-0.9 – 9.8
%MIP-Borg 5	46 (16.5)	41.3 - 50.7	50.2 (18.6)	44.9 - 55.4	4.2 (17.1)	-0.7 – 9.0
%MIP-Borg 6	48.4 (18.0)	43.7 - 54.0	53.5 (16.0)	49.0 - 58.1	4.7 (19.7)	-0.9 – 10.3
%MIP-Borg 7	55.9 (21.1)	49.9 - 61.9	60.7 (14.5)	56.5 - 64.8	4.7 (21.2)	-1.3 – 10.8
%MIP-Borg 8	60.2 (19.7)	54.6 - 65.8	60.7 (17.1)	55.8 - 65.6	0.5 (23.4)	-6.1 – 7.1
%MIP-Borg 9	72.0 (19.8)	66.3 - 77.6	74.7 (14.4)	70.6 - 78.8	2.7 (22.3)	-3.6 – 9.0
%MIP-Borg 10	92.9 (16.1)	88.3 - 97.5	92.9 (13.4)	89.1 - 96.7	0.1 (16.6)	-4.6 – 4.8

%MIP-Borg: the corresponding percentage of maximal inspiratory pressure at different Borg scores.
N: the number of values for each Borg score
CI: Confident interval

Table 3. The difference in the corresponding percentage of maximal inspiratory pressure across the Borg CR10 scale

	First visit										Second visit									
%MIP Borg	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
1	-	○	♦	♦	♦	♦	♦	♦	♦	♦	-	○	♦	♦	♦	♦	♦	♦	♦	♦
2	○	-	○	♦	♦	♦	♦	♦	♦	♦	○	-	○	♦	♦	♦	♦	♦	♦	♦
3	♦	○	-	○	♦	♦	♦	♦	♦	♦	♦	○	-	○	♦	♦	♦	♦	♦	♦
4	♦	♦	○	-	○	○	♦	♦	♦	♦	♦	♦	○	-	○	♦	♦	♦	♦	♦
5	♦	♦	♦	○	-	○	○	♦	♦	♦	♦	♦	♦	○	-	○	♦	♦	♦	♦
6	♦	♦	♦	○	○	-	○	♦	♦	♦	♦	♦	♦	♦	○	-	○	○	♦	♦
7	♦	♦	♦	♦	○	○	-	○	♦	♦	♦	♦	♦	♦	♦	○	-	○	♦	♦
8	♦	♦	♦	♦	♦	♦	○	-	♦	♦	♦	♦	♦	♦	♦	○	○	-	♦	♦
9	♦	♦	♦	♦	♦	♦	♦	♦	-	♦	♦	♦	♦	♦	♦	♦	♦	♦	-	♦
10	♦	♦	♦	♦	♦	♦	♦	♦	♦	-	♦	♦	♦	♦	♦	♦	♦	♦	♦	-

○: No statistically significant difference

♦: Statistically significant difference

%MIP Borg: the corresponding percentage of maximal inspiratory pressure at Borg score

Reliability

The ICC was 0.87, with a mean bias of 2.9% and the limits of agreement ranging from -38.2% to 32.4%. However, the depth analysis demonstrated different levels of reliability across the Borg CR10 scale. The limit agreements are shown in Table 4. There is no proportional bias for the Borg scores with statistically significant reliability (all p values of linear regression > 0.05).

The Receiver Operating Characteristic analysis

The AUC of the ROC analysis with the cut-off threshold of 30% MIP was 0.89 (95% IC: 0.85 – 0.92) and 0.90 (95% IC: 0.87 – 0.93) for the first and second visit, respectively (Figure 2). The coordinates of the curve are shown in Table 5.

The curve estimation analysis and the explained model for predicting %MIP based on the Borg score

The curve estimation analysis was conducted for the first and second visits. The results indicated that the linear, cubic, and quadratic models could all be used to explain the relationship between %MIP and the Borg score. The adjusted R-squared values were relatively consistent across the three models (linear, cubic, and quadratic) and between the two visits (Table 6).

Table 4. The reliability of the corresponding percentage of maximal inspiratory pressure at each Borg score

	ICC	95% IC	Bias	Upper limit	Lower limit
%MIP-Borg 1	0.53*	0.18 – 0.74	1.4	24.9	-22.1
%MIP-Borg 2	0.66*	0.40 – 0.80	3.0	27.7	-21.7
%MIP-Borg 3	0.75*	0.56 – 0.86	3.3	29.8	-23.2
%MIP-Borg 4	0.65*	0.40 – 0.81	4.4	41.2	-32.4
%MIP-Borg 5	0.68*	0.46 – 0.82	4.2	37.7	-29.3
%MIP-Borg 6	0.49*	0.12 – 0.71	4.7	43.3	-33.9
%MIP-Borg 7	0.48*	0.09 – 0.70	4.7	46.3	-36.9
%MIP-Borg 8	0.33	-0.20 – 0.63	0.5	46.4	-45.4
%MIP-Borg 9	0.30	-0.24 – 0.60	2.7	46.4	-41.0
%MIP-Borg 10	0.55*	0.20 – 0.75	0.06	32.6	-32.5

*: $p < 0.05$

ICC: intraclass correlation coefficient (absolute agreement)

%MIP-Borg N (N:1-10): corresponding %MIP at each Borg score

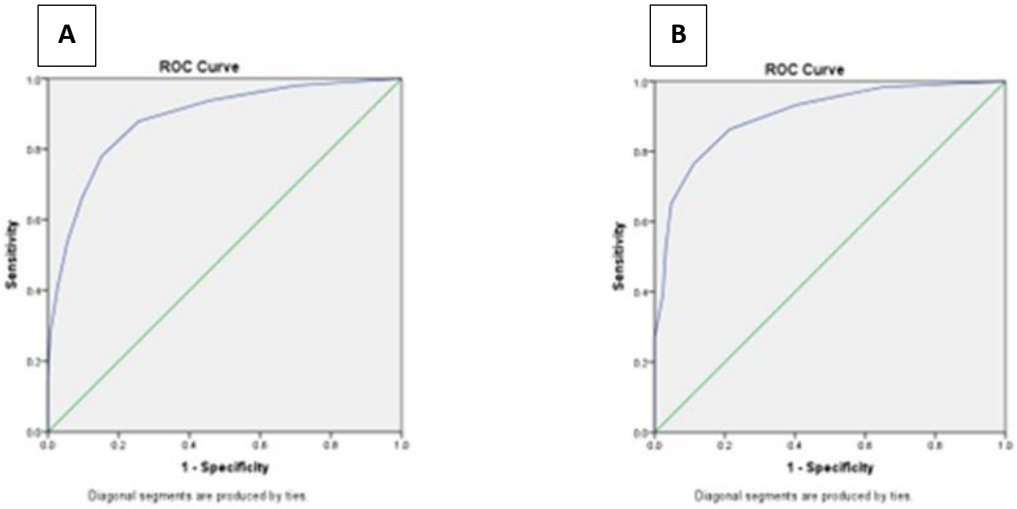


Figure 2. The ROC analysis: A: first visit (AUC = 0.89); B: Second visit (AUC = 0.9)

Table 5. Coordinates of the curve

Positive of archive > 30% MIP if greater than or equal to the Borg score of	First visit		Second visit	
	Sensitivity	1 - Specificity	Sensitivity	1 - Specificity
1.5	0.98	0.70	0.98	0.65
2.5	0.94	0.46	0.93	0.41
3.5	0.88	0.26	0.86	0.21
4.5	0.78	0.15	0.77	0.11
5.5	0.66	0.10	0.65	0.05
6.5	0.54	0.06	0.52	0.03
7.5	0.41	0.03	0.39	0.02
8.5	0.28	0.007	0.27	0.001
9.5	0.14	0.001	0.13	0.001

Table 6. Predictive models for %MIP using Borg score					
Model	Equations	Adjusted R^2 (%)	SEE	p	No.
Linear	%MIP = 8.2 + 7.4 x Borg-score	59	17.6	0.001	First visit
Quadratic	%MIP = 17.3 + 2.8 x Borg-score + 0.4 x Borg-score²	60	17.3	0.001	
Cubic	%MIP = 4.1 + 14.2 x Borg-score - 2.1 x Borg-score² + 0.15 x Borg-score³	61	17.1	0.001	
Linear	%MIP = 11.9 + 7.2 x Borg-score	63	15.8	0.001	Second visit
Quadratic	%MIP = 17.2 + 4.5 x Borg-score + 0.2 x Borg-score²	64	15.7	0.001	
Cubic	%MIP = 3.8 + 16.4 x Borg-score - 2.3 x Borg-score² + 0.16 x Borg-score³	65	15.5	0.001	
%MIP: percentage of maximal inspiratory pressure, SEE: standard error of the estimate, No.: number of values.					

Discussion

The results indicate that the Borg scale is a valid method for anchoring the intensity of inspiratory muscle training, with high, similar correlations observed across both visits (r = 0.79 and 0.80). The results align with findings from a systematic review that examined using the Borg scale to determine the intensity of resistance training for limb muscles, which reported a good validity coefficient (r = 0.88) [234]. Additionally, the participants performed two measurements unaware of the intensity, reinforcing the good validity of anchoring the %MIP with the Borg score. However, in both visits, a warm-up with 3 repetitions of maximal inspiratory effort was conducted to anchor the maximal effort, suggesting that a warm-up with maximal effort may be necessary to anchor the effort, like the ERS recommendation [149].

The box plot demonstrated a clear trend of increasing %MIP proportional to the rise in Borg score. However, the relationship between the Borg score and the corresponding %MIP was not straightforward, but rather exhibited a more complex pattern. Specifically, the Borg 3, Borg 4, and Borg 10 corresponded to approximately 30% MIP, 40% MIP, and 100% MIP, respectively. Meanwhile, Borg 1 and 2 were equivalent to around 20% MIP, and Borg 5 through Borg 9 fell within the 45% to 70% MIP range. Interestingly, this pattern is similar to findings from other studies that used the Borg CR10 scale to determine intensity in peripheral muscle resistance training [235] and aerobic training [236]. The reason may be attributed to the ambiguity of the higher half of the Borg CR10 scale, where some Borg scores lack descriptors (Borg scores of 6 and 8), leading to biased responses [237]. At the same time, the descriptions for Borg 7 (very hard) and Borg 9 (very, very hard) can be confusing and overlap. The paired t-test revealed no significant difference in %MIP at each Borg score between the two visits, suggesting no training effect. Noticeably, the ANOVA result (Table 3) recommended that a change of one Borg score may not lead to a significant change in IMT intensity from Borg 1 to Borg 7. Accordingly, changing at least

2 Borg score units may be more clinically meaningful in training. For the Borg scores from 8 to 10, a change of one score unit may be sufficient to result in a real change in the corresponding %MIP. Despite the imperfections, the Borg CR10 scale can consistently anchor the relative intensity of inspiratory muscle training across participants.

In our study, the Borg CR10 scale has excellent reliability in IMT, similar to a previous study on peripheral muscle resistance training (ICC = 0.88) [238]. However, a further analysis revealed varying levels of reliability across the Borg CR10 scale. Most Borg scores (1, 2, 4, 5, and 10) demonstrated acceptable reliability in replicating the corresponding %MIP (ICCs > 0.5) [152]. The Borg score of 3 showed good reliability (ICC =0.75). In contrast, the Borg scores of 6 and 7 exhibited poor reliability (ICCs < 0.5), and the Borg scores of 8 and 9 were unreliable (p > 0.5). The low reliability to unreliability of the Borg scores from 6 to 9 may be attributed to the ambiguous nature of those scale points, as mentioned in the discussion of the scale's validity. To our knowledge, this is the first study to investigate the reliability of each level of the Borg scale, highlighting the importance of considering the methodological approach when investigating the reliability of a multi-level perception scale. The results also showed a small systematic error of less than 5% MIP in most Borg levels (except 8 and 9), which is less than 10% of the accepted error by the ERS guidelines [149]. However, the general random error of the scale was higher, likely due to the subjective nature of the measurement. In summary, the reliability of the Borg scale in determining IMT intensity is acceptable, but the potential for random errors should be aware, particularly for Borg scores from 6 to 9.

The ROC analysis demonstrated good AUC in both visits [233]. The curve coordinates for each Borg score of both visits were almost the same, demonstrating the scale's stability. This study suggested that around 86 to 88% sensitivity (meaning the Borg score correctly identifies this percentage of people at the target intensity) with a Borg

score of 3 or 4 can anchor the intensity of 30% MIP, which is widely recommended as a minimal intensity for IMT [139,216,222].

Although the 30% MIP is commonly used as the initial target intensity for inspiratory muscle training, the appropriate training demand may need to be adjusted based on the individual's underlying health conditions [139,216,220]. Therefore, curve fit analysis was performed to establish predictive models for %MIP using the Borg scale. This study demonstrated that the linear, the cubic, and the quadratic models can all be used to predict the %MIP based on the Borg score, with adjusted R^2 above 60%. Furthermore, the results across the two visits were highly consistent, indicating the stability and reliability of the predictive equations. The slightly higher adjusted R2 in the second visit may be attributed to the participants' improved ability to anchor their effort with repetition. However, the study's methodology cannot investigate this possibility. Nevertheless, the reported level of variance explanation is higher than the generally accepted value of 25% to 40% used in medicine [239]. Compared to another study using the Borg scale to predict heart rate and lactate levels in aerobic training, our study reported a higher adjusted R-squared value [236]. The Borg scale may be more efficient for monitoring inspiratory muscle training than aerobic training, as aerobic training involves a more global response rather than a focused effort, which can be affected by various factors [240]. The standard error of the estimate (SEM) was 15-17%, close to the recommended 10% acceptable error [149], supporting the use of equations to predict the %MIP. While the adjusted R-squared values were similar across the three models, the linear model can be easily computed and adjusted for clinical practice. This finding proposed a new approach for assessing respiratory muscle function at the submaximal levels, particularly for individuals with contraindications to maximal respiratory muscle testing [149]. With a simple device that displays the absolute pressure value at a given Borg CR10 score, along with the corresponding %MIP to that Borg score calculated using the equation, the estimated MIP can be easily calculated. For instance, if a person were asked

to inspire with a device that can quantify the pressure at Borg 5, and the result is 30 cmH2O. Then, using the linear regression equation from this study, the estimated %MIP at Borg 5 would be 8.2 + (7.4 x 5) = 45.2%. Therefore, the estimated MIP for this case would be 30 cmH2O / 0.452 = 66.4 cmH2O.

This study has several limitations. First, the participants were generally healthy adults, so the results may not apply to other populations. For instance, the subjective dyspnea associated with respiratory diseases can impair the anchor effort capacity of the Borg CR10 scale in people with respiratory conditions [241]. Future studies are needed to confirm these findings on other populations. Additionally, the small sample size precluded the establishment of a multiple linear regression model, which could potentially provide a more accurate estimation of the Borg score by incorporating additional parameters like age, height, weight, and gender. Furthermore, because the experiment for anchoring the % MIP with the Borg scale was conducted in order from maximum to minimum score, the study could not investigate the effect of randomization in anchoring the %MIP by Borg score.

In conclusion, this study suggests that the Borg CR10 scale can be used to identify the IMT intensity and estimate the MIP. This provides a practical approach for clinicians to assess and adjust respiratory muscle training at submaximal levels.

Chapter 2: Functional Exercise Capacity Assessment for persons with heart failure

In cardiopulmonary disorders, functional exercise capacity is an excellent indicator of mortality rate, hospitalization, quality of life, and dependency. Therefore, a main role of cardiopulmonary rehabilitation is to monitor the trajectory and improve functional exercise capacity. Functional exercise capacity is fundamentally determined by oxygen transportation and utilization efficiency. The cardiopulmonary exercise test is considered the gold standard for assessing exercise capacity with comprehensive information about the integrated physiological response to exercise. However, due to the high cost and specialized facility requirements, its application in routine practice remains limited, especially in settings like Vietnam. Therefore, the development and validation of more accessible assessment tools to comprehensively evaluate cardiopulmonary function during exercise is needed. Considering the time and infrastructure constraints within the Vietnamese context, the 1-minute sit-to-stand test was selected as the focus for a validation project. Among cardiopulmonary diseases, individuals with heart failure were selected as the target population due to the high prevalence rate, not only in Vietnam but also worldwide.

The Validity, Test-Retest Reliability, and Learning Effect on the One-Minute Sit-To-Stand Test in Individuals with Heart Failure

Ngoc-Minh Nguyen*; Hien Thi Le*; Bich-Ly Thi Ngo; Hung Phi Trương; Gregory Reychler; Gael Deboeck

* Hien Thi Le and Ngoc-Minh Nguyen contributed equally as the co-first authors to this work

Adapted from: Archives of Physical Medicine and Rehabilitation, May 2025

DOI: 10.1016/j.apmr.2025.05.004

Abstract

Objective: This study aimed to evaluate the reliability, the minimally important difference (MID), and the learning effect of the one-minute sit-to-stand test (1STST) in assessing functional exercise capacity in patients with heart failure (HF).

Design: A cross-sectional study

Setting: Two hospitals

Participants: Patients with a HF diagnosis by a cardiologist following the 2021 European Society of Cardiology Heart Failure guideline. Most participants were male (60%). Most participants had normal BMI (62%). All participants were classified as having New York Heart Association Functional Class II or III.

Interventions: Not applicable

Main outcome measures: Each patient performed a 6-minute walk test (6MWT) and two 1STST on two occasions spaced one month apart. Test-retest reliability between occasions was evaluated using intraclass correlation coefficients (ICC), Bland-Altman plot analysis, and linear regression. The minimal important difference (MID) was determined using a distribution-based method. To assess the learning effect, we conducted a paired t-test comparing the two 1STST on each occasion, and the magnitude of the learning effect

was quantified using Cohen's d. The correlation between the 1STST and 6MWT was performed to confirm validity.

Results: 1STST showed good test-retest reliability (ICC = 0.98; p<0.001) with good agreement between the two measurements, without proportional bias (p>0.05). The MID was 1.1 repetitions. The learning effect of 1STST was observed (p<0.001) with a large effect size when repeated a month later (Cohen's d >0.8). The 1STST strongly correlated with the 6MWT (r = 0.74, p = 0,001).

Conclusions: The 1STST is a valid and reliable measure of functional exercise capacity in patients with HF. Due to the learning effect, this study recommended performing two trials to capture the true value.

Keywords: heart failure, 1-minute-sit-to-stand test, reliability, validity, learning effect.

Introduction

Heart failure (HF) is a leading cause of mortality worldwide, with a significant impact on public health [242]. The disease is characterized by the heart's inability to pump blood effectively, resulting in poor exercise tolerance, reduced functional capacity, and diminished quality of life for those affected [243]. Furthermore, the healthcare costs associated with HF are rising significantly, posing a growing burden on healthcare systems [244]. Among the contributing factors, decreased exercise capacity is a strong predictor of morbidity and mortality in patients with HF [245,246]. Therefore, assessing exercise functional capacity is essential for evaluating disease severity, determining prognosis, and guiding treatment in this patient population [246]. Besides, evaluating exercise functional capacity provides insights that can guide the establishment of appropriate rehabilitation interventions.

The cardiopulmonary exercise test is considered the gold standard for assessing physical capacity with comprehensive information about the integrated physiological response to exercise [246]. However,

the test requires specialized equipment with trained staff and is time-consuming, making it impractical for routine clinical use in resource-limited settings [247]. Thus, submaximal tests like the 6-minute walk test (6MWT) are commonly used for measuring functional exercise capacity [248]. However, the 6MWT has several limitations, including the need for a long hallway and the time required to conduct the test [249]. Therefore, the 6MWT challenges resource-constrained settings, where hospital space and staffing might be limited. The one-minute sit-to-stand test (1STST) has emerged as a promising alternative to the 6MWT for patients with HF [250]. This simple assessment of functional exercise capacity may be well-suited for use in settings with limited resources. Previous research found the one-minute sit-to-stand test to be a reliable and valid assessment of functional exercise capacity in patients with HF, without a significant learning effect [251]. However, the reported lack of a learning effect for the 1STST is open to question and debate, given the current evidence on this topic in other populations [252-254]. Additionally, this study assessed the learning effect in participants who were already familiar with the 1STST, which raises skepticism about ruling out the potential for a learning effect and reduces confidence in the reliability finding. Moreover, the minimal important difference (MID) of 1STST in patients with HF has not yet been established. Therefore, this study aimed to expand on previous research by investigating the reliability, MID, learning effect, and learning effect size of the 1STST in patients with HF. The secondary objective was to confirm the validity of the 1STST.

Methods

Ethics statement

This cross-sectional study was approved by the Ethics Committee at University of Medicine and Pharmacy at Ho Chi Minh City with the approval No.1016/HĐĐĐ-ĐHYD. All participants were comprehensively informed about the protocol and provided written consent. The data of this study was securely stored and accessible only with proper authorization. All stages of the study were

conducted following the Declaration of Helsinki and Clinical Good Practice guidelines.

Participants

The patients with HF were recruited from November 2023 to July 2024. They were receiving outpatient treatment at Cardiology Clinics of the University Medical Center Ho Chi Minh City and Vinmec Times City International Hospital (Hanoi), Vietnam. Inclusion criteria were to be above 18 years old with diagnosed HF according to the 2021 ESC Heart Failure guideline [255] by cardiologists, clinically stable, and able to provide informed consent.

Patients were excluded if they had recently experienced a cardiac event within 1 month or if they had a medical record of cognitive impairments or psychiatric disorders that could affect their cooperation. Individuals with contraindications for weight-bearing and walking, as confirmed by a primary care physician, were also excluded. Additionally, participants were not included in the reliability assessment if they experienced changes in their condition between the two 1STST evaluations, such as acute exacerbation, changes in HF classification, medication alterations, participation in a rehabilitation program, or new comorbidities that could impact their physical condition.

Protocol and measurements

During the first visit, demographic data and comorbidities were collected. The participants' New York Heart Association functional class (NYHA) and left ventricular ejection fraction (LVEF), obtained within a month, were extracted from their medical records. The 1STST and 6MWT were performed in a randomized order, following standard protocols [250,256]. To investigate the test-retest reliability, the 1STST was repeated within one month (1STST1 and 1STST2). All tests were conducted by the same investigator in the same setting. Participants who experienced changes in their health conditions, such as medication modifications, shifts in their NYHA class, or any concerns about changes in their health status reported by their

primary care physician or themselves, were excluded from the reliability analysis. Within a visit, the 1STST was performed twice (a and b) to verify the learning effect. Participants were given at least 30 minutes of rest between the tests. The vital signs, including blood pressure, heart rate, oxygen saturation, and Borg's scale, were measured before and after each test.

The one-minute sit-to-stand test

None of the participants had prior experience with the 1STST. They were instructed to sit on a standard chair (46cm in height) without armrests, with their back against the chair, arms crossed on their chest, and feet flat on the floor. They were familiarized with one full cycle of standing up and sitting down. When the test began, participants were asked to stand up fully and then sit down as many times as possible in one minute, without using their arms for support. The number of completed cycles was counted and recorded [150]. The 1STST was performed twice (1STST1a and 1STST1b for the first visit, 1STST2a and 1STST2b for the second visit). The test with the greater number of repetitions was used for the test-retest reliability and validity analysis.

The 6-minute walking test

The 6MWT was used to assess criterion validity. The test was conducted in a 30m flat, level hallway with good lighting and a quiet environment. The participants were instructed to walk with or without walking assistive devices as far as possible within 6 minutes. The distance was recorded [256]. The 6MWT was performed once at the first visit, as the patients were familiar with it from their routine follow-ups.

Statistical analysis

The data analysis was performed using SPSS version 29.0. Descriptive statistics were used to summarize the participants' characteristics. The normality of the data was assessed using the Shapiro-Wilk test [151].

To determine the required sample size, the researchers calculated an expected intraclass correlation coefficient (ICC) of 0.9, set a minimum acceptable ICC of 0.75, and aimed for a power of 80% and an alpha level of 0.05. Anticipating a 20% dropout rate, the final sample size needed was 42 participants [2].

The test-retest reliability of the 1STST was examined using the ICC with a two-way mixed, absolute agreement, single-measurement model [2]. The ICC values greater than 0.5 are deemed satisfactory, values over 0.8 suggest the test has good reliability, and those above 0.9 are considered excellent [152]. The level of agreement was determined with the Bland-Altman plot to evaluate the consistency between the best 1STST in the first and second visits[230]. The linear regression was performed to check for proportional bias [230]. The MID was defined using the distribution-based method, calculated as MID = 0.5 x SD of the change in the number of repetitions between the two 1STST assessments performed during the separate visits [2]. This distribution-based approach provided an estimation of the smallest change in 1STST performance that could be considered meaningful. Besides, the minimal detectable change (MDC) was computed as $1.96 \times SEM \times \sqrt{2}$ [257]. The SEM was calculated using the equation $SEM = SD \times \sqrt{(1 - ICC)}$. The MDC represents the smallest change that exceeds the measurement error, allowing for the assessment of whether the MID is larger than the measurement error and, thus, represents a true change.

The learning effect was evaluated using a paired t-test between two performances within each visit. The effect size was calculated using Cohen's d to determine the learning effect size [258]. The effect size was categorized as small, moderate, and large, with Cohen's d values less than 0.2, between 0.2 and 0.5, and greater than 0.8, respectively [259].

The criterion validity of the 1STST was confirmed with Spearman’s correlation coefficients between the 1STST and the 6MWT [2]. The correlation is typically classified as low when $r < 0.5$, moderate when

between 0.5 and 0.69, and high when 0.7 or greater [154]. Additionally, pre- and post-test physiological adaptation between 6MWT and 1STST were compared using a paired-sample t-test or Wilcoxon signed rank test, based on the distribution of the variables with statistical significance set at $p<0.05$.

Results

Demographic

Forty-seven participants who met the inclusion criteria completed the 6MWT and 1STST test during the first visit. Five patients missed their follow-up appointment, one participant died due to coronary thrombosis, and one experienced an exacerbation before the second visit. In total, forty patients completed all the tests during the second visit. The majority of participants were male (60%) with a normal BMI (62%) (Table 1). The Shapiro-Wilk test revealed that the data for the 1STST and 6MWT were normally distributed ($p > 0.05$). No adverse events were observed during or after the tests, apart from some participants reporting leg fatigue and dyspnoea. The simplicity and detailed instructions of the two tests resulted in high participant compliance without any need for significant correction or repetition.

Table 1. Characteristic of participants (n=47)

Demographics	Value (Mean ± SD)	Minimum - Maximum
Age (year)	66.0 ± 10.1	39.0 – 86.0
Sex (n)		
Male/Female	28/19	
BMI (kg/m2)	23.7 ± 3.6	15.1 – 33.7
NYHA classification		
NYHA II (n,%)	36 (77%)	
NYHA III (n,%)	11 (23%)	
LVEF group (%)		
HFrEF	11 (23%)	
HFmrEF	9 (19%)	
HFpEF	27 (58%)	
6 MWT (m)	403.5 ± 96.0	181 – 612
6MWT% predicted	83 ± 18	48 – 141
1 STST1 (repetition)	29.1 ± 7.9	12 – 44
1STST % predicted	100 ± 26	38 – 151
Comorbidities (n,%)	*43 (94%)*	
Hypertension	24 (51%)	
Diabetes	17 (36%)	
Coronary artery disease	17 (36%)	
Lipid disorder	11 (23%)	

n: number of subjects; BMI: body mass index; NYHA: New York Heart Association; LVEF: left ventricular ejection fraction; HFrEF: Heart failure with reduced ejection fraction; HFmrEF: Heart failure with mildly reduced ejection fraction; HFpEF: Heart failure with preserved ejection fraction; 6MWT: 6-minute walk test, 1STST: 1-minute sit-to-stand test

Reliability

The mean number of the best 1STST repetitions was 28.9 ± 8.1 on the first visit and 29.3 ± 8.2 on the second visit without significant difference ($p > 0.05$), and the ICC between these two visits was 0.98 (95% CI: 0.96 - 0.99). Noticeably, the second administration was constantly better in both visits. The SD and SEM of the difference in repetitions between the best 1STST from the first and second visits were 2.2 and 0.34, respectively. The Bland-Altman plot showed good agreement, with a mean difference of 0.4 repetitions and limits of agreement ranging from -3.8 to 4.5 repetitions between the two visits (Figure 1). The linear regression model indicated no proportional bias ($p > 0.05$). Based on the distribution-based approach, the MID and MDC for the 1STST were estimated to be 1.1 and 0.94 repetitions, respectively.

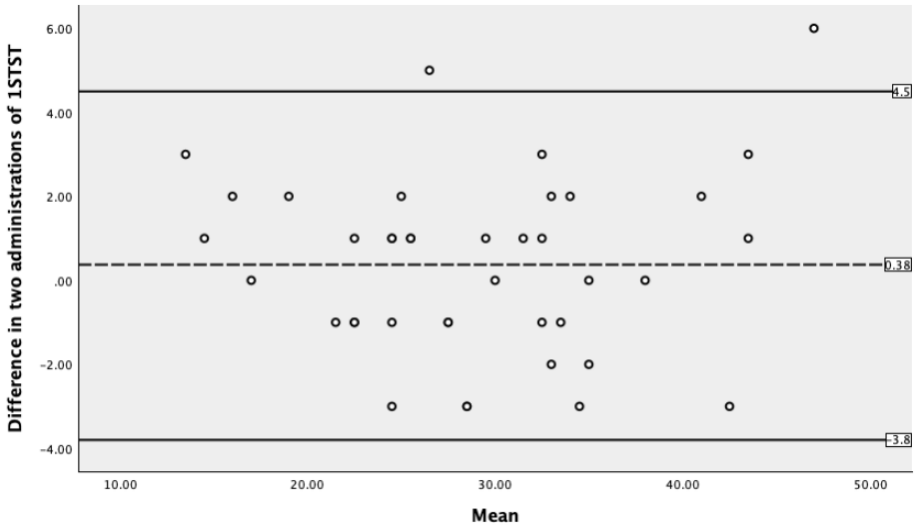


Figure 1. The Bland-Altman plot for the two best 1-minute sit-to-stand performances of each visit.

Learning effect

There was a significant increase in 1STST repetitions from the first to the second administration within each visit ($p < 0.001$). The effect size was large, with Cohen's d =1.24 for the first visit and 1.78 for the second visit. (Figure 2)

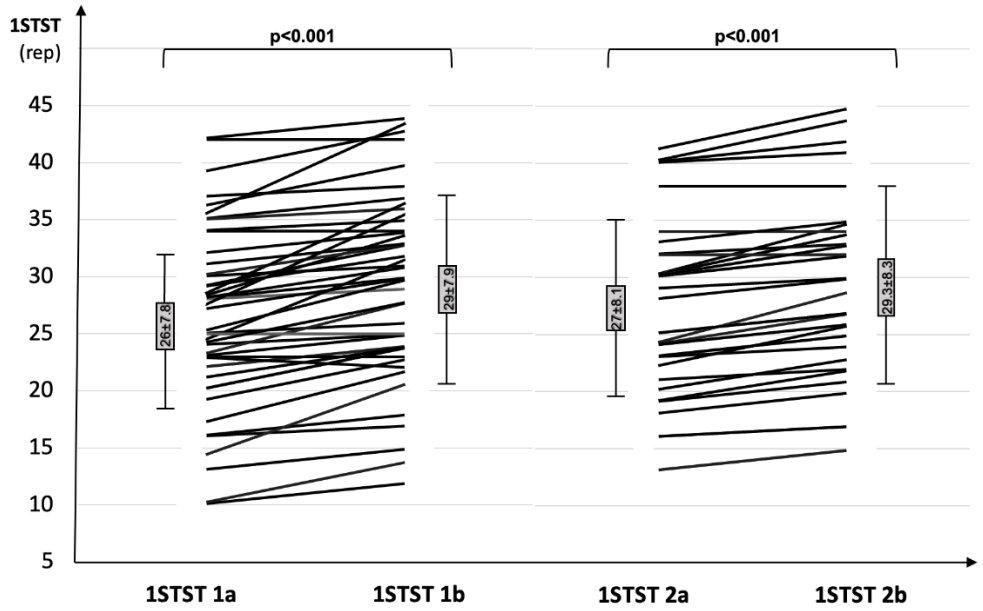


Figure 2. The difference in 1-minute sit-to-stand tests between the two performances in each visit

Validity

The results demonstrated a strong correlation between the number of 1STST repetitions and 6MWT distance (rho = 0.74, 95% CI: 0.55 – 0.86; $p < 0.001$) (Figure 3). The pre-test heart rate, systolic pressure, oxygen saturation, dyspnea, and fatigue levels were similar before both tests. Moreover, the physiological adaptations observed during the 1STST1 and 6MWT were comparable and not statistically significant ($p>0.05$), as shown in Table 2.

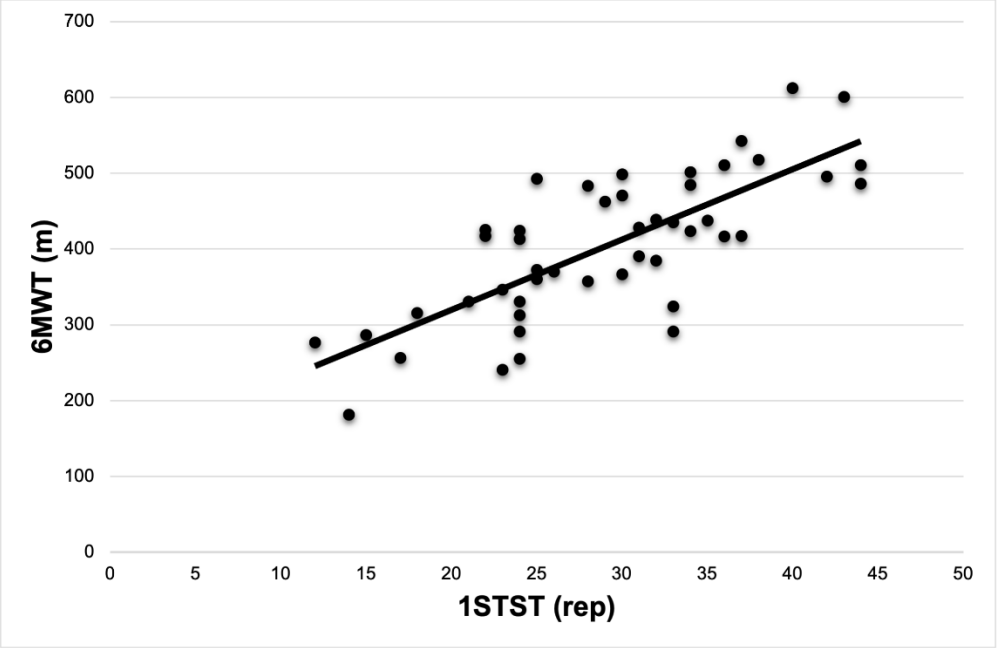


Figure 3. The correlation plot between the 6-minute walk test and the 1-minute sit-to-stand test.

Table 2. Pre and post-test physiological adaptation between 6MWT and 1STST (n=47).

	Pre-test		P	Post-test		ΔPre-post test		p
	6MWT	1STST		6 MWT	1STST	6MWT	1STST	
HR (bpm)	75.0±12.7	75.6±13.0	0.44*	88.5±15.3	90.4±13.4	13.5±7.4	14.8±8.6	0.25†
SBP (mmHg)	111.7±1.4	112.0±14.7	0.54*	122.2±15.3	124.7±15.5	10.5±9.4	13.4±7.3	0.07*
SpO_2(%)	97.6±1.8	97.7±1.5	0.88†	97.8±1.9	97.9±1.5	0.2±1.4	0.2±0.9	0.95†
Borg dyspnea	0.1±0.2	0.1±0.2	1.0†	1.57±1.6	1.6±1.3	1.5±1.5	1.6±1.3	0.42†
Borg fatigue	0.1±0.3	0.1±0.2	0.57†	1.8±1.3	1.9±1.1	1.7±1.3	1.9±1.1	0.10†

6MWT: 6-minute walk test; 1STST: 1-minute sit-to-stand test; Δ: change from before to after test;
HR: Heart rate; bpm: beat per minute; SBP: systolic blood pressure; SpO_2: Saturation of peripheral oxygen;
*: p value of paired T-test;
†: p value of the Wilcoxon signed-rank test.

Discussion

This study examines the reliability and the learning effect of the 1STST in the time in patients with HF. To our knowledge, this is the first study on test-retest reliability of the 1STST between two separate visits in CHF with an ICC of 0.98 [152]. This is slightly better than previously reported by another study (ICC = 0.93) [251]. However, in the mentioned study, the reliability of the 1STST was assessed on the same day, with two test administrations performed 30 minutes apart [251]. This relatively short interval may not be sufficient to fully evaluate the reliability from a clinical and methodological perspective [2,260]. The method of the Bland-Altman plot highlighted a small systematic error of 0.38 repetition between the two visits [2]. The random error ranged from -3.8 to 4.5 repetitions. The small systematic and random error may have been attributed to conducting two performances per visit, which allowed participants to anchor their performance and reduce the learning-effect variability associated with the test. Moreover, we observed no proportional bias, suggesting that the reliability of 1STST was not affected by the number of repetitions [261]. Those results are consistent with the evidence from other populations [150].

The distribution-based determination of the MID [2] estimated a difference of 1.1 repetitions, which, from a clinical perspective, could be rounded to 2 repetitions. The result suggests that even small changes in performance on this measure may be clinically meaningful in patients with HF. Additionally, the MID exceeds the observed MDC, suggesting that the MID can reliably detect the true change. The MID observed in this study is relatively similar to that reported for the Chronic lung disease population [148,163]. However, the MDC of our study was lower than a previously suggested MDC of 3.79 repetitions in the same patient group [251]. The smaller MDC observed in our study was due to the higher ICC, which mathematically reduced the SEM, thus the MDC. Although the distribution-based method used to estimate the MID is commonly employed, it lacks the strong theoretical basis of anchor-based

methods, which directly assess the patient's perspective for a meaningful change [262]. Therefore, the application of MID in this study in clinical practice should be cautious.

This study reported the learning effect associated with the 1STST of 2 to 3 repetitions between the first and second administrations, realised 30 minutes apart on patients having no experience of 1STST. This result contradicts a previous study [251]. In addition, Cohen's d exceeded 0.8 in both visits, indicating a large learning effect size. The discrepancies between our results and the previous study [251] may be attributed to the subjects' prior familiarity with the test before participating in the referenced research. Indeed, a learning effect has also been reported in other populations in previous studies using the 1STST [263] and other functional tests like the 6MWT [264]. While the progression of this learning effect over time remains unknown, our findings suggest that administering the 1STST twice is necessary to capture the patient's true functional exercise capacity.

Furthermore, our findings confirmed that the 1STST is a valid measure for assessing functional exercise capacity in this population. The strong correlation between the 1STST and the 6MWT, along with the 95% CI of the correlation being larger than 0.5, suggests that the 1STST is an alternative method for assessing functional exercise capacity in this population, aligning with supportive evidence for the 1STST in other populations [150]. Additionally, fatigue and dyspnoea level scores were also well correlated to 1STST (Table 2). These findings are consistent with a previous study [251], which also reported similar physiological responses between the 1STST and 6MWT. Nevertheless, it has been suggested that the shorter duration of the 1STST, which is less than the recommended minimum of 3 minutes needed to stabilize heart rate [265], may limit the ability to reflect chronotropic competence [266]. However, as Tanriverdi et al.[10], we observe similar HR responses between both tests. This rapid increase in HR could be related to higher exercise intensity during the 1STST, which relates indeed more directly to lower body strength

and muscle function [267]. This could moreover partially explain the lower mean sit-to-stand performance for patient performing similar 6MWD, as their patients presented with higher BMI than ours [10]. The 1STST is then probably complementary to the 6MWT, which also involves integrated assessment of cardiovascular, respiratory, and musculoskeletal systems. Therefore, the 1STST may provide a more targeted evaluation of functional exercise capacity related to lower extremity strength, which are other important factors predicting survival rate and incident risk in patients with HF [268,269]. Thus, the 1STST can be a useful tool in rehabilitation programs for patients with HF, allowing for a more focused assessment and targeting of lower extremity function, which is crucial for improving functional exercise capacity and overall outcomes in this population [270].

Study Limitations

This study has several limitations. First, this study only included patients with NYHA class II and III who exhibited high predicted performance on 6MWD and 1STST. This may restrict the generalizability of the findings to patients with more advanced HF, who could display different feasibility and reliability characteristics for the 1STST. Another limitation is the lack of a comprehensive assessment of other important functional outcomes, such as health-related quality of life, activities of daily living, and muscle strength, which could have provided a more complete evaluation of the validity of the 1STST [137,251].

Conclusion

The 1STST is valid and reliable for the assessment of functional exercise capacity in patients with HF. The learning effect of 1STST requires performing at least two trials to capture the true functional capacity in patients with HF during each visit.

Chapter 3: Translation, Validation, and Reliability Investigation of Questionnaires in Vietnamese.

In clinical practice, many outcomes are obtained via questionnaires due to the indirect nature of assessment for constructs such as quality of life, child development, or aspects of psychology. To ensure the accuracy and cultural relevance of research and clinical assessments in the new language, the cross-cultural validity and reliability of questionnaires must be rigorously evaluated. This process typically involves translation, adaptation, and validation to ensure that the questionnaire accurately captures the intended constructs within the specific cultural context. In Vietnam, the accessibility of valid and reliable questionnaires for physiotherapists is limited. Therefore, it is essential to increase the availability of questionnaire assessments in Vietnamese, which may improve the quality of rehabilitation services in Vietnam. Our project chose two questionnaires to investigate: the Paediatric Evaluation of Disability Inventory and the Nijmegen Questionnaire for hyperventilation.

The Cross-Cultural Validity and Reliability of the Vietnamese Version of the Paediatric Evaluation of Disability Inventory

Ngoc-Minh Nguyen; Thi-Quynh-Nhu Do; Quan Minh Hien Nguyen; Hanh Thi Bich Tran; Thy Hoang Phuong Dao; Gregory Reychler; Cynthia Dodds, Patty Coker-Bolt

Adapted from: Child: care, health and development, March 2025

DOI: 10.1111/cch.70066

Abstract

Background: To translate and investigate the validity and test-retest reliability of the Vietnamese Pediatric Evaluation of Disability Inventory (vPEDI).

Methods: The PEDI was translated and adapted following established guidelines, including forward translation, reconciliation, and back translation. Content validity was assessed by expert panels and the Item-Content Validity Index (I-CVI), Universal Agreement among experts Scale-Content Validity Index (S-CVI/UA), and the Average CVI (S-CVI/Ave) were calculated. The Kappa statistics tested the level of agreement among content experts. The face validity was assessed by determining the percentage of each level of the rating of easiness of understanding as rated by 32 caregivers. A total of 446 Vietnamese children aged 6 to 90 months were recruited to assess the normal raw scores by administering the Vietnamese PEDI. From this total, 50 children were evaluated twice within two weeks to examine the test-retest reliability of the vPEDI using intraclass correlation coefficients (ICC) and Bland-Altman analysis.

Results: The vPEDI required minor modifications to be responsive to the culture and typical daily activities in Vietnam. The I-CVI for all domains was above 0.8. The S-CVI/UA, and S-CVI/Ave for clarity and relevance were from 0.78 to 0.98. Face validity ratings indicated high understandability. The test-retest reliability of all domains was excellent, with ICCs above 0.93.

Conclusion: The vPEDI is a valid and reliable tool for assessing functional abilities in Vietnamese children. Healthcare providers can use the vPEDI to set individual goals and guide intervention strategies for contexts and environments relevant to Vietnam.

Keywords: Cross-culture, Validity, reliability, Pediatric Evaluation of Disability Inventory, norm values distribution.

Introduction

Childhood disability and developmental delay demonstrate an increasing trend globally, affecting nearly 240 million children [271]. In middle-income countries like Vietnam, as the neonatal survival rate increases, the rate of childhood disability has also increased [272,273]. Vietnamese neonatal disability rates have increased, reaching 37/1000 live births [274] and a significant number of children are living with various disabilities that impose a heavy burden on families, society, and the healthcare system [275,276]. There is a need to determine the most effective and targeted interventions, which will require comprehensive, practical, and culturally relevant assessment tools, to support children with disabilities (CWD) and their families [277]. Relevant assessment measures can provide valuable insights into the current functional abilities of children, informing and guiding intervention strategies to help children and their families achieve individualized goals within the context of their homes and communities [278-280]. Unfortunately, there are limited standardized assessment tools that fully assess the capacities of Vietnamese children, leading to inconsistencies in the identification of children with developmental delays and potentially impacting the delivery of critical rehabilitation services. Without reliable and culturally relevant assessment tools, interventions and support for children with atypical development and their families may not be adequately tailored to their specific needs or provided in a timeframe to maximize a child's overall potential.

Among the popular assessment tools, the Pediatric Evaluation of Disability Inventory (PEDI) is a widely used pediatric assessment due

to the alignment with the framework of the International Classification of Functioning, Disability, and Health [281] and its comprehensive examination of a child's activity performance, participation in daily tasks, and the environmental factors that influence performance [282,283]. Additionally, it offers normative scores for children ages 6 months to 7.5 years, as well as criterion-referenced or scaled scores that can be used for children up to 18 years of age, enabling its use across a wide developmental range [283,284]. The PEDI was developed and standardized for children living in the United States (US), whose living conditions, customs, and social behaviours differ from those living in non-US cultures. However, the PEDI has shown good reliability, validity, and responsiveness in cross-cultural variations, which have successfully adapted items to fit the unique experiences and contexts of children in many different countries [285-289]. This study aimed to translate and culturally adapt the PEDI for use with Vietnamese children, and to investigate the reliability and validity of the new Vietnamese version. A secondary objective was to explore the distribution of scores across normative age groups of typically developing Vietnamese children aged 6 months to 7.5 years.

Method

Ethics statement

The study was approved by the Ethics Committee at the University of Medicine and Pharmacy at Ho Chi Minh City (402/HĐĐĐ-ĐHYD). The participants of the study were parents of children without disabilities. The informed consent was explained to the participants before participation. The study follows the Declaration of Helsinki and Clinical Good Practice [290].

Participants

Using social networks and health stations, caregivers of typically developing children ages 6 months to 7.5 years old without disability, or delayed development, as determined by Vietnamese medical professionals, were recruited from three regions of Vietnam

(Northern, Central, and Southern). Researchers contacted the participants and scheduled appointments after obtaining their contact information. All assessments were conducted in the children's homes. The sample size was determined following the COSMIN guidance, requiring at least 30 caregivers for the face validity assessment and 50 children for the test-retest reliability investigation [2]. Additionally, the researchers conducted a prior power analysis, which determined that at least 23 children per age subgroup were needed to achieve the desired error rates. This was based on the expected large effect size and a power of 80% for the normative score [291]. The sample was divided into 14 age groups, each spanning 6 months, ranging from 6 months to 7.5 years of age, consistent with the approach used in the original study [283]. The recruitment of participants continued until each age group reached the minimum number of 23 children in each age subgroup.

Measurement: Pediatric Evaluation of Disability Inventory

The PEDI is a comprehensive assessment tool that examines a child's functional capabilities, the level of independence, and the extent of modifications required to perform functional activities in three domains including self-care, mobility, and social function. The Functional Skills Scale (FSS) assesses the child's performance across these domains with 197 items related to self-care (73 items), mobility (59 items), and social function (65 items), with dichotomous scoring (unable/capable to perform the task). The Caregiver Assistance Scale (CAS) measures the level of caregiver support required for the child across the same three domains with 20 items, using a Likert scale (0: total assistance required; 5: totally independent). For the FSS and CAS, higher scores indicate a higher level of performance in the respective domains. The Modifications Scale (MS) focuses on the environmental adaptations needed to support the child's function, with four discrete categories reported as frequency counts (N: no modifications; C: child-oriented modifications; R: specialized rehabilitation equipment; E: extensive modifications). The separate scales can provide insight into the

child's limitations and strengths across the three domains. Together, these subscales provide a holistic understanding of the child's activity performance, participation in daily tasks, and the environmental factors that influence performance [283]. The PEDI is scored through either direct observation by a healthcare provider or caregiver reports. Although the original PEDI design does not include the calculation of overall summary scores for FS and CA, we chose to calculate these scores to provide additional information about children's overall performance.

Protocol for Translation and Cross-Cultural Validity

Permission to translate the PEDI into Vietnamese was obtained from the primary author of the original U.S. version of the PEDI prior to the start of the study. The PEDI translation and back translation were performed following the guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures [292]. The PEDI was translated into Vietnamese by two translators skilled in both English and Vietnamese. The first draft (V1) was synthesized after the review process with two translators and one rehabilitation expert. Two new and distinct translators conducted the back translation of V1. A meeting was organized between the two translators who completed the back translation and a rehabilitation expert to finalize the translated PEDI (E1). The rehabilitation expert conferred with the two translators to reach a consensus on the translation in case of disagreement. The V1 and E1 versions were sent to a panel of 5 Vietnamese pediatric rehabilitation experts to assess the quality, clarity, and relevance of the V1 version. The comments from this expert panel were collected to determine any necessary modifications for V1. Then, a separate and new panel of 5 Vietnamese pediatric rehabilitation experts assessed the content validity of the modified V1 using a 4-point Likert scale (from 1: cannot be used, not relevant/clear to 4: highly relevant/clear) [293,294]. All experts were physicians and physiotherapists with at least five years of experience. The last step of the translation involved correspondence with a primary author of the original U.S. version of

the PEDI to confirm and accept any modifications to the final version of the Vietnamese PEDI (vPEDI).

To evaluate the face validity of the translated PEDI, a convenience sample of at least 30 Vietnamese caregivers of typically developing children aged 6 months to 7.5 years was recruited through a social network, and the questionnaire was then sent via email [292]. The participants were asked to rate how easy each PEDI item was to understand, using a Likert scale from 1 to 5 (with 5 indicating "very easy to understand"). This process was crucial in demonstrating that the translated and adapted vPEDI was culturally appropriate, understandable, and relevant to the intended population [2] and parental feedback helped identify any issues or concerns with the adapted vPEDI, enhancing its suitability for Vietnamese children.

Reliability

The test-retest reliability of the vPEDI was explored by administering the PEDI assessment to 50 participants on two separate occasions, with a two-week interval between the assessments. This approach enabled an evaluation of the stability and consistency of the vPEDI score over a brief period of time, ensuring the reliable administration of the measure [2]. This step involved a convenience sample of families who volunteered to participate in this part of the study.

The distribution of scores across normative age groups investigation

To investigate the distribution of scores across normative age groups, in addition to the 50 participants assessed for test-retest reliability, a larger sample of children was recruited until at least 23 children were evaluated in each age group. The score was described by mean and standard deviation.

PEDI assessment procedure

Two U.S.-trained pediatric therapists (one physical therapist, and one occupational therapist) provided in-country PEDI training to all Vietnamese assessment therapists. The vPEDI was administered by therapists who completed this training and included three

experienced Vietnamese physiotherapists and two Vietnamese students in their final year of study in the field of rehabilitation. To ensure reliable administration of the vPEDI, the Vietnamese assessment therapists had to reach a predetermined level of acceptable reliability with the U.S. trainers before conducting the field assessments. Specifically, the Vietnamese assessors were required to demonstrate 90% agreement with the U.S. trainers in individual item scoring on three practice assessments before being permitted to carry out the field tests independently. The assessment procedure included structured interviews with each child's caregivers to gather information and direct observation of the child in their home environments following the manual instruction of the original PEDI by one assessor.

Data analysis

To investigate the content validity, we examined the proportion of agreement on the relevancy of the items in each subscale, known as the item content index (I-CVI), and the proportion of items for a subscale that achieved a rating of 3 or 4, denoted as the scale-level content index (S-CVI) [293]. For the item level content validity, I-CVI was calculated by dividing the number of experts rating the relevancy of each item as level 3 or 4 by the total number of experts. Assessments of each item using the I-CVI were made given the guidelines that I-CVI > 0.79 indicates the item is appropriate, 0.7-0.79 suggests the item needs revision, and < 0.7 implies the item should be eliminated For the scale-level content validity, S-CVI was rated by the Universal Agreement among experts (S-CVI/UA), calculated by dividing the total number of items rated as valid content by all experts by the total number of items, and the Average CVI (S-CVI/Ave), calculated by dividing the sum of the I-CVIs for each item by the total number of items. Assessments of the two S-CVI measures were made using the guidelines that S-CVI/UA ≥ 0.8 and S-CVI/Ave ≥ 0.9 indicate agreement among experts on the validity of the content. To adjust for the chance agreement among experts, the level of agreement was tested by the Kappa statistic, with the values

> 0.74 being excellent, 0.6-0.74 being good, and 0.4-0.59 being fair. The face validity was assessed by examining the percentage of participants who selected each level on a scale of ease of understanding [2].

We examined the test-retest reliability using the absolute agreement intraclass correlation coefficient (ICC) [295]. The ICC ranges from 0 to 1, where ICC between 0.5-0.75 indicates moderate reliability, > 0.75 indicates good consistency, and > 0.9 suggests excellent reliability [152]. The bias and limits of agreement were assessed with the Bland-Altman plot [296]. The agreement of the modification scale was investigated with Cohen's Kappa or weighted statistic [297]. Descriptive statistics were calculated on the raw score of each age group for each scale. All analyses were conducted using the Statistical Package for the Social Sciences (SPSS-IBM) software.

Result

Participants

Thirty-two caregivers took part in the face validity test. The vPEDI was administered to a total of 446 typically developing Vietnamese children (198 girls and 248 boys) with an age range from 6 months (0.5 years) to 90 months (7.5 years) and a median age of 43 months. From this total, 50 children were evaluated twice within two weeks to examine the test-retest reliability. The majority of the families and children resided in central and southern Vietnam.

The Vietnamese PEDI translation and cross-cultural validity

The Vietnamese PEDI translation went through a rigorous process to ensure cultural appropriateness and relevance for the Vietnamese population. The vPEDI was modified based on feedback from a panel of five Vietnamese pediatric rehabilitation experts, who assessed the quality, clarity, and relevance of the initial Vietnamese version. Their comments were used to make necessary revisions. The final step involved collaboration with one of the primary U.S. authors of the original PEDI to confirm and accept any modifications to the resulting

final Vietnamese PEDI version. The modifications made to the vPEDI, as shown in Table 1, incorporated considerations of common modes of transportation and culturally popular daily activities in Vietnam to ensure the translated instrument was culturally relevant and appropriate for the target population in all regions of the country.

Table 1. The modifications of the Vietnamese PEDI

Domain	Item	The PEDI	The Vietnamese PEDI
Self-care	A. Food texture	2. Eats ground/lumpy food	2. Thức ăn xay – **Minced food**
	B. Use of utensils	9. Use a knife to butter bread, cut soft food	9. Sử dụng đũa - **Use chopstick**
	I.Pullover/Front-opening Garments	42. Puts on and remove front-opening shirts, not including fasteners	42. Mặc và cởi áo mở ở phía trước, không có **khuy**, khóa - Puts on and remove front-opening shrits, not include **BUTTONS**/ fasteners
		43. Put on and remove front-opening shirts, including fasteners	43. Mặc và cởi áo mở ở phía trước, có **khuy**, khóa - Puts on and remove front-opening shrits, include **BUTTONS**/fasteners
Mobility	C. Car transfers	C. Car transfers	C. Di chuyển vào và ra xe hơi/taxi/xe máy - C. Car/**Taxi/Motorbike** transfers
	E. Tub transfers	E. Tub transfers	E. Chuyển thế trong bồn tắm/vòi sen/chậu tắm - E. Tub/**Shower**/ **tub bath** transfers

Content validity

Results provide strong evidence of content validity. The clarity level was excellent with S-CVI/UA > 0.8 and S-CVI/Ave > 0.9. [293] The Relevance level was less robust with S-CVI/UA < 0.8, but S-CVI/Ave was still > 0.9. [293] The I-CVI for each domain is shown in Table 2.

Table 2. Content validity *estimates* of Vietnamese PEDI

Item	Clarity				Relevance			
	A	I-CVI	Pc	K	A	I-CVI	Pc	K
FSS- SC	5/5	1.00	0.03	1.00	5/5	1.00	0.03	1.00
FSS-M	5/5	1.00	0.03	1.00	5/5	1.00	0.03	1.00
FSS-SF	5/5	1.00	0.03	1.00	5/5	1.00	0.03	1.00
CAS-SC	5/5	1.00	0.03	1.00	5/5	1.00	0.03	1.00
CAS-M	5/5	1.00	0.03	1.00	5/5	1.00	0.03	1.00
CAS-SF	5/5	1.00	0.03	1.00	4/5	0.80	0.16	0.76
MS-SC	5/5	1.00	0.03	1.00	5/5	1.00	0.03	1.00
MS-M	5/5	1.00	0.03	1.00	5/5	1.00	0.03	1.00
MS-SF	4/5	0.80	0.16	0.76	4/5	0.80	0.16	0.76
	S-CVI/UA = 0.89				S-CVI/UA = 0.78			
	S-CVI/Ave = 0.98				S-CVI/Ave = 0.96			

A: The number of experts rated the item as a 3 or 4 on a 4-point scale; I-CVI: Item-content validity index; S-CVI: Scale-level-content validity index; S-CVI/UA: universal agreement among experts; S-CVI/Ave is averages of the item-level CVIs; Pc: Probability of chance agreement; K: Modified kappa statistic coefficient; FSS-SC: Functional Skill scale-self-care; FSS-M: Functional Skill scale-Mobility; FSS-SF: Functional Skill scale-social function; CAS-SC: Caregiver Assistance scale-self-care; CAS-M: Caregiver Assistance scale-Mobility; CAS-SF: Caregiver Assistance scale-social function.

The Face validity

Thirty-two Vietnamese caregivers of typically developing children evaluated the face validity. No item was rated at levels 1, 2, or 3. In summary, across all items, eight caregivers rated the vPEDI items at

level 4 (25% - easy to understand), and 28 rated items at level 5 (75% - very easy to understand).

The test-retest reliability

Scores from all scales had high levels of reliability with ICC ranging from 0.932 to 0.994. The ICCs of all domains of FSS and CAS were excellent with all levels above 0.9 (Table 3). The Bland-Altman method between the two assessments was reported in Table 4. The effect size was investigated with the result of FSS-SC (d=0.45), FSS-M (d=0.19), FSS-SF (d=0.36), total FSS (d=064), CAS-SC (d=0.54), CAS-M (d=37), CAS-SF (d=0.33), and total CAS (d=0.88). For the categorical MS, The Kappa (k) were good to excellent for all domains (Table 5).

Table 3. Test-retest reliability for FSS and CAS

PEDI Domains	**ICC**	**95% CI**
Functional skill scale		
Self-care	0.982*	0.961-0,991
Mobility	0.994*	0.990-0.997
Social function	0.983 *	0.965-0.991
Caregiver assistance scale		
Self-care	0.945*	0.877-0.973
Mobility	0.952*	0.916-0.972
Social function	0.932*	0.884-0.961

FSS: Functional Skill scale; CAS: Caregiver Assistance scale; ICC: intraclass correlation coefficient, CI: confident interval
**: P<0.0001*

Table 4. Bias and Limit of agreement for test-retest reliability for FSS and CAS

Limit of agreement	Bias	Upper	Lower
Functional skill scale			
Self-care	-1.56	4.777	-7.897
Mobility	0.06	2.698	-2.578
Social function	-1.04	3.92	-6
Total	-2.54	6.448	-11.528
Caregiver assistance scale			
Self-care	-1.98	5.559	-9.519
Mobility	0	5.178	-5.178
Social function	-0.28	4.685	-4.965
Total	-2.26	9.962	-14.483

FSS: Functional Skill scale; CAS: Caregiver Assistance scale

Table 5. Test-retest reliability for the Modifications scale (Cohen's Kappa)

Domains of Modifications scale	Items								
	1	2	3	4	5	6	7	8	Mean
Self-care	0.86	0.91	1	1	1	0.93	0.87	1	0.94
Mobility	0.89	1	1	0.79	1	1	1	_	0.95
Social function	1	1	1	1	0.83	_	_	_	0.96

Normal value distribution throughout the age groups

The mean standard score of the PEDI for each 6-month age group is presented in Table 6.

Table 6. Distribution of raw score totals across age groups

Age groups (years)	n = 446	Gender (F/M)	FSS-SC (SD)	FSS-M (SD)	FSS-SF (SD)	FSS-total (SD)	CAS-SC (SD)	CAS-M (SD)	CAS-SF (SD)	CAS-total (SD)
0.5 – 0.9	51	27/24	6.6 (3.7)	8.9 (6.7)	5.7 (4.0)	21.2 (13.0)	1.0 (2.2)	2.6 (3.7)	0.8 (1.5)	4.4 (6.8)
1.0 – 1.4	38	12/26	19.5 (7)	36.8 (11.4)	16.7 (6.6)	73 (19.7)	5.5 (5.7)	16.3 (6.4)	5.7 (4.1)	27.5 (13.5)
1.5 – 1.9	28	13/15	34.0 (12.0)	44.6 (6.0)	28.9 (7.9)	107.6 (21.3)	9.9 (6.2)	22.2 (6.6)	10.5 (5.0)	42.6 (14.9)
2.0 – 2.4	32	12/20	39.9 (5.3)	48.3 (8.7)	36.6 (5.3)	124.7 (13.3)	16.2 (9.4)	26.5 (4.7)	14.5 (4.7)	57.2 (15.4)
2.5 – 2.9	38	22/16	48.1 (10.3)	49.6 (5.4)	41.9 (7.4)	139.6 (18.5)	21.6 (10.6)	28.7 (4.7)	16.8 (4.2)	67.1 (16.6)
3.0 – 3.4	28	15/13	52.9 (6.9)	51.3 (4.0)	46.9 (5.7)	151.1 (9.8)	27.7 (9.2)	29.4 (3.9)	18.4 (4.0)	75.4 (13.9)
3.5 – 3.9	35	16/19	60.4 (7.8)	51.8 (8.2)	52.5 (6.8)	164.7 (16.3)	29.2 (7.1)	31.3 (3.2)	20.4 (3.1)	80.8 (9.9)
4.0 – 4.4	27	12/15	63.2 (7.0)	55.4 (2.8)	54.8 (5.7)	173.3 (13.2)	30.9 (6.7)	32.0 (2.8)	21.3 (2.4)	84.1 (9.9)
4.5 – 4.9	34	14/20	63.5 (5.8)	54.2 (7.1)	57.2 (4.4)	174.9 (13.6)	31.2 (5.5)	32.9 (1.8)	22.1 (3.3)	86.2 (8.5)
5.0 – 5.4	25	10/15	66.0 (5.9)	54.4 (6.8)	56.6 (5.9)	176.9 (15.3)	32.1 (6.8)	32.7 (2.2)	20.9 (3.3)	85.8 (9.5)
5.5 – 5.9	36	14/22	67.3 (6.6)	56.0 (6.7)	58.4 (5.6)	181.7 (14.5)	35.0 (5.0)	33.8 (2.0)	22.9 (3.0)	91.7 (8.5)

6.0 – 6.4	24	11/13	70.5 (2.4)	57.7 (2.9)	61.5 (2.2)	189.7 (4.8)	37.3 (3.7)	34.4 (0.9)	23.1 (2.5)	94.8 (5.9)
6.5 – 6.9	27	12/15	68.0 (13.1)	56.9 (4.6)	59.4 (10.8)	184.3 (27.9)	36.4 (8.9)	33.3 (5.4)	22.5 (5.4)	92.3 (19.4)
7.0 +	23	8/15	72.1 (1.2)	56.7 (7.8)	61.3 (1.8)	192.0 (8.6)	38.7 (2.8)	34.7 (0.6)	24 (1.3)	97.4 (3.4)

The standard scores are presented as Mean (standard deviation)

FSS-SC: Functional Skill scale-self-care; FSS-M: Functional Skill scale-Mobility; FSS-SF: Functional Skill scale-social function; FSS-total: Functional Skill scale-total score; CAS-SC: Caregiver Assistance scale-self-care; CAS-M: Caregiver Assistance scale-Mobility; CAS-SF: Caregiver Assistance scale-social function, CAS-total: Caregiver Assistance scale-total score; F/M: female/male

Discussion

The Vietnamese version of the PEDI was translated and cross-culturally validated through a rigorous translation, reconciliation, and back-translation process [292]. The vPEDI has been successfully adapted to the unique culture and environment of Vietnam through minor modifications. The newly adapted vPEDI demonstrates excellent content validity, as evidenced by the high I-CVI, S-CVI/UA, and S-CVI/Ave scores [293]. This result is not surprising due to the strong content validity of the original PEDI [283]. The vPEDI's incorporation of culturally relevant activities and modes of transportation provided a comprehensive evaluation of a child's functional abilities within their everyday routines and environments. This will enable early intervention programs and providers to deliver services that are culturally responsive to the unique needs and experiences of Vietnamese children and their families.

The inclusion of Vietnamese families in determining face validity also enhances the usability of the vPEDI. Since the PEDI uses a structured parent interview format and clinical observations of the child, it is crucial that the caregivers can comprehend the items of the vPEDI [298]. Our study found the vPEDI to have excellent face validity, with all caregivers rating the Vietnamese version as understandable or easy to understand. The authors assured the vPEDI only included activities that occur routinely in the Vietnamese culture and avoided the use of complex medical terminologies. The integration of culturally relevant activities and modes of transportation further contributes to the vPEDI's strong face validity and will support the ability of rehabilitation specialists to accurately relate the results of the assessment to essential life activities in Vietnam. The strong face validity of the vPEDI has important implications for early intervention programs in Vietnam. By ensuring the assessment items are meaningful and understandable to Vietnamese caregivers, the vPEDI can facilitate effective communication and trust between healthcare providers and families, engaging a higher level of parent participation during the assessment process. This, in turn, can lead to more

engagement of families in early therapy services, support family-centered care, and lead to more personalized goal-setting and treatment plans that align with the child's and family's individual priority needs [299].

The test-retest reliability of the vPEDI was excellent, consistent with findings from other studies that examined cross-cultural translations of the PEDI [289,300]. For FSS and CAS domains, the ICC of all domains was higher than 0.9. The Bland-Altman analysis indicated a low risk of systematic bias between the two administrations, with all interval confidences of FSS and CAS containing zero [2,301], suggesting clinicians can have confidence in vPEDI scores. Notably, the mobility domain of CAS and FSSS demonstrated a low bias, which was zero or extremely close to zero. This implies the vPEDI can be very sensitive to distinguish changes in ambulatory skills in children [302], which is similar to the result of a prior study [303]. These findings suggest the vPEDI can be a reliable tool for assessing functional changes in Vietnamese children, which could elevate current service delivery models by allowing clinicians to track progress over time and make informed and more accurate decisions about intervention strategies. The excellent test-retest reliability across the domains indicates that the vPEDI can be used consistently to evaluate a child's functional abilities over time, enabling healthcare providers to monitor their progress and adapt treatment plans accordingly. While our study also reported the effect size of the Bland-Altman analysis, we could not find other comparable studies to compare these results.

The Kappa values indicated good to excellent agreement of the MS across the two-time assessments [297]. However, since the vPEDI was administered to typically developing children, the majority of scores for the MS domain were reported as "No Modification" or "Child-oriented," with no families reporting the need for "Rehabilitation Equipment" or "Extension Modifications". Therefore, caution is warranted when using the MS to assess children with disabilities, as

reliability is a characteristic of an instrument used within a specific population [2].

Our study also reports the vPEDI raw scores for more than 400 typical developing Vietnamese children, with caregivers as the main respondents. This data provides estimated raw scores means and standard deviations for each age group. Compared to the norm raw scores from the original study [283], we found that Vietnamese children under 1 year old had lower raw scores in FSS and CAS, while the scores were comparable across the other age groups. This difference may be attributed to the variations in caregiving practices (i.e. feeding and eating, sleep positions, bathing, grooming, and play) commonly observed in Vietnamese culture during the first year of a child's life, which could impact specific exposure to a broader range of activities identified in the PEDI scales. Prior cross-cultural studies have found that caregiving practices influence children's development with respect to developmental dynamics, the timing of developmental milestones, developmental gestalts, and precursors and consequences of developmental achievements [304,305].

Although a minor difference exists between scores of Vietnamese children under 1 year of age versus Western counterparts, the age-related trends observed in the vPEDI raw scores show a gradual and consistent increase across the different age groups over time. This pattern aligns with expectations, as one would anticipate children's functional development to progress steadily with age. The increasing raw scores across the age groups suggest the vPEDI effectively captures this developmental trajectory within a population of typically developing Vietnamese children. This provides preliminary evidence for the measure's ability to detect typical functional milestones in this cultural context.

The study had several limitations. The majority of families and children who participated in the study resided in central and southern Vietnam, with only a few families who participated from the north. This regional difference in lifestyle poses a potential risk to the

generalizability of the normative raw scores, although this issue could not be examined in the current study. Future studies with Rasch analysis are needed to determine the overall stability of the instrument and to develop best-fitting hierarchical models of functional development. The next step is to study the vPEDI's internal validity and item response pattern and to fully evaluate the vPEDI's psychometric properties to strengthen the interpretation of the caregiver-reported data [306]. The authors are currently examining the Rasch item difficulty estimates and will consider similarities or differences between U.S. and Vietnamese cohorts as part of a subsequent paper.

Conclusion

The vPEDI is a valid and reliable tool to assess the functional abilities of Vietnamese children in the domains of self-care, mobility, and social function. The newly validated vPEDI tool could be used in both clinical practice and research to evaluate children with various developmental delays and disabilities throughout Vietnam, and healthcare providers can use the vPEDI to set individual goals and guide intervention strategies for contexts and environments relevant to Vietnam. It will be important to train current and future Vietnamese practitioners about the vPEDI to enhance the relevance and applicability of this study's findings. Future studies should continue to explore the psychometric properties of the vPEDI and consider studies of the vPEDI to evaluate children with various developmental delays and conditions.

The Cross-cultural Validity and Reliability of the Vietnamese Nijmegen Questionnaire

Ngoc-Minh Nguyen; Hanh Thi-Bich Tran; Thi-Quynh-Nhu Do; Nicolas Audag; Giuseppe Liistro, Philippe Fait; Gregory Reychler

Adapted from: Cureus, 2025 Feb 7;17(2):e78709

DOI: 10.7759/cureus.78709

Abstract

Introduction: The Nijmegen Questionnaire assesses hyperventilation syndrome (HVS), but a validated Vietnamese version is lacking. This study investigates the cross-cultural validity, structural validity, and reliability of the Vietnamese Nijmegen questionnaire (VNQ) for screening HVS in primary healthcare settings.

Methods: Following Beaton's guidelines, translation and adaptation involved two independent expert panels for content validity. Exploratory factor analysis (EFA)and confirmatory factor analysis (CFA) examined the VNQ's structure. Reliability was assessed via Cronbach's alpha, intraclass correlation coefficients, Bland-Altman analysis, and linear regression.

Results: Content validity was excellent (all I-CVIs > 0.79 and all S-CVIs > 0.8). EFA revealed a four-factor structure, including the original three NQ factors and a new "psychology" factor, confirmed by CFA. Cronbach's alpha and ICC values exceeded 0.7 and 0.8, respectively, indicating good internal consistency and test-retest reliability. Bland-Altman analysis showed low systematic error, and linear regression revealed no proportional bias ($P > 0.05$).

Conclusion: The VNQ demonstrates excellent content validity, acceptable structural validity, and reliable psychometric properties for HVS screening in primary healthcare settings.

Keywords: cross-cultural validity, hyperventilation, Nijmegen, reliability, Vietnamese

Introduction

Dysfunctional breathing significantly deteriorates health conditions [307], with hyperventilation syndrome (HVS) being the most prevalent type. This latter manifests through a wide spectrum of symptoms such as stress, depression, anxiety, chest pain, and light-headedness, which collectively impair the quality of life (QoL) [308]. These symptoms are often variable, nonspecific, and overlap with many other diseases, complicating the diagnosis process [309]. The physiological basis of these symptoms is believed to be a decrease in arterial carbon dioxide due to excessive alveolar ventilation exacerbated by psychological aspects such as anxiety, stress, and anxiety-related problems [310]. Despite its heavy impact on quality of life (QoL) and an estimated prevalence of around 20% [311], HVS remains an underrecognized condition. Enhanced screening is thus essential for appropriate treatment and early rehabilitation [309], especially given the increased prevalence of HVS post-COVID-19 pandemic [312].

Traditionally, HVS diagnosis involves the exclusion of other organic diseases, posing additional challenges to HVS screening [313]. Moreover, HVS can occur in people without any underlying disease[311]. Delayed diagnosis can exacerbate symptoms of HVS, leading to conditions such as tetany [314]. Currently, there is no gold standard for HVS diagnosis, but a combination of tools is recommended, including the Nijmegen questionnaire (NQ), end-tidal CO_2 ($PETCO_2$) measurement, holding breath test (HBT), hyperventilation provocation test (HVPT), and cardio-pulmonary exercise test (CPET) [315].

The NQ stands out as a practical tool due to its simplicity and proven validity [316]. Compared to other tools, the NQ has several advantages. The NQ is a simple questionnaire with only 16 items, reducing the time required for assessment without incurring any costs. Additionally, it can be self-administered by individuals with hyperventilation syndrome. Moreover, the NQ has high sensitivity (86% - 91%) and specificity (79% - 92%), depending on health

conditions and cut-off threshold [317]. Additionally, the NQ is suitable for screening HVS in healthy populations [311]. Despite its proven utility, the NQ has not yet been translated into Vietnamese.
This study aims to translate the NQ into Vietnamese and to examine its cross-cultural validity, structural validity, and reliability, enabling its use for screening hyperventilation syndrome in primary healthcare settings throughout Vietnam.

Methodology

Ethics statement

The study was approved by the Ethics Committee at the University of Medicine and Pharmacy at Ho Chi Minh City (approval No. 1187 HĐĐĐ-ĐHYD). All participants were fully informed about the study protocol and provided written informed consent. Participant information will be stored securely and accessible only with a password. All stages of the study were conducted following the Declaration of Helsinki and Clinical Good Practice guidelines.

Participants

Participants were randomly recruited in the community and were required to be literate in Vietnamese. All participants were over 18 years old and had the appropriate consciousness level to answer the questionnaire.

The Nijmegen questionnaire

The NQ rates 16 HVS-related symptoms on a 5-point scale across three factors: shortness of breath (Items 1, 2, 6, 7, 8, 11, 15), peripheral tetany (items 10, 12, 13, 14), and central tetany (items 1, 3, 4, 5, 9), with a total possible score of 64 [318].

Translation and cross-cultural validity

The translation and cross-cultural validity of the NQ were conducted following the COSMIN guidelines [2,298]. First, permission was obtained from the original authors of the NQ, which is free to use without copyright restrictions [316]. Then, the translation and back-translation

processes adhered to the guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures [292]. The NQ was initially translated into Vietnamese by two translators. A review meeting involving these translators and an expert was held to synthesize the Vietnamese version of the NQ (V-NQ). The V-NQ was then back-translated by two different translators. Experts reviewed both the original and back-translated versions to create a synthesized back-translated NQ (B-NQ). The V-NQ and B-NQ were reviewed by a committee of five experts, whose feedback was used to finalize the Vietnamese NQ (VNQ). A different panel of five experts subsequently assessed the content validity of the VNQ. The clarity and relevance of the VNQ were rated on a 4-point Likert scale (from 1: cannot be used, not relevant/clear to 4: highly relevant/clear). Feedback from the expert panel was collected and incorporated to refine the translated version before final approval [293].
The face validity of the Vietnamese NQ was conducted with 30 participants [2]. Participants rated the comprehensibility of the NQ from 1 (unable to understand) to 4 (easy to understand). Their feedback was used to make further modifications to VNQ.

Structural validity

To examine the structural validity of the VNQ, exploratory factor analysis was conducted to either confirm the original 3-factor structure (shortness of breath, peripheral tetany, and central tetany) or establish a new structure [316]. This was followed by a confirmatory factor analysis to assess the structural validity of the V-NQ using the original or newly identified factor model [2].

Reliability

The internal consistency and test-retest reliability of the final VNQ were evaluated with a sample of at least 50 participants [2]. The VNQ was distributed via email or Google Forms, and participants completed the questionnaire twice within a two-week interval. Participants also completed the 11-point Global Rating of Change (GRC) during the second administration to detect any changes in

health status. The GRC is a visual analog scale for recalling the health status at different time points. Participants with a non-zero GRC score were excluded from the reliability analysis [319].

Data analysis

Content validity was assessed using the item content validity index (I-CVI) and scale-level content validity index (S-CVI) [320]. The I-CVI was calculated by dividing the number of experts rating an item as relevant/clear (score 3 or 4) by the total number of experts [293]. An I-ICV score greater than 0.79 is considered excellent, between 0.7 to 0.79 is average, and less than 0.7 is poor [293]. The S-CVI was evaluated through two indices: Universal Agreement among experts (S-CVI/UA) and Average CVI (S-CVI/Ave) [293]. The S-CVI/UA is the ratio of items with an I-CVI of 1 to the total number of items [293]. Content validity is considered excellent if S-CVI/UA ≥ 0.8, and S-CVI/Ave ≥ 0.9 [293]. The risk chance agreement was assessed using the Kappa statistic, with values > 0.74 being excellent, 0.6-0.74 good, and 0.4-0.59 fair [293].

The face validity of the Vietnamese NQ was examined by having 30 participants rate the comprehensibility of the questionnaire.

The exploratory factor analysis was conducted using the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity to confirm the assumptions [321]. The KMO value assessed the sampling adequacy, while Bartlett's test evaluated the presence of sufficient correlations among the items to proceed with the factor analysis. The rotation of sums of squared loadings was used to determine the total variance explained by each factor. Factor loadings above 0.4 were considered acceptable.

The confirmatory factor analysis was performed to evaluate the model fit of the VNQ, using indicators such as the Chi-squared, Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMSR). These model fit indices were assessed based on

the COSMIN guidelines and general recommendations for CFA [2]. For the CFI, values of 0.95 or higher are considered to indicate good model fit, while values between 0.90 and 0.95 suggest adequate fit, and below 0.90 indicates poor fit [322]. For the RMSEA, values of 0.05 or lower suggest a good model fit, and values up to 0.08 are considered acceptable [322]. Values above 0.10 indicate poor fit. For the SRMSR, values of 0.08 or lower are considered a good model fit [322].

The internal consistency of each exploratory factor was evaluated using Cronbach's Alpha [323]. An alpha value above 0.7 indicates acceptable internal consistency [323].

Test-retest reliability was analyzed using a two-way mixed-effects absolute agreement intraclass correlation coefficient (ICC) [324]. The ICCs were calculated for the total score as well as for the confirmed factors. An ICC above 0.5 is considered acceptable, above 0.8 indicates good reliability, and above 0.9 is excellent [152]. To assess bias and limits of agreement, a Bland-Altman plot was utilized [296]. Proportional bias was further investigated using linear regression [155].

The statistical analyses were performed using SPSS software, while the confirmatory factor analysis was assessed using AMOS software (IBM Corp. Released 2023. IBM SPSS Statistics for Windows, Version 29.0.2.0, Armonk, NY: IBM Corp).

Results

Participants

A total of 82 participants (40 females and 42 males) were recruited for the study. None reported any cardiopulmonary or health issues within the preceding 6 weeks. Thirty participants were involved solely in the face validity investigation. Fifty-five participants were recruited to investigate test-retest reliability, but three participants were excluded due to changes in the GRC score, leaving 52 participants for the reliability assessment. Participants' ages ranged from 24 to 59 years.

Translation

A free translation method was suggested for the “feeling tense” item, to describe a stuffy feeling in Vietnamese. The expert panel recommended removing “deeper” from the “Faster/Deeper breathing” item to avoid possible confusion with the “unable to breathe deeply” item. The VNQ was ultimately accepted (supplemental table).

Content validity

The content validity of the VNQ was found to be excellent overall (Table 1). The clarity I-CVI for the “Feeling tense” item was poor (0.4), while all other I-CVIs for clarity and relevance were above 0.79. All items' modified Kappa statistic coefficient was above 0.74, except for the clarity of the “Feeling tense” item (0.37). All S-CVIs were above 0.8.

Face validity

Thirty participants rated the ease of understanding of the VNQ. None rated it at levels 1 and 2. Seven participants (23%) rated it at level 3, and 23 participants (77%) rated it at level 4.

Table 1. The content validity of the Vietnamese NQ

Item	Clarity				Relevance			
	A	I-CVI	Pc	K	A	I-CVI	Pc	K
Chest pain	5/5	1.00	0.03	1.00	5/5	1.00	0.03	1.00
Feeling tense	2/5	0.4	0.05	0.37	4/5	0.80	0.16	0.76
Blurred vision	5/5	1.00	0.03	1.00	5/5	1.00	0.03	1.00
Dizzy spells	5/5	1.00	0.03	1.00	5/5	1.00	0.03	1.00
Feeling confused	5/5	1.00	0.03	1.00	5/5	1.00	0.03	1.00
Faster breathing	5/5	1.00	0.03	1.00	4/5	1.00	0.03	1.00
Short of breath	5/5	1.00	0.03	1.00	5/5	1.00	0.03	1.00
Tight feelings in the chest	5/5	1.00	0.03	1.00	5/5	1.00	0.03	1.00
Bloated feeling in the stomach	4/5	0.80	0.16	0.76	4/5	0.80	0.16	0.76
Tingling fingers	5/5	1.00	0.03	1.00	5/5	1.00	0.03	1.00
Unable to breathe deeply	5/5	1.00	0.03	1.00	5/5	1.00	0.03	1.00
Stiff fingers or arms	5/5	1.00	0.03	1.00	5/5	1.00	0.03	1.00
Tight feelings around the mouth	5/5	1.00	0.03	1.00	5/5	1.00	0.03	1.00
Cold hands or feet	5/5	1.00	0.03	1.00	5/5	1.00	0.03	1.00
Palpitations	5/5	1.00	0.03	1.00	5/5	1.00	0.03	1.00
Feelings of anxiety	5/5	1.00	0.03	1.00	4/5	1.00	0.16	0.76
	S-CVI/UA = 0.88				S-CVI/UA = 0.88			
	S-CVI/Ave = 0.95				S-CVI/Ave = 0.98			

A: The number of experts rated the item as a 3 or 4 on a 4-point scale; I-CVI: Item-content validity index; S-CVI: Scale-level-content validity index; S-CVI/UA: universal agreement among experts; S-CVI/Ave is the average of the item-level CVIs; Pc: Probability of chance agreement; K: Modified kappa statistic coefficient.

Structural validity

The Kaiser-Meyer-Olkin measure of sampling adequacy had a value of 0.795, and Bartlett's test of sphericity was statistically significant ($p < 0.001$), indicating that the assumptions were met to proceed with the exploratory factor analysis. The eigenvalues suggested the presence of four factors. Accordingly, the researchers proposed adding a "psychology" factor as a fourth latent factor for the VNQ, besides the three original factors. For each latent factor, any items with a factor loading below 0.4 were removed. The total variance explained by each factor and the rotated component matrix are presented in Table 2.

The structural validity of the VNQ was evaluated using confirmatory factor analysis, as presented in Figure 1. Based on the four-factor structure established with EFA, the results showed that the factor loadings ranged from 0.31 to 1.72 for the shortness of breath factor, 0.90 to 1.49 for the peripheral tetany factor, 0.24 to 1.50 for the central tetany, and 0.26 to 1.00 for the psychology factor. The model fit indices suggested that the model fit the data well, with a Chi-squared of 104.6 ($p = 0.2$), CFI of 0.96, RMSEA of 0.04, and SRMSR of 0.0785.

Table 2. Rotated Factor Matrix with Kaiser Normalization and Variance explained for each factor

Rotated Factor Matrix [a]					
		Factor			
	Items	Short of breath	Peripheral tetany	Central tetany	Psychology
11	Unable to breathe deeply	0.815	NA	NA	NA
6	Faster breathing	0.783	NA	NA	NA
7	Short of breath	0.522	NA	NA	NA
2	Felling tense	0.500	NA	NA	0.482
15	Palpitations	0.482	NA	NA	0.504
5	Feeling confused	0.420	NA	0.476	NA
12	Stiff fingers or arms	NA	0.738	NA	NA
13	Tight feelings around the mouth	NA	0.706	NA	NA
10	Tingling fingers	NA	0.624	NA	NA
8	Tight feelings in the chest	NA	0.612	NA	NA
14	Cold hands or feet	NA	0.608	NA	0.404
1	Chest pain	NA	NA	0.818	NA
3	Blurred vision	NA	NA	0.787	NA
4	Dizzy spells	NA	NA	0.560	NA
9	Bloated feeling in the stomach	NA	NA	NA	0.833
16	Feelings of anxiety	NA	NA	NA	0.784
Rotation Sums of Squared Loadings	% variance explained	17.618	16.602	14.736	14.078
	Cumulative %			63.034	

Extraction Method: Principal Factor Analysis.
Rotation Method: Varimax with Kaiser Normalization. a
a. Rotation converged in 6 iterations.

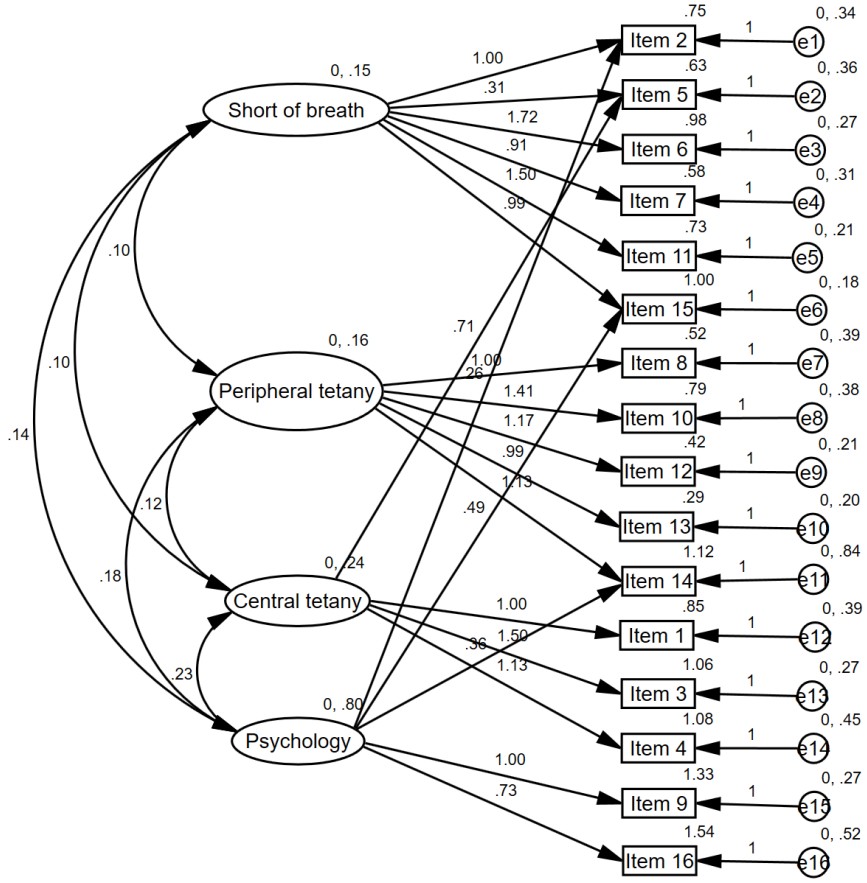


Figure 1. Structural validity for the Vietnamese Nijmegen questionnaire with four factors solution.

Reliability

The internal consistency and test-retest reliability of the VNQ were evaluated using data from 52 participants. The Cronbach's Alpha value for the total score and the individual factors were all above 0.7 (Table 3), indicating good internal consistency. The correlation between each factor and the total score exceeded 0.8.

The test-retest reliability analysis revealed statistically significant ICCs above 0.8 for the total score and factor scores, suggesting good reliability (Table 3). The Bland-Altman analysis showed a bias ranging from -0.58 to 0.08 for the total and factor scores (Table 4, Figure 2). The standard error measurement (SEM) was reported as 0.61 for the total VNQ, 0.29 for shortness of breath, 0.26 for peripheral tetany, 0.24 for central tetany, and 0.24 for psychology. Linear regression analysis demonstrated no proportional bias, with all p values above 0.05 (Table 4).

Table 3. Internal consistency and Test-retest reliability for the VNQ

	Correlation	Cronbach's α	ICC	95% IC
Total		0.89	0.93*	0.87 – 0.96
Shortness of breath	0.89	0.82	0.91*	0.84 – 0.95
Peripheral tetany	0.82	0.74	0.87*	0.77 – 0.92
Central tetany	0.78	0.75	0.85*	0.74 – 0.92
Psychology	0.89	0.80	0.93*	0.89 – 0.96

VNQ: Vietnamese Nijmegen questionnaire
*: p value < 0.001
ICC: Intraclass Correlation Coefficient, IC: Interval Confidence

Table 4. Bias and Limit of agreement for test-retest reliability for VNQ

Limit of agreement	Bias	Upper	Lower	P-value of linear regression
Total	-0.58	8.1	-9.2	0.54
Shortness of breath	0.08	4.20	-4.04	0.22
Peripheral tetany	-0.31	3.43	-4.05	0.10
Central tetany	-0.23	3.14	-3.60	0.10
Psychology	-0.40	3.00	-3.80	0.65

VNQ: Vietnamese Nijmegen questionnaire

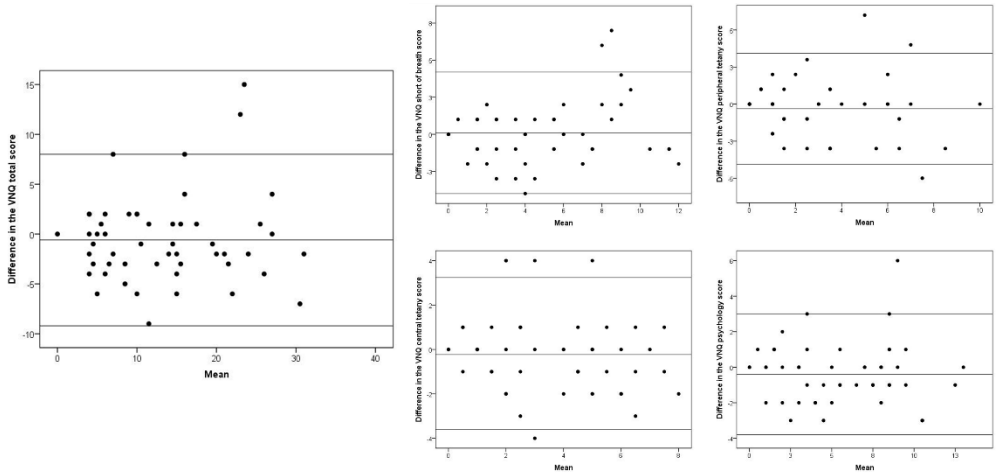


Figure 2. The Bland-Altman plot of total and factor scores for VNQ. Solid lines represent the bias mean. Upper and Lower dashed lines represent the upper and lower limits of agreement. VNQ: Vietnamese Nijmegen questionnaire

Discussion

Due to the unavailability of gold measurements for HVS [310,315], and the lack of guidance for HVS diagnosis in Vietnam, this study investigated the cross-cultural (content, face) and structural validity. The translation process of the VNQ strictly adhered to Beaton's guidelines. This study employed two independent expert panels: one to collect opinions and modify the VNQ, and another to investigate content validity. Based on the first panel's recommendations, minor changes were made to the questionnaire. The second expert panel then evaluated the content validity. This approach enhanced the objective assessment of content validity.

The content validity results indicated an excellent corresponding index. Only the "feeling tense" item had a poor I-CVI for clarity, while the other I-CVIs were excellent in both clarity and relevance [293]. The overall content validity of the questionnaire was excellent, with both S-CVI/UA and S-CVI/Ave for clarity and relevance above 0.8 and 0.9, respectively [293]. Compared to a critical review of the Nijmegen questionnaire, our study reported better content validity than the original study [325]. This may be due to a better understanding of HVS and the NQ. This study also reported good face validity, with all the participants rating the VNQ as understandable or easy to understand. According to our search, this is the first study to report face validity through an easy-to-understand level from a target population. This result was not surprising since the questionnaire only included symptom listing without any complex medical terminologies or ambiguous sentences.

The EFA results showed that the hypothesized 3-factor structure of the original NQ was not replicated in the VNQ. The result suggested the presence of a fourth latent factor. This fourth factor included 5 explained items (Feeling tense, Bloated feeling in the stomach, Cold hands or feet, palpitations, and Feelings of anxiety), which the researchers named the "psychology" factor. The new factor finding is not surprising considering that HVS is commonly accompanied by psychological symptoms such as anxiety and panic [326]. This finding is

supported by another study that also detected new factors compared to the original version [327], suggesting the cross-cultural adaptations of questionnaires may reveal new factors that better fit the target population. Additionally, the cumulative explained variance was 64%, higher than that reported in two validation studies of the NQ for patients with asthma (54% to 58%), potentially attributable to the influence of the disease characteristics [327,328]. The results of the CFA showed that the four-factor model, consisting of the three original factors of the NQ (shortness of breath, peripheral tetany, and central tetany) and an additional "psychology" factor, provided a good fit model to the observed data for the VNQ. The factor loadings for the four factors were acceptable [322]. The model fit indices, including a Chi-squared, CFI, RMSEA, and SRMSR, all indicated a good fit for the model [322], supporting the 4-factor structure of the VNQ.

The reliability analysis showed excellent results, with very high Cronbach's Alpha values for total score and factor scores. All factor values exceeded the acceptable level of 0.7, and the total score was good (> 0.8) [323]. This study reported similar internal consistency results to the original study [318], though the high number of items raises the potential risk of high-value bias [2].

This study also reported excellent ICC values for absolute agreement for the total score and good ICC values for each factor [152]. The results were comparable to another study [329]. To enhance the accuracy of the reliability assessment as required by the COSMIN [2,298], the study used the GRC despite the potential risk of "recall bias" associated with this method [319]. The Bland-Altman analysis demonstrated a low systematic error, with biases of -0.58 for the total score, 0.08 for shortness of breath, -0.31 for peripheral tetany, -0.23 for central tetany, and -0.4 for psychology. The random errors were relatively wide for the total score (-9 to 8), and narrower for each factor, suggesting a cautious interpretation of VNQ total score changes in the healthy population. The Bland-Altman plots indicated that the increased mean difference does not change the test-retest

difference [296]. Additionally, the statistically non-significant result of the linear regression analysis rejected a proportional bias [2].

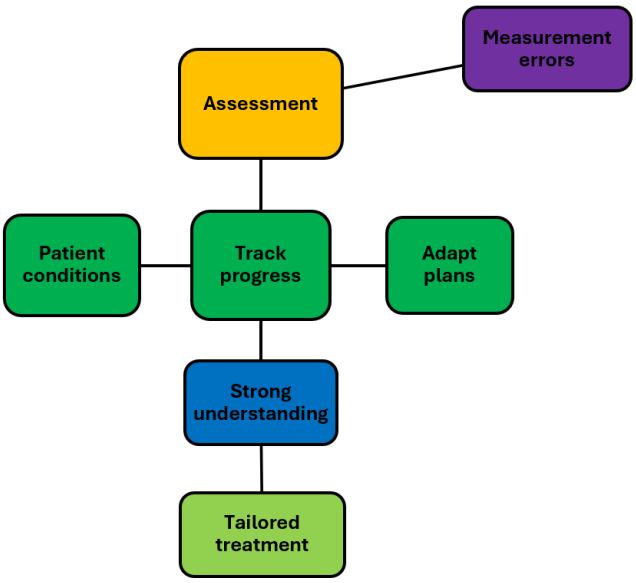
This study employed a rigorous methodology, featuring two expert panels for assessing content validity. Furthermore, the structural validity analysis identified a new latent psychological factor, potentially improving targeted interventions for individuals with HVS. The study also applied control tests when assessing reliability, enhancing confidence in the results. However, the small sample size limited conclusions about parameter estimates and construct validity. Additionally, the study was conducted on participants without reported health issues, so caution is warranted when applying the VNQ to other populations, such as those with asthma or other chronic lung diseases.

Conclusion

The VNQ has excellent content validity, as shown by the thorough translation process and two expert panels evaluating the questionnaire items' clarity and relevance. The study also found the VNQ has acceptable structural validity, with a four-factor model fitting the data well and including an additional "psychology" factor not in the original Nijmegen Questionnaire. This suggests the VNQ may better capture Hyperventilation Syndrome's complexity. Additionally, the VNQ has reliable psychometric properties, with high internal consistency and test-retest reliability, indicating a consistent measurement. These findings make the VNQ a valuable tool for primary healthcare providers in Vietnam to effectively screen and identify individuals with Hyperventilation Syndrome, a crucial step in delivering targeted interventions and improving patient outcomes.

Chapter 4: General discussion and future studies

Three years ago, I began my thesis, driven by the central question of how to improve physiotherapy services in Vietnam. It quickly became clear that to be effective, physiotherapists must possess a strong understanding of their objectives and the rationale behind their interventions. This requires a clear understanding of the patient's current condition, the ability to track progress methodically, and the capacity to adapt intervention plans appropriately; assessment, therefore, plays a critical role. Moreover, a thorough assessment can lead to improved clinical reasoning, empowering therapists to tailor treatment plans to the individual conditions and expectations of their patients. Thus, investigating and adapting excellent available assessment tools was the main aim of the thesis from the very beginning. However, when I was introduced to the field of medical measurement, I was surprised to find that it is not as straightforward as it appears.



Assessment role in Physiotherapy

First, the validity of an assessment is highly context-specific; an assessment that works excellently in one population may be inappropriate in a different one [330]. For instance, questions about walking capacity can be valuable for assessing the quality of life in individuals with stroke, but may not be meaningful for individuals with gastroesophageal reflux due to the difference in health impairment between the two populations. Another example is how we can ensure that a list of activities evaluates a child's development, considering that development is a brain-learning process requiring experiential exposure, which can vary tremendously across cultures [305]. The overprotective nature in Asian countries, for example, can delay development compared to Western countries, where children are encouraged to be independent from a young age [331]. Secondly, due to the error inherent in measurement and the natural variability of patients, interpreting assessment outcomes is not always straightforward or certain. If a change happens, how can we interpret whether it is a true change due to the impact of interventions or merely measurement error? Considering that a measurement result can be influenced by many factors, such as the environment, the assessor’s experience, the patient’s motivation and concentration, equipment conditions, medication usage, and the time of day that the assessment is given. These factors can introduce systematic or random errors into the measurement process, affecting the reliability of the results [298]. The potential for erroneous measurements to cause medical errors and adverse events presents challenges in clinical practice [332]. Fortunately, the science of validity and reliability was introduced to mitigate error and improve accuracy. However, identifying the minimal detectable change and the minimal clinically important difference is crucial for accurately interpreting changes in patient status and ensuring that detected improvements are both statistically significant and practically meaningful. Surprisingly, determining true changes is not easy, especially given the inherent variability in human performance and the contextual influences on task execution. This complexity is further compounded by diversity

and disagreement in methodology and statistical techniques employed across studies to establish these thresholds. For example, distribution-based methods alone have more than five ways to calculate, leading to confusion, and the inconsistent application of these methods hampers the establishment of universally accepted thresholds for MDC and MID. Another issue in assessment is the lack of reference norms for different ethnicities, particularly in low- and middle-income countries, which is due to limited research capacity. Without reference norms, some assessment tools may become less useful. For instance, the same spirometry result obtained from a Vietnamese individual and a Caucasian individual with the same height, age, and gender may have different implications. However, if the same reference range, which has not been investigated in Vietnamese populations, is used, the Vietnamese individual might receive the same medical treatment, which could be unnecessary or even harmful. Thus, this thesis represents a modest step towards enriching the assessment tools for physiotherapists in Vietnam, based on the medical measurement science.

Respiratory assessment is one of the areas this thesis addresses. The respiratory system requires harmony between the lungs, the mechanical pump, and the central control to maintain optimal function. However, in Vietnam, physiotherapists often overemphasize the mucus in the management of respiratory disorders, while the central control and mechanical pump receive inadequate attention. In 2024, a survey investigated the existing clinical practice in respiratory physiotherapy in Vietnam, revealing that most responses focused on concerns about mucus retention without mentioning the respiratory muscle. This discrepancy highlights the need for a more holistic approach to respiratory physiotherapy, one that considers all components of the respiratory system. To address this gap, a study was conducted to investigate respiratory muscle norms and establish predictive values for MIP, MEP, and SNIP, aiming to create an initial database for respiratory muscle assessment, to be followed by

targeted training. However, given that predictive values are influenced by physical condition, a revisit study should be conducted every 10 to 20 years to update the predictive equation, taking into account improvements in physical condition alongside socioeconomic development. For instance, according to the Ministry of Health, there have been dramatic increases in the average height, a predictive variable, and overall nutrition of Vietnamese individuals in the last 20 years.

After completing a study on predicting values for MIP, MEP, and SNIP, we realized that while we have established a necessary reference norm, a sufficient condition - standardized equipment for measuring respiratory pressure - remains limitedly accessible in Vietnam and potentially other low-income countries. This barrier sparked a desire to find an alternative method, particularly for inspiratory muscle assessment and training, given its prominent role in current literature. Consequently, we tested the idea of using the Borg CR10 scale to identify IMT intensity and estimate MIP, with encouraging results for using the Borg CR10 scale as an alternative method in healthy subjects. This new approach diversifies assessing and training options, therefore, enhancing accessibility to IMT in various settings. This is especially relevant in low- and middle-income countries where access to gold-standard equipment is challenging. However, these results cannot yet be extrapolated to other populations, particularly individuals with dyspnea symptoms that could impair their perception of exertion, such as persons with COPD or HF. Further studies employing the same protocol should be conducted across different populations to validate these findings.

Another tool investigated in this thesis is the 1STST, which was the subject of two studies. The first study, focusing on the application of the 1STST in individuals with heart failure, revealed that a difference of 2 repetitions can serve as a minimal important difference for a quick evaluation of exercise capacity in this population. Considering the context of persistent hospital overload in Vietnam and a

healthcare worker-to-habitat ratio of approximately 10 per 10,000, clinical implementation could be significantly impacted. Notably, exercise capacity is a reliable indicator for predicting hospitalization and mortality in this population, potentially empowering persons with HF to easily self-monitor their exercise capacity. In scientific terms, these findings offer a valuable reference point for determining sample sizes in future clinical trials. An ambitious study predicting mortality and hospitalization using the 1STST could be an interesting and practical project. The second study examined the correlation between 1STST and MIP in persons with COPD. The low correlation suggests that inspiratory muscle strength may contribute to exercise capacity. However, the influence of inspiratory muscle strength on the 1STST may be less than in the walking test, which showed a higher correlation [134]. This underscores the natural difference between the walking test, which involves more global effort, and the 1STST, which involves more localized effort in the leg, as demonstrated by their correlation with VO2max in previous studies [333,334]. Therefore, test selection and interpretation should be carefully considered. For instance, if testing time and environment allow, the 6MWT may be the best field test for assessing global effort in exercise capacity. However, if a simpler method is needed to empower individuals with chronic diseases to self-manage, the 1STST is more appropriate.

The thesis also included the cross-cultural validation of the Vietnamese versions of the Paediatric Evaluation of Disability Inventory and the Nijmegen Questionnaire, which further confirms the necessity of a rigorous and comprehensive translation and validation process to ensure appropriate modifications that align with the new culture. This approach is vital to ensure that the tools are not only linguistically accurate but also culturally relevant and contextually appropriate, as potential latent factors can appear in new cultures and populations. To ensure the validity and reliability of these tools in the new cultural context, it is essential to consider the specific culture, lifestyle, and linguistic characteristics of the

population. This involves adapting the tools to account for differences in language, customs, and social norms, ensuring that they accurately capture the intended constructs. For example, children in Vietnam rarely use forks for eating; thus, any evaluation relating to fork usage may not capture their real function, not to mention that it is not useful to evaluate that activity. Furthermore, rigorous statistical analyses should be conducted to assess the psychometric properties of the adapted tools, including their reliability and validity, and sensitivity of changes. By following these steps, researchers and clinicians can ensure that the assessment tools are appropriate for use in the new cultural context and provide meaningful information about the health and well-being of individuals.

In conclusion, this body of work contributes to the advancement of physiotherapy practice in Vietnam by providing validated assessment tools, representing an initial effort to address the assessment gap in Vietnamese physiotherapy and potentially enhancing service quality. The researchers hope that these initial studies will encourage other physiotherapists to validate additional questionnaires or assessments necessary within their scope of practice. Furthermore, we advocate for further research into alternative assessment methods suitable for low-resource settings. After all, without proper assessment and evaluation, interventions are like a game of luck.

“A smart mother often makes a better diagnosis than a poor doctor.”

August Bier

References:

1. French MA, Roemmich RT, Daley K, et al. Precision rehabilitation: optimizing function, adding value to health care. *Arch Phys Med Rehabil*. 2022;103(6):1233-1239. https://doi.org/10.1016/j.apmr.2022.01.154
2. De Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine*. Cambridge university press; 2011.
3. Al-Worafi YM. Physiotherapy Research in Developing Countries. *Handbook of Medical and Health Sciences in Developing Countries: Education, Practice, and Research*. Springer; 2024:1-28.
4. Terwee C, Dekker F, Wiersinga W, Prummel M, Bossuyt P. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Qual Life Res*. 2003;12:349-362. https://doi.org/10.1023/a:1023499322593
5. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-249. https://doi.org/10.3322/caac.21660
6. Association AL. *State of Lung Cancer: 2020 Report*. 2020. 2021. Accessed January 19, 2022. https://www.lung.org/getmedia/381ca407-a4e9-4069-b24b-195811f29a00/solc2020-report-final.pdf
7. Howlader N, Forjaz G, Mooradian MJ, et al. The effect of advances in lung-cancer treatment on population mortality. *New England Journal of Medicine*. 2020;383(7):640-649.
8. Jones GS, Baldwin DR. Recent advances in the management of lung cancer. *Clin Med (Northfield Il)*. 2018;18(2):s41-s46. https://doi.org/10.7861/clinmedicine.18-2-s41
9. Rueth NM, Parsons HM, Habermann EB, et al. The long-term impact of surgical complications after resection of stage I nonsmall cell lung cancer: a population-based survival analysis. *Annals of surgery*. 2011;254(2):368-374.
10. Detillon DD, Veen EJ. Postoperative outcome after pulmonary surgery for non-small cell lung cancer in elderly patients. *The Annals of Thoracic Surgery*. 2018;105(1):287-293. https://doi.org/10.1016/j.athoracsur.2017.07.032
11. Poghosyan H, Sheldon LK, Leveille SG, Cooley ME. Health-related quality of life after surgical treatment in patients with non-small cell lung cancer: a systematic review. *Lung Cancer*. 2013;81(1):11-26. https://doi.org/10.1016/j.lungcan.2013.03.013
12. Jones LW, Hornsby WE, Goetzinger A, et al. Prognostic significance of functional capacity and exercise behavior in patients with metastatic non-small cell lung cancer. *Lung Cancer*. 2012;76(2):248-252. https://doi.org/10.1016/j.lungcan.2011.10.009

13. Jones LW, Watson D, Herndon JE, et al. Peak oxygen consumption and long-term all-cause mortality in nonsmall cell lung cancer. *Cancer*. 2010;116(20):4825-4832. https://doi.org/10.1002/cncr.25396
14. Varela G, Jiménez MF, Novoa N, Aranda JL. Estimating hospital costs attributable to prolonged air leak in pulmonary lobectomy. *Eur J Cardiothorac Surg*. 2005;27(2):329-333. https://doi.org/10.1016/j.ejcts.2004.11.005
15. Varela G, Novoa NM, Agostini P, Ballesteros E. Chest physiotherapy in lung resection patients: state of the art. Elsevier; 2011:297-306.
16. Buchman A, Boyle P, Wilson R, Leurgans S, Shah R, Bennett D. Respiratory muscle strength predicts decline in mobility in older persons. *Neuroepidemiology*. 2008;31(3):174-180. https://doi.org/10.1159/000154930
17. Greising SM, Ottenheijm CA, O'Halloran KD, Barreiro E. Diaphragm plasticity in aging and disease: therapies for muscle weakness go from strength to strength. *J Appl Physiol*. 2018;125(2):243-253. https://doi.org/10.1152/japplphysiol.01059.2017
18. Mans CM, Reeve JC, Gasparini CA, Elkins MR. Postoperative outcomes following preoperative inspiratory muscle training in patients undergoing open cardiothoracic or upper abdominal surgery: protocol for a systematic review. *Systematic reviews*. 2012;1(1):63. https://doi.org/10.1186/2046-4053-1-63
19. Pegorari M, dos Santos NO, Monteiro R, Matos A, Jamami M. Respiratory muscle strength as a discriminator of sarcopenia in community-dwelling elderly: a cross-sectional study. *The Journal of nutrition, health and aging*. 2018;22(8):952-958. https://doi.org/10.1007/s12603-018-1079-4
20. Wijdicks EF. The neurology of acutely failing respiratory mechanics. *Ann Neurol*. 2017;81(4):485-494. https://doi.org/10.1002/ana.24908
21. Wang YQ, Liu X, Jia Y, Xie J. Impact of breathing exercises in subjects with lung cancer undergoing surgical resection: A systematic review and meta-analysis. *J Clin Nurs*. 2019;28(5-6):717-732. https://doi.org/10.1111/jocn.14696
22. Kendall F, Oliveira J, Peleteiro B, Pinho P, Bastos PT. Inspiratory muscle training is effective to reduce postoperative pulmonary complications and length of hospital stay: a systematic review and meta-analysis. *Disabil Rehabil*. 2018;40(8):864-882. https://doi.org/10.1080/09638288.2016.1277396
23. Rosero ID, Ramírez-Vélez R, Lucia A, et al. Systematic review and meta-analysis of randomized, controlled trials on preoperative physical exercise interventions in patients with non-small-cell lung cancer. *Cancers (Basel)*. 2019;11(7):944. https://doi.org/10.3390/cancers11070944
24. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009;339https://doi.org/10.7326/0003-4819-151-4-200908180-00135

25. Kwon Y, Lemieux M, McTavish J, Wathen N. Identifying and removing duplicate records from systematic review searches. *Journal of the Medical Library Association: JMLA*. 2015;103(4):184. https://doi.org/10.3163/1536-5050.103.4.004
26. Sherrington C, Herbert R, Maher C, Moseley A. PEDro. A database of randomized trials and systematic reviews in physiotherapy. *Man Ther*. 2000;5(4):223-226. https://doi.org/10.1054/math.2000.0372
27. Verhagen AP, De Vet HC, De Bie RA, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol*. 1998;51(12):1235-1241. https://doi.org/10.1016/s0895-4356(98)00131-0
28. Valkenet K, van de Port IG, Dronkers JJ, de Vries WR, Lindeman E, Backx FJ. The effects of preoperative exercise therapy on postoperative outcome: a systematic review. *Clin Rehabil*. 2011;25(2):99-111. https://doi.org/10.1177/0269215510380830
29. Liu J-F, Kuo N-Y, Fang T-P, Chen J-O, Lu H-I, Lin H-L. A six-week inspiratory muscle training and aerobic exercise improves respiratory muscle strength and exercise capacity in lung cancer patients after video-assisted thoracoscopic surgery: A randomized controlled trial. *Clin Rehabil*. 2021;35(6):840-850. https://doi.org/10.1177/0269215520980138
30. Brocki BC, Andreasen JJ, Langer D, Souza DSR, Westerdahl E. Postoperative inspiratory muscle training in addition to breathing exercises and early mobilization improves oxygenation in high-risk patients after lung cancer surgery: a randomized controlled trial. *Eur J Cardiothorac Surg*. 2015;49(5):1483-1491. https://doi.org/10.1093/ejcts/ezv359
31. Brocki BC, Andreasen JJ, Westerdahl E. Inspiratory muscle training in high-risk patients following lung resection may prevent a postoperative decline in physical activity level. *Integr Cancer Ther*. 2018;17(4):1095-1102. https://doi.org/10.1177/1534735418796286
32. Messaggi-Sartor M, Marco E, Martínez-Téllez E, et al. Combined aerobic exercise and high-intensity respiratory muscle training in patients surgically treated for non-small cell lung cancer: a pilot randomized clinical trial. *Eur J Phys Rehabil Med*. 2018;55(1):113-122. https://doi.org/10.23736/S1973-9087.18.05156-0
33. Taşkin H, Telli Atalay O, Yuncu G, Taşpinar B, Yalman A, Şenol H. Postoperative respiratory muscle training in addition to chest physiotherapy after pulmonary resection: A randomized controlled study. *Physiother Theory Pract*. 2020;36(3):378-385. https://doi.org/10.1080/09593985.2018.1488189
34. Organization WH. Obesity: preventing and managing the global epidemic. 2000;

35. Nieuwenhuijs D, Bruce J, Drummond G, Warren P, Wraith P, Dahan A. Ventilatory responses after major surgery and high dependency care. *Br J Anaesth*. 2012;108(5):864-871. https://doi.org/10.1093/bja/aes017
36. Agostini P, Cieslik H, Rathinam S, et al. Postoperative pulmonary complications following thoracic surgery: are there any modifiable risk factors? *Thorax*. 2010;65(9):815-818. https://doi.org/10.1136/thx.2009.123083
37. Tenório LHS, Santos AC, Câmara Neto JB, et al. The influence of inspiratory muscle training on diaphragmatic mobility, pulmonary function and maximum respiratory pressures in morbidly obese individuals: a pilot study. *Disabil Rehabil*. 2013;35(22):1915-1920. https://doi.org/10.3109/09638288.2013.769635
38. Daniel M, Lang E, Huynh TM, et al. Prevalence and time-course of diaphragmatic dysfunction following lung resection: A repeated ultrasonic assessment. *Anaesthesia Critical Care & Pain Medicine*. 2022;41(2):101024. https://doi.org/10.1016/j.accpm.2022.101024
39. Jiao W, Zhao Y, Wang M, et al. A retrospective study of diaphragmatic motion, pulmonary function, and quality-of-life following video-assisted thoracoscopic lobectomy in patients with nonsmall cell lung cancer. *Indian J Cancer*. 2014;51(6):45. https://doi.org/10.4103/0019-509X.151991
40. Souza H, Rocha T, Pessoa M, et al. Effects of inspiratory muscle training in elderly women on respiratory muscle strength, diaphragm thickness and mobility. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*. 2014;69(12):1545-1553. https://doi.org/10.1093/gerona/glu182
41. Cesario A, Ferri L, Galetta D, et al. Post-operative respiratory rehabilitation after lung resection for non-small cell lung cancer. *Lung Cancer*. 2007;57(2):175-180. https://doi.org/10.1016/j.lungcan.2007.02.017
42. Kendall F, Abreu P, Pinho P, Oliveira J, Bastos P. The role of physiotherapy in patients undergoing pulmonary surgery for lung cancer. A literature review. *Revista Portuguesa de Pneumologia (English Edition)*. 2017;23(6):343-351. https://doi.org/10.1016/j.rppnen.2017.05.003
43. Pasqua F, Geraneo K, Nardi I, Lococo F, Cesario A. Pulmonary rehabilitation in lung cancer. *Monaldi Archives for Chest Disease*. 2013;79(2)https://doi.org/10.4081/monaldi.2013.95
44. Santos SC, Fréz AR, Ruaro JA, Wouk J, Daniel CR. Risk factors and postoperative complications in oncologic surgeries. *Fisioterapia em Movimento*. 2020;33https://doi.org/10.1590/1980-5918.033.AO53
45. Vassilakopoulos T, Mastora Z, Katsaounou P, et al. Contribution of pain to inspiratory muscle dysfunction after upper abdominal surgery: a randomized controlled trial. *Am J Respir Crit Care Med*. 2000;161(4):1372-1375. https://doi.org/10.1164/ajrccm.161.4.9907082

46. Brocki BC, Westerdahl E, Langer D, Souza DS, Andreasen JJ. Decrease in pulmonary function and oxygenation after lung resection. *ERJ open research*. 2018;4(1)https://doi.org/10.1183/23120541.00055-2017
47. Avancini A, Sartori G, Gkountakos A, et al. Physical activity and exercise in lung cancer care: will promises be fulfilled? *The oncologist*. 2020;25(3):e555-e569. https://doi.org/10.1634/theoncologist.2019-0463
48. Bickel CS, Cross JM, Bamman MM. Exercise dosing to retain resistance training adaptations in young and older adults. *Med Sci Sports Exerc*. 2011;43(7):1177-87. https://doi.org/10.1249/MSS.0b013e318207c15d
49. Martín-Valero R, Jimenez-Cebrian AM, Moral-Munoz JA, de-la-Casa-Almeida M, Rodriguez-Huguet M, Casuso-Holgado MJ. The efficacy of therapeutic respiratory muscle training interventions in people with bronchiectasis: a systematic review and meta-analysis. *Journal of Clinical Medicine*. 2020;9(1):231.
50. Padilha CS, Ribeiro AS, Fleck SJ, et al. Effect of resistance training with different frequencies and detraining on muscular strength and oxidative stress biomarkers in older women. *Age*. 2015;37(5):1-9. https://doi.org/10.1007/s11357-015-9841-6
51. Kadi F, Schjerling P, Andersen LL, et al. The effects of heavy resistance training and detraining on satellite cells in human skeletal muscles. *The Journal of physiology*. 2004;558(3):1005-1012. https://doi.org/10.1113/jphysiol.2004.065904
52. Blocquiaux S, Gorski T, Van Roie E, et al. The effect of resistance training, detraining and retraining on muscle strength and power, myofibre size, satellite cells and myonuclei in older men. *Exp Gerontol*. 2020;133:110860. https://doi.org/10.1016/j.exger.2020.110860
53. Tokmakidis SP, Kalapotharakos VI, Smilios I, Parlavantzas A. Effects of detraining on muscle strength and mass after high or moderate intensity of resistance training in older adults. *Clin Physiol Funct Imaging*. 2009;29(4):316-319. https://doi.org/10.1111/j.1475-097X.2009.00866.x
54. Braun DP, Gupta D, Staren ED. Quality of life assessment as a predictor of survival in non-small cell lung cancer. *BMC cancer*. 2011;11(1):1-9.
55. Agostini PJ, Naidu B, Rajesh P, et al. Potentially modifiable factors contribute to limitation in physical activity following thoracotomy and lung resection: a prospective observational study. *Journal of cardiothoracic surgery*. 2014;9(1):1-6.
56. Granger CL, Parry SM, Edbrooke L, et al. Improving the delivery of physical activity services in lung cancer: a qualitative representation of the patient's perspective. *European Journal of Cancer Care*. 2019;28(1):e12946. https://doi.org/10.1111/ecc.12946

Abstract

57. Sloan JA, Cheville AL, Liu H, et al. Impact of self-reported physical activity and health promotion behaviors on lung cancer survivorship. *Health and quality of life outcomes*. 2016;14(1):1-14. https://doi.org/10.1186/s12955-016-0461-3
58. Möller A, Sartipy U. Long-term health-related quality of life following surgery for lung cancer. *Eur J Cardiothorac Surg*. 2012;41(2):362-367. https://doi.org/10.1016/j.ejcts.2011.05.055
59. Witt JD, Guenette JA, Rupert JL, McKenzie DC, Sheel AW. Inspiratory muscle training attenuates the human respiratory muscle metaboreflex. *The Journal of physiology*. 2007;584(3):1019-1028. https://doi.org/10.1113/jphysiol.2007.140855
60. Rogers LJ, Bleetman D, Messenger DE, et al. The impact of enhanced recovery after surgery (ERAS) protocol compliance on morbidity from resection for primary lung cancer. *The Journal of thoracic and cardiovascular surgery*. 2018;155(4):1843-1852. https://doi.org/10.1016/j.jtcvs.2017.10.151
61. Bade BC, Thomas DD, Scott JB, Silvestri GA. Increasing physical activity and exercise in lung cancer: reviewing safety, benefits, and application. *J Thorac Oncol*. 2015;10(6):861-871. https://doi.org/10.1097/JTO.0000000000000536.
62. Cooper BG. Republished review: An update on contraindications for lung function testing. *Postgrad Med J*. 2011;87(1032):724-733. https://doi.org/10.1136/pgmj.2010.139881rep
63. Ponikowski P, Anker SD, AlHabib KF, et al. Heart failure: preventing disease and death worldwide. *ESC heart failure*. 2014;1(1):4-25.
64. Benjamin EJ, Muntner P, Alonso A, et al. Heart disease and stroke Statistics-2019 update a report from the American Heart Association. *Circulation*. 2019;https://doi.org/10.1161/CIR.0000000000000659
65. Maggioni AP, Dahlström U, Filippatos G, et al. EURObservational Research Programme: regional differences and 1-year follow-up results of the Heart Failure Pilot Survey (ESC-HF Pilot). *Eur J Heart Fail*. 2013;15(7):808-817.
66. García-Olmos L, Batlle M, Aguilar R, et al. Disability and quality of life in heart failure patients: a cross-sectional study. *Fam Pract*. 2019;36(6):693-698. https://doi.org/10.1093/fampra/cmz017
67. Ponikowski P, Voors AA, Anker SD, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur J Heart Fail*. 2016;18(8):891-975. https://doi.org/10.1002/ejhf.592
68. Shih ML, Chen HM, Chou FH, Huang YF, Lu C-H, Chien HC. Quality of life and associated factors in patients with heart failure. *Hu li za zhi The Journal of Nursing*. 2010;57(6):61-71.

69. Seto E, Leonard KJ, Cafazzo JA, Masino C, Barnsley J, Ross HJ. Self-care and quality of life of heart failure patients at a multidisciplinary heart function clinic. *J Cardiovasc Nurs*. 2011;26(5):377-385. https://doi.org/10.1097/JCN.0b013e31820612b8
70. Long L, Mordi IR, Bridges C, et al. Exercise-based cardiac rehabilitation for adults with heart failure. *Cochrane Database Syst Rev*. 2019;(1)https://doi.org/10.1002/14651858.CD003331.pub5
71. Jones J, Hayward E, Hinton S. *Physical Activity and Exercise in the Management of Cardiovascular Disease*. Bristish Association for Cardiovascular Prevention and Rehabilitation; 2018.
72. Balady GJ, Williams MA, Ades PA, et al. Core components of cardiac rehabilitation/secondary prevention programs: 2007 update: A scientific statement from the american heart association exercise, cardiac rehabilitation, and prevention committee, the council on clinical cardiology; the councils on cardiovascular nursing, epidemiology and prevention, and nutrition, physical activity, and metabolism; and the american association of cardiovascular and pulmonary rehabilitation. *Circulation*. 2007;115(20):2675-2682.
73. Agiovlasitis S, Baruth M, Baynard T, et al. *ACSM's guidelines for exercise testing and prescription*. 10th ed. Wolters Kluwer; 2017.
74. De Vos C, Li X, Van Vlaenderen I, et al. Participating or not in a cardiac rehabilitation programme: factors influencing a patient's decision. *European journal of preventive cardiology*. 2013;20(2):341-348. https://doi.org/10.1177/2047487312437057
75. Daw P, Withers TM, van Zanten JJ, Harrison A, Greaves CJ. A systematic review of provider-and system-level factors influencing the delivery of cardiac rehabilitation for heart failure. *BMC Health Serv Res*. 2021;21(1):1-14. https://doi.org/10.1186/s12913-021-07174-w
76. Meyer, Zugck C, Haass M, et al. Inefficient ventilation and reduced respiratory muscle capacity in congestive heart failure. *Basic Res Cardiol*. 2000;95(4):333-342. https://doi.org/10.1007/s003950070053
77. Krishnan, Wang Y. Inspiratory Muscle Weakness and Dyspnea in Chronic Heart Failure. *Am Rev Respir Dis*. 1992;146:467-472. https://doi.org/10.1164/ajrccm/146.2.467
78. Yamada K, Kinugasa Y, Sota T, et al. Inspiratory muscle weakness is associated with exercise intolerance in patients with heart failure with preserved ejection fraction: a preliminary study. *Journal of Cardiac Failure*. 2016;22(1):38-47.
79. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. 2018;169(7):467-473. https://doi.org/10.7326/M18-0850

80. Deeks JJ, Higgins JP. Statistical algorithms in review manager 5. *Statistical Methods Group of The Cochrane Collaboration*. 2010;1(11)
81. Wallace BC, Lajeunesse MJ, Dietz G, et al. Open MEE: Intuitive, open-source software for meta-analysis in ecology and evolutionary biology. *Methods Ecol Evol*. 2017;8(8):941-947. https://doi.org/10.1111/2041-210X.12708
82. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol*. 2005;5(1):1-10. https://doi.org/10.1186/1471-2288-5-13
83. Haidich A-B. Meta-analysis in medical research. *Hippokratia*. 2010;14(Suppl 1):29.
84. Deeks JJ, Higgins JP, Altman DG, Group CSM. Analysing data and undertaking meta-analyses. *Cochrane handbook for systematic reviews of interventions*. 2019:241-284. https://doi.org/10.1002/9781119536604.ch10
85. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557-560. https://doi.org/10.1136/bmj.327.7414.557
86. Higgins JP, Savović J, Page MJ, Elbers RG, Sterne JA. Assessing risk of bias in a randomized trial. *Cochrane handbook for systematic reviews of interventions*. 2019:205-228. https://doi.org/10.1002/9781119536604.ch8
87. Bosnak-Guclu M, Arikan H, Savci S, et al. Effects of inspiratory muscle training in patients with heart failure. *Respir Med*. 2011;105(11):1671-1681. https://doi.org/10.1016/j.rmed.2011.05.001
88. Laoutaris I, Dritsas A, Brown MD, Manginas A, Alivizatos PA, Cokkinos DV. Inspiratory muscle training using an incremental endurance test alleviates dyspnea and improves functional status in patients with chronic heart failure. *European Journal of Preventive Cardiology*. 2004;11(6):489-496. https://doi.org/10.1097/01.hjr.0000152242.51327.63
89. Laoutaris ID, Dritsas A, Brown MD, et al. Effects of inspiratory muscle training on autonomic activity, endothelial vasodilator function, and N-terminal pro-brain natriuretic peptide levels in chronic heart failure. *J Cardiopulm Rehabil Prev*. 2008;28(2):99-106. https://doi.org/10.1097/01.HCR.0000314203.09676.b9
90. Laoutaris ID, Dritsas A, Brown MD, et al. Immune response to inspiratory muscle training in patients with chronic heart failure. *European Journal of Preventive Cardiology*. 2007;14(5):679-686. https://doi.org/10.1097/HJR.0b013e3281338394
91. Palau P, Domínguez E, Núñez E, et al. Effects of inspiratory muscle training in patients with heart failure with preserved ejection fraction. *European journal of preventive cardiology*. 2014;21(12):1465-1473. https://doi.org/10.1177/2047487313498832

92. Mello PR, Guerra GM, Borile S, et al. Inspiratory muscle training reduces sympathetic nervous activity and improves inspiratory muscle weakness and quality of life in patients with chronic heart failure: a clinical trial. *J Cardiopulm Rehabil Prev*. 2012;32(5):255-261. https://doi.org/10.1097/HCR.0b013e31825828da
93. Dall'Ago P, Chiappa GR, Guths H, Stein R, Ribeiro JP. Inspiratory muscle training in patients with heart failure and inspiratory muscle weakness: a randomized trial. *J Am Coll Cardiol*. 2006;47(4):757-763. https://doi.org/10.1016/j.jacc.2005.09.052
94. Seidell JC, Flegal KM. Assessing obesity: classification and epidemiology. *Br Med Bull*. 1997;53(2):238-252. https://doi.org/10.1093/oxfordjournals.bmb.a011611
95. Rodrigues A, Da Silva ML, Berton DC, et al. Maximal inspiratory pressure: does the choice of reference values actually matter? *Chest*. 2017;152(1):32-39. https://doi.org/10.1016/j.chest.2016.11.045
96. Laoutaris ID, Dritsas A, Brown MD, Manginas A, Alivizatos PA, Cokkinos DV. Inspiratory muscle training using an incremental endurance test alleviates dyspnea and improves functional status in patients with chronic heart failure. *European Journal of Cardiovascular Prevention Rehabilitation*. 2004;11(6):489-496. https://doi.org/10.1097/01.hjr.0000152242.51327.63
97. Callegaro CC, Martinez D, Ribeiro PA, Brod M, Ribeiro JP. Augmented peripheral chemoreflex in patients with heart failure and inspiratory muscle weakness. *Respiratory Physiology Neurobiology*. 2010;171(1):31-35. https://doi.org/10.1016/j.resp.2010.01.009
98. Chiappa GR, Roseguini BT, Vieira PJ, et al. Inspiratory muscle training improves blood flow to resting and exercising limbs in patients with chronic heart failure. *J Am Coll Cardiol*. 2008;51(17):1663-1671. https://doi.org/10.1016/j.jacc.2007.12.045
99. Ciani O, Piepoli M, Smart N, et al. Validation of exercise capacity as a surrogate endpoint in exercise-based rehabilitation for heart failure: a meta-analysis of randomized controlled trials. *JACC: Heart Failure*. 2018;6(7):596-604. https://doi.org/10.1016/j.jchf.2018.03.017
100. Demers C, McKelvie RS, Negassa A, Yusuf S, Investigators RPS. Reliability, validity, and responsiveness of the six-minute walk test in patients with heart failure. *Am Heart J*. 2001;142(4):698-703. https://doi.org/10.1067/mhj.2001.118468
101. Ferreira JP, Duarte K, Graves TL, et al. Natriuretic peptides, 6-min walk test, and quality-of-life questionnaires as clinically meaningful endpoints in HF trials. *J Am Coll Cardiol*. 2016;68(24):2690-2707. https://doi.org/10.1016/j.jacc.2016.09.936

102. Giannitsi S, Bougiakli M, Bechlioulis A, Kotsia A, Michalis LK, Naka KK. 6-minute walking test: a useful tool in the management of heart failure patients. *Ther Adv Cardiovasc Dis*. 2019;13:1753944719870084. https://doi.org/10.1177/1753944719870084
103. Ingle L, Shelton RJ, Rigby AS, Nabb S, Clark AL, Cleland JG. The reproducibility and sensitivity of the 6-min walk test in elderly patients with chronic heart failure. *Eur Heart J*. 2005;26(17):1742-1751. https://doi.org/10.1093/eurheartj/ehi259
104. O'Neill JO, Young JB, Pothier CE, Lauer MS. Peak oxygen consumption as a predictor of death in patients with heart failure receiving β-blockers. *Circulation*. 2005;111(18):2313-2318. https://doi.org/10.1161/01.CIR.0000164270.72123.18
105. Keteyian SJ, Patel M, Kraus WE, et al. Variables measured during cardiopulmonary exercise testing as predictors of mortality in chronic systolic heart failure. *J Am Coll Cardiol*. 2016;67(7):780-789. https://doi.org/10.1016/j.jacc.2015.11.050
106. Myers J, Kaykha A, George S, et al. Fitness versus physical activity patterns in predicting mortality in men. *The American Journal of Medicine*. 2004;117(12):912-918. https://doi.org/10.1016/j.amjmed.2004.06.047
107. Feuerstadt P, Chai A, Kligfield P. Submaximal effort tolerance as a predictor of all-cause mortality in patients undergoing cardiac rehabilitation. *Clinical Cardiology: An International Indexed Peer-Reviewed Journal for Advances in the Treatment of Cardiovascular Disease*. 2007;30(5):234-238. https://doi.org/10.1002/clc.20076
108. Enright SJ, Unnithan VB, Heward C, Withnall L, Davies DH. Effect of high-intensity inspiratory muscle training on lung volumes, diaphragm thickness, and exercise capacity in subjects who are healthy. *Phys Ther*. 2006;86(3):345-354.
109. Bassett DR, Howley ET. Limiting factors for maximum oxygen uptake and determinants of endurance performance. *Med Sci Sports Exerc*. 2000;32(1):70-84. https://doi.org/10.1097/00005768-200001000-00012
110. Wong CY, Chaudhry SI, Desai MM, Krumholz HM. Trends in comorbidity, disability, and polypharmacy in heart failure. *The American journal of medicine*. 2011;124(2):136-143. https://doi.org/10.1016/j.amjmed.2010.08.017
111. Slater M, Perruccio AV, Badley EM. Musculoskeletal comorbidities in cardiovascular disease, diabetes and respiratory disease: the impact on activity limitations; a representative population-based study. *BMC Public Health*. 2011;11(1):1-7. https://doi.org/10.1186/1471-2458-11-77
112. Nielen MM, van Sijl AM, Peters MJ, Verheij RA, Schellevis FG, Nurmohamed MT. Cardiovascular disease prevalence in patients with inflammatory arthritis, diabetes mellitus and osteoarthritis: a cross-sectional

study in primary care. *BMC Musculoskelet Disord*. 2012;13(1):1-5. https://doi.org/10.1186/1471-2474-13-150
113. Hwang SL, Liao WC, Huang TY. Predictors of quality of life in patients with heart failure. *Jpn J Nurs Sci*. 2014;11(4):290-298. https://doi.org/10.1111/jjns.12034
114. Terakado S, Takeuchi T, Miura T, et al. Early occurrence of respiratory muscle deoxygenation assessed by near-infrared spectroscopy during leg exercise in patients with chronic heart failure. *Jpn Circ J*. 1999;63(2):97-104. https://doi.org/10.1253/jcj.63.97
115. Tanaka T, Basoudan N, Melo LT, et al. Deoxygenation of inspiratory muscles during cycling, hyperpnoea and loaded breathing in health and disease: a systematic review. *Clin Physiol Funct Imaging*. 2018;38(4):554-565. https://doi.org/10.1111/cpf.12473
116. Mcparland C, Krishnan B, Wang Y, Gallagher CG. Inspiratory muscle weakness and dyspnea in chronic heart failure. *Am Rev Respir Dis*. 1992;146(2):467-472. https://doi.org/10.1164/ajrccm/146.2.467
117. Mancini DM, Henson D, LaManca J, Levine S. Respiratory muscle function and dyspnea in patients with chronic congestive heart failure. *Circulation*. 1992;86(3):909-918. https://doi.org/10.1161/01.cir.86.3.909
118. Kupper N, Bonhof C, Westerhuis B, Widdershoven J, Denollet J. Determinants of dyspnea in chronic heart failure. *J Card Fail*. 2016;22(3):201-209. https://doi.org/10.1016/j.cardfail.2015.09.016
119. Ramasamy R, Hildebrandt T, O'Hea E, et al. Psychological and social factors that correlate with dyspnea in heart failure. *Psychosomatics*. 2006;47(5):430-434. https://doi.org/10.1176/appi.psy.47.5.430
120. Reyes A, Ziman M, Nosaka K. Respiratory muscle training for respiratory deficits in neurodegenerative disorders: a systematic review. *Chest*. 2013;143(5):1386-1394. https://doi.org/10.1378/chest.12-1442
121. Schoenfeld BJ. Is there a minimum intensity threshold for resistance training-induced hypertrophic adaptations? *Sports Med*. 2013;43(12):1279-1288. https://doi.org/10.1007/s40279-013-0088-z
122. Morton RW, Colenso-Semple L, Phillips SM. Training for strength and hypertrophy: an evidence-based approach. *Current Opinion in Physiology*. 2019;10:90-95. https://doi.org/10.1016/j.cophys.2019.04.006
123. Miyamoto T, Sotobayashi D, Ito G, et al. Physiological role of anticipatory cardiorespiratory responses to exercise. *Physiological Reports*. 2022;10(5):e15210. https://doi.org/10.14814/phy2.15210
124. Jensen MT. Resting heart rate and relation to disease and longevity: past, present and future. *Scand J Clin Lab Invest*. 2019;79(1-2):108-116. https://doi.org/10.1080/00365513.2019.1566567

125. Vogel CU, Wolpert C, Wehling M. How to measure heart rate? *Eur J Clin Pharmacol*. 2004;60(7):461-466. https://doi.org/10.1007/s00228-004-0795-3
126. Bundgaard JS, Thune JJ, Gislason G, et al. Quality of life and the associated risk of all-cause mortality in nonischemic heart failure. *Int J Cardiol*. 2020;305:92-98. https://doi.org/10.1016/j.ijcard.2020.02.008
127. Arnold M, Rajda M, Ignaszewski A, et al. Changes in the Minnesota living with heart failure questionnaire score and clinical outcomes in a large contemporary population of ambulatory heart failure patients in the Canadian Heart Failure Network. *J Card Fail*. 2012;18(8):S79. https://doi.org/10.1016/j.cardfail.2012.06.493
128. Hoekstra T, Jaarsma T, van Veldhuisen DJ, Hillege HL, Sanderman R, Lesman-Leegte I. Quality of life and survival in patients with heart failure. *Eur J Heart Fail*. 2013;15(1):94-102. https://doi.org/10.1093/eurjhf/hfs148
129. Azambuja AdCM, de Oliveira LZ, Sbruzzi G. Inspiratory muscle training in patients with heart failure: what is new? systematic review and meta-analysis. *Phys Ther*. 2020;100(12):2099-2109. https://doi.org/10.1093/ptj/pzaa171
130. von Hippel PT. The heterogeneity statistic I2 can be biased in small meta-analyses. *BMC Med Res Methodol*. 2015;15(1):1-8. https://doi.org/10.1186/s12874-015-0024-z
131. Heron M. Deaths: Leading Causes for 2019. *Natl Vital Stat Rep*. Jul 2021;70(9):1-114.
132. Punekar YS, Riley JH, Lloyd E, Driessen M, Singh SJ. Systematic review of the association between exercise tests and patient-reported outcomes in patients with chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis*. 2017:2487-2506. https://doi.org/10.2147/COPD.S100204
133. de Souza Y, Suzana ME, Medeiros S, Macedo J, da Costa CH. Respiratory muscle weakness and its association with exercise capacity in patients with chronic obstructive pulmonary disease. *Clin Respir J*. Feb 2022;16(2):162-166. https://doi.org/10.1111/crj.13449
134. Tudorache V, Oancea C, Mlădinescu OF. Clinical relevance of maximal inspiratory pressure: determination in COPD exacerbation. *Int J Chron Obstruct Pulmon Dis*. 2010:119-123.
135. Silva RN, Goulart CdL, de Oliveira CR, et al. Respiratory muscle strength can improve the prognostic assessment in COPD. *Sci Rep*. 2024;14(1):12360. https://doi.org/10.1038/s41598-024-54264-w
136. Polkey MI, Spruit MA, Edwards LD, et al. Six-minute-walk test in chronic obstructive pulmonary disease: minimal clinically important difference for death or hospitalization. *Am J Respir Crit Care Med*. Feb 15 2013;187(4):382-6. https://doi.org/10.1164/rccm.201209-1596OC

137. Crook S, Busching G, Schultz K, et al. A multicentre validation of the 1-min sit-to-stand test in patients with COPD. *Eur Respir J*. Mar 2017;49(3)https://doi.org/10.1183/13993003.01871-2016
138. Figueiredo RI, Azambuja AM, Cureau FV, Sbruzzi G. Inspiratory muscle training in COPD. *Respir Care*. 2020;65(8):1189-1201. https://doi.org/10.4187/respcare.07098
139. Nguyen N-M, Vancraeynest D, Pouleur A-C, Beaumont M, Fait P, Reychler G. The impact of isolated inspiratory muscle training on exercise capacity, dyspnea, and quality of life in heart failure patients: a systematic review and meta-analysis. *Phys Ther Rev*. 2025:1-13. https://doi.org/10.1080/10833196.2025.2475643
140. Silveira BMF, Pereira HLA, Chaves G, Ho DGC, Parreira VF. Reliability and Validity of Maximal Respiratory Pressures. *Respir Care*. Jun 28 2024;69(7):881-890. https://doi.org/10.4187/respcare.10641
141. Larson JL, Covey MK, Vitalo CA, Alex CG, Patel M, Kim MJ. Maximal inspiratory pressure. Learning effect and test-retest reliability in patients with chronic obstructive pulmonary disease. *Chest*. Aug 1993;104(2):448-53. https://doi.org/10.1378/chest.104.2.448
142. Larson JL, Kim MJ. Reliability of maximal inspiratory pressure. *Nurs Res*. Sep-Oct 1987;36(5):317-9.
143. Lopez-de-Uralde-Villanueva I, Fabero-Garrido R, Rivera EAR, et al. New Protocol for Evaluating Maximum Inspiratory Pressure: Concurrent Validity and Test-Retest Reliability. *Phys Ther*. Sep 3 2024:pzae124. https://doi.org/10.1093/ptj/pzae124
144. Formiga MF, Roach KE, Vital I, et al. Reliability and validity of the test of incremental respiratory endurance measures of inspiratory muscle performance in COPD. *Int J Chron Obstruct Pulmon Dis*. 2018;13:1569-1576. https://doi.org/10.2147/COPD.S160512
145. World Medical A. World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. *Nurs Ethics*. Jan 2002;9(1):105-9. https://doi.org/10.1191/0969733002ne486xx
146. Agustí A, Celli BR, Criner GJ, et al. Global initiative for chronic obstructive lung disease 2023 report: GOLD executive summary. *Am J Respir Crit Care Med*. 2023;207(7):819-837. https://doi.org/10.1164/rccm.202301-0106PP
147. Stanojevic S, Kaminsky DA, Miller MR, et al. ERS/ATS technical standard on interpretive strategies for routine lung function tests. *Eur Respir J*. Jul 2022;60(1)https://doi.org/10.1183/13993003.01499-2021
148. Vaidya T, De Bisschop C, Beaumont M, et al. Is the 1-minute sit-to-stand test a good tool for the evaluation of the impact of pulmonary rehabilitation?

Determination of the minimal important difference in COPD. *Int J Chron Obstruct Pulmon Dis*. 2016:2609-2616. https://doi.org/10.2147/COPD.S115439
149. Laveneziana P, Albuquerque A, Aliverti A, et al. ERS statement on respiratory muscle testing at rest and during exercise. *Eur Respir J*. 2019;53(6)https://doi.org/10.1183/13993003.01214-2018
150. Bohannon RW, Crouch R. 1-Minute Sit-to-Stand Test: SYSTEMATIC REVIEW OF PROCEDURES, PERFORMANCE, AND CLINIMETRIC PROPERTIES. *J Cardiopulm Rehabil Prev*. Jan 2019;39(1):2-8. https://doi.org/10.1097/HCR.0000000000000336
151. Yazici B, Yolacan S. A comparison of various tests of normality. *Journal of statistical computation and simulation*. 2007;77(2):175-183. https://doi.org/10.1080/10629360600678310
152. Liljequist D, Elfving B, Skavberg Roaldsen K. Intraclass correlation - A discussion and demonstration of basic features. *PLoS One*. 2019;14(7):e0219854. https://doi.org/10.1371/journal.pone.0219854
153. Mouelhi Y, Jouve E, Castelli C, Gentile S. How is the minimal clinically important difference established in health-related quality of life instruments? Review of anchors and methods. *Health and quality of life outcomes*. 2020;18(1):136. https://doi.org/10.1186/s12955-020-01344-w
154. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Med J*. Sep 2012;24(3):69-71.
155. Martínez À, del Río FJ, Riu J, Rius FX. Detecting proportional and constant bias in method comparison studies by using linear regression with errors in both axes. *Chemometrics and Intelligent Laboratory Systems*. 1999;49(2):179-193. https://doi.org/10.1016/S0169-7439(99)00036-2
156. Amagliani RM, Peterella JK, Jung AP. Type of Encouragement Influences Peak Muscle Force in College-Age Women. *Int J Exerc Sci*. 2010;3(4):165-173.
157. Halperin I, Pyne DB, Martin DT. Threats to internal validity in exercise science: a review of overlooked confounding variables. *Int J Sports Physiol Perform*. Oct 2015;10(7):823-9. https://doi.org/10.1123/ijspp.2014-0566
158. Bohannon RW. Considerations and practical options for measuring muscle strength: a narrative review. *BioMed Research International*. 2019;2019(1):8194537. https://doi.org/10.1155/2019/8194537
159. Beaumont M, Couasnon C, Péran L, Berriet AC, Ber CL, Pichon R. Determination of the minimal important difference for inspiratory muscle strength in people with severe and very severe COPD. *Clin Rehabil*. 2023;37(11):1521-1532. https://doi.org/10.1177/02692155231174124
160. Miranda S, Machado A, Melro H, Oliveira A, Valente C, Marques ASPDD. Is the 1-minute sit-to-stand test related to respiratory muscle strength in patients

with COPD? : Eur Respiratory Soc; 2017. https://doi.org/10.1183/1393003.congress-2017.OA1772
161. Tudorache V, Oancea C, Mladinescu OF. Clinical relevance of maximal inspiratory pressure: determination in COPD exacerbation. *Int J Chron Obstruct Pulmon Dis*. May 6 2010;5:119-23. https://doi.org/10.2147/copd.s9194
162. Khalil M, Wagih K, Mahmoud O. Evaluation of maximum inspiratory and expiratory pressure in patients with chronic obstructive pulmonary disease. *Egyptian Journal of Chest Diseases and Tuberculosis*. 2014;63(2):329-335. https://doi.org/10.1016/j.ejcdt.2014.01.010
163. Schneeberger T, Gloeckl R, Jarosch I, Drechsel F, Koczulla AR, Kenn K. The minimal important difference for the 1-minute sit-to-stand test following pulmonary rehabilitation in patients with COPD–a prospective observational trial. Eur Respiratory Soc; 2018. https://doi.org/10.1183/13993003.congress-2018.PA1431
164. COSMIN. COSMIN Helps you Select the Most Suitable Outcome Measurement Instruments. *URL: www cosmin nl/(accessed 27 February 2020)*. 2020;
165. Jolley C, Moxham J. Respiratory muscles, chest wall, diaphragm, and other: Clinical implications. *Encyclopedia of Respiratory Medicine, Four-Volume Set*. Elsevier Inc.; 2006:632-643.
166. Habedank D, Meyer FJ, Hetzer R, Anker SD, Ewert R. Relation of respiratory muscle strength, cachexia and survival in severe chronic heart failure. *J Cachexia Sarcopenia Muscle*. Dec 2013;4(4):277-85. https://doi.org/10.1007/s13539-013-0109-7
167. Hodgev VA, Kostianev SS. Maximal inspiratory pressure predicts mortality in patients with chronic obstructive pulmonary disease in a five-year follow-up. *Folia Med (Plovdiv)*. 2006;48(3-4):36-41.
168. Schoser B, Fong E, Geberhiwot T, et al. Maximum inspiratory pressure as a clinically meaningful trial endpoint for neuromuscular diseases: a comprehensive review of the literature. *Orphanet J Rare Dis*. 2017;12(1):1-12. https://doi.org/10.1186/s13023-017-0598-0
169. Tzanis G, Vasileiadis I, Zervakis D, et al. Maximum inspiratory pressure, a surrogate parameter for the assessment of ICU-acquired weakness. *BMC Anesthesiol*. Jun 26 2011;11(1):14. https://doi.org/10.1186/1471-2253-11-14
170. Hamazaki N, Kamiya K, Matsuzawa R, et al. Prevalence and prognosis of respiratory muscle weakness in heart failure patients with preserved ejection fraction. *Respir Med*. Jan 2020;161:105834. https://doi.org/10.1016/j.rmed.2019.105834

171. Society ER, Society AT. ATS/ERS Statement on respiratory muscle testing. *Am J Respir Crit Care Med*. 2002;166(4):518-624. https://doi.org/10.1164/rccm.166.4.518
172. Meyer FJ, Lossnitzer D, Kristen AV, et al. Respiratory muscle dysfunction in idiopathic pulmonary arterial hypertension. *Eur Respir J*. Jan 2005;25(1):125-30. https://doi.org/10.1183/09031936.04.00095804
173. Meyer FJ, Borst MM, Zugck C, et al. Respiratory muscle dysfunction in congestive heart failure: clinical correlation and prognostic significance. *Circulation*. May 1 2001;103(17):2153-8. https://doi.org/10.1161/01.cir.103.17.2153
174. Meade M, Guyatt G, Cook D, et al. Predicting success in weaning from mechanical ventilation. *Chest*. Dec 2001;120(6 Suppl):400S-24S. https://doi.org/10.1378/chest.120.6_suppl.400s
175. Dai Y, Liu Z, Zhan H, et al. Reduced inspiratory muscle strength increases pneumonia in patients with acute myocardial infarction. *Ann Phys Rehabil Med*. Mar 2022;65(2):101511. https://doi.org/10.1016/j.rehab.2021.101511
176. Chatwin M, Ross E, Hart N, Nickol AH, Polkey MI, Simonds AK. Cough augmentation with mechanical insufflation/exsufflation in patients with neuromuscular weakness. *Eur Respir J*. Mar 2003;21(3):502-8. https://doi.org/10.1183/09031936.03.00048102
177. Hutcheson KA, Barrow MP, Warneke CL, et al. Cough strength and expiratory force in aspirating and nonaspirating postradiation head and neck cancer survivors. *Laryngoscope*. Jul 2018;128(7):1615-1621. https://doi.org/10.1002/lary.26986
178. Koo P, Oyieng'o DO, Gartman EJ, Sethi JM, Eaton CB, McCool FD. The maximal expiratory-to-inspiratory pressure ratio and supine vital capacity as screening tests for diaphragm dysfunction. *Lung*. 2017;195:29-35. https://doi.org/10.1007/s00408-016-9959-z
179. Nève V, Edmé J-L, Matran R. Earlier decline in sniff nasal inspiratory pressure than peak expiratory flow in children with Duchenne muscular dystrophy. *Eur Respir J*. 2014;44(5):1361-1363. https://doi.org/10.1183/09031936.00043214
180. Capozzo R, Quaranta VN, Pellegrini F, et al. Sniff nasal inspiratory pressure as a prognostic factor of tracheostomy or death in amyotrophic lateral sclerosis. *J Neurol*. 2015;262:593-603. https://doi.org/10.1007/s00415-014-7613-3
181. Weidner NJ. Developing an interdisciplinary palliative care plan for the patient with muscular dystrophy. *Pediatr Ann*. Jul 2005;34(7):546-52. https://doi.org/10.3928/0090-4481-20050701-12
182. Emerson RW. Convenience sampling, random sampling, and snowball sampling: How does sampling affect the validity of research? *Journal of visual*

impairment & blindness. 2015;109(2):164-168. https://doi.org/10.1177/0145482X1510900215
183. Bujang MA, Sa'at N, Bakar TMITA, Joo LC. Sample size guidelines for logistic regression from observational studies with large population: emphasis on the accuracy between statistics and parameters based on real life clinical data. *The Malaysian journal of medical sciences: MJMS*. 2018;25(4):122. https://doi.org/: https://doi.org/10.21315/mjms2018.25.4.12
184. Ponce MC, Sharma S. Pulmonary function tests. *StatPearls [internet]*. StatPearls Publishing; 2021.
185. Hounkpatin H, Stuart B, Zhu S, et al. Post-consultation acute respiratory tract infection recovery: a latent class-informed analysis of individual patient data. *Br J Gen Pract*. Mar 2023;73(728):e196-e203. https://doi.org/10.3399/BJGP.2022.0229
186. Degryse J, Buffels J, Van Dijck Y, Decramer M, Nemery B. Accuracy of office spirometry performed by trained primary-care physicians using the MIR Spirobank hand-held spirometer. *Respiration*. 2012;83(6):543-52. https://doi.org/10.1159/000334907
187. Graham BL, Steenbruggen I, Miller MR, et al. Standardization of Spirometry 2019 Update. An Official American Thoracic Society and European Respiratory Society Technical Statement. *Am J Respir Crit Care Med*. Oct 15 2019;200(8):e70-e88. https://doi.org/10.1164/rccm.201908-1590ST
188. Man WD, Kyroussis D, Fleming TA, et al. Cough gastric pressure and maximum expiratory mouth pressure in humans. *Am J Respir Crit Care Med*. Sep 15 2003;168(6):714-7. https://doi.org/10.1164/rccm.200303-334BC
189. Berger VW, Zhou Y. Kolmogorov–smirnov test: Overview. *Wiley statsref: Statistics reference online*. 2014;
190. Mordkoff JT. A simple method for removing bias from a popular measure of standardized effect size: Adjusted partial eta squared. *Advances in Methods and Practices in Psychological Science*. 2019;2(3):228-232. https://doi.org/10.1177/2515245919855053
191. Cohen J. Statistical power for the behavioural sciences. Hilsdale. *NY: Lawrence Erlbaum*. 1988;58(1):7-19.
192. Štrumbelj E, Kononenko I. Explaining prediction models and individual predictions with feature contributions. *Knowledge and information systems*. 2014;41:647-665.
193. Solutions S. Correlation. *Pearson, Kendall, Spearman) Retrieved April*. 2019;16:2018.
194. Enright PL, Kronmal RA, Manolio TA, Schenker MB, Hyatt RE. Respiratory muscle strength in the elderly. Correlates and reference values. Cardiovascular

Health Study Research Group. *Am J Respir Crit Care Med*. Feb 1994;149(2 Pt 1):430-8. https://doi.org/10.1164/ajrccm.149.2.8306041
195. Harik-Khan RI, Wise RA, Fozard JL. Determinants of maximal inspiratory pressure. The Baltimore Longitudinal Study of Aging. *Am J Respir Crit Care Med*. Nov 1998;158(5 Pt 1):1459-64. https://doi.org/10.1164/ajrccm.158.5.9712006
196. Chen H-I, Kuo C-S. Relationship between respiratory muscle function and age, sex, and other factors. *J Appl Physiol*. 1989;66(2):943-948. https://doi.org/10.1152/jappl.1989.66.2.943
197. Johan A, Chan C, Chia H, Chan O, Wang Y. Maximal respiratory pressures in adult Chinese, Malays and Indians. *Eur Respir J*. 1997;10(12):2825-2828. https://doi.org/10.1183/09031936.97.10122825
198. Evans JA, Whitelaw WA. The assessment of maximal respiratory mouth pressures in adults. *Respir Care*. Oct 2009;54(10):1348-59.
199. Gopalakrishna A, Vaishali K, Prem V, Aaron P. Normative values for maximal respiratory pressures in an Indian Mangalore population: A cross-sectional pilot study. *Lung India*. Oct 2011;28(4):247-52. https://doi.org/10.4103/0970-2113.85684
200. Uldry C, Fitting JW. Maximal values of sniff nasal inspiratory pressure in healthy subjects. *Thorax*. Apr 1995;50(4):371-5. https://doi.org/10.1136/thx.50.4.371
201. Araujo PR, Resqueti VR, Nascimento Junior J, et al. Reference values for sniff nasal inspiratory pressure in healthy subjects in Brazil: a multicenter study. *J Bras Pneumol*. Nov-Dec 2012;38(6):700-7. https://doi.org/10.1590/s1806-37132012000600004
202. Kamide N, Ogino M, Yamashina N, Fukuda M. Sniff nasal inspiratory pressure in healthy Japanese subjects: mean values and lower limits of normal. *Respiration*. 2009;77(1):58-62. https://doi.org/10.1159/000167321
203. Huang C-H, Yang G-G, Chen T-W. Sniff nasal inspiratory pressure does not decrease in elderly subjects. *Journal of Physical Therapy Science*. 2014;26(9):1509-1513. https://doi.org/10.1589/jpts.26.1509
204. Braun L, Wolfgang M, Dickersin K. Defining race/ethnicity and explaining difference in research studies on lung function. *Eur Respir J*. 2013;41(6):1362-1370. https://doi.org/10.1183/09031936.00091612
205. Diaz T, Strong KL, Cao B, et al. A call for standardised age-disaggregated health data. *Lancet Healthy Longev*. Jul 2021;2(7):e436-e443. https://doi.org/10.1016/S2666-7568(21)00115-X
206. Geifman N, Cohen R, Rubin E. Redefining meaningful age groups in the context of disease. *Age (Dordr)*. Dec 2013;35(6):2357-66. https://doi.org/10.1007/s11357-013-9510-6

207. Watsford ML, Murphy AJ, Pine MJ. The effects of ageing on respiratory muscle function and performance in older adults. *J Sci Med Sport*. Feb 2007;10(1):36-44. https://doi.org/10.1016/j.jsams.2006.05.002
208. Simoes RP, Deus AP, Auad MA, Dionisio J, Mazzonetto M, Borghi-Silva A. Maximal respiratory pressure in healthy 20 to 89 year-old sedentary individuals of central Sao Paulo State. *Rev Bras Fisioter*. Jan-Feb 2010;14(1):60-7. https://doi.org/10.1590/s1413-35552010000100010
209. Caruana EJ, Roman M, Hernandez-Sanchez J, Solli P. Longitudinal studies. *J Thorac Dis*. Nov 2015;7(11):E537-40. https://doi.org/10.3978/j.issn.2072-1439.2015.10.63
210. Goodpaster BH, Park SW, Harris TB, et al. The loss of skeletal muscle strength, mass, and quality in older adults: the health, aging and body composition study. *J Gerontol A Biol Sci Med Sci*. Oct 2006;61(10):1059-64. https://doi.org/10.1093/gerona/61.10.1059
211. Bordoni B, Morabito B, Simonelli M. Ageing of the diaphragm muscle. *Cureus*. 2020;12(1)https://doi.org/10.7759/cureus.6645
212. Quanjer PH, Tammeling G, Cotes J, Pedersen O, Peslin R, Yernault J. Lung volumes and forced ventilatory flows. *Eur Respir J*. 1993;6(Suppl 16):5-40.
213. Katagiri M, Abe T, Yokoba M, Dobashi Y, Tomita T, Easton PA. Neck and abdominal muscle activity during a sniff. *Respir Med*. Sep 2003;97(9):1027-35. https://doi.org/10.1016/s0954-6111(03)00133-1
214. Goligher EC, Dres M, Fan E, et al. Mechanical Ventilation-induced Diaphragm Atrophy Strongly Impacts Clinical Outcomes. *Am J Respir Crit Care Med*. Jan 15 2018;197(2):204-213. https://doi.org/10.1164/rccm.201703-0536OC
215. Abd Rahman NI, Md Dawal SZ, Yusoff N, Mohd Kamil NS. Anthropometric measurements among four Asian countries in designing sitting and standing workstations. *Sādhanā*. 2018;43:1-9. https://doi.org/10.1007/s12046-017-0768-8
216. Beaumont M, Forget P, Couturaud F, Reychler G. Effects of inspiratory muscle training in COPD patients: A systematic review and meta-analysis. *Clin Respir J*. Jul 2018;12(7):2178-2188. https://doi.org/10.1111/crj.12905
217. Elkins M, Dentice R. Inspiratory muscle training facilitates weaning from mechanical ventilation among patients in the intensive care unit: a systematic review. *J Physiother*. 2015;61(3):125-134.
218. Watson K, Egerton T, Sheers N, et al. Respiratory muscle training in neuromuscular disease: a systematic review and meta-analysis. *Eur Respir Rev*. Dec 31 2022;31(166)https://doi.org/10.1183/16000617.0065-2022
219. Lasevicius T, Ugrinowitsch C, Schoenfeld BJ, et al. Effects of different intensities of resistance training with equated volume load on muscle strength and hypertrophy. *Eur J Sport Sci*. Jul 2018;18(6):772-780. https://doi.org/10.1080/17461391.2018.1450898

220. Seixas MB, Almeida LB, Trevizan PF, et al. Effects of Inspiratory Muscle Training in Older Adults. *Respir Care*. Apr 2020;65(4):535-544. https://doi.org/10.4187/respcare.06945
221. Vázquez-Gandullo E, Hidalgo-Molina A, Montoro-Ballesteros F, Morales-González M, Muñoz-Ramírez I, Arnedillo-Muñoz A. Inspiratory muscle training in patients with chronic obstructive pulmonary disease (COPD) as part of a respiratory rehabilitation program implementation of mechanical devices: a systematic review. *Int J Environ Res Public Health*. 2022;19(9):5564.
222. McConnell A, Romer LM, Weiner P. Inspiratory muscle training in obstructive lung disease. *Breathe*. 2005;2(1):38-49. https://doi.org/10.1183/18106838.0201.38
223. Palermo P, Corra U. Exercise prescriptions for training and rehabilitation in patients with heart and lung disease. *Annals of the American Thoracic Society*. 2017;14(Supplement 1):S59-S66. https://doi.org/10.1513/AnnalsATS.201702-160FR
224. Ritchie C. Rating of perceived exertion (RPE). *J Physiother*. 2012;58(1):62. https://doi.org/10.1016/S1836-9553(12)70078-4
225. Buckley JP, Borg GA. Borg's scales in strength training; from theory to practice in young and older adults. *Appl Physiol Nutr Metab*. Oct 2011;36(5):682-92. https://doi.org/10.1139/h11-078
226. McConnell A. *Respiratory muscle training: theory and practice*. 1 ed. Elsevier Health Sciences; 2013.
227. Borg G. Psychophysical scaling with applications in physical work and the perception of exertion. *Scand J Work Environ Health*. 1990:55-58. https://doi.org/10.5271/sjweh.1815
228. Mukaka MM. A guide to appropriate use of correlation coefficient in medical research. *Malawi Med J*. 2012;24(3):69-71.
229. Mrkvicka T, Myllymaki M, Jilek M, Hahn U. A one-way ANOVA test for functional data with graphical interpretation. *arXiv preprint arXiv:161203608*. 2016;https://doi.org/10.48550/arXiv.1612.03608
230. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. Jun 1999;8(2):135-60. https://doi.org/10.1177/096228029900800204
231. Unal I. Defining an optimal cut-point value in ROC analysis: an alternative approach. *Comput Math Methods Med*. 2017;2017(1):3762651. https://doi.org/10.1155/2017/3762651
232. del Mar Seguí M, Cabrero-García J, Crespo A, Verdú J, Ronda E. A reliable and valid questionnaire was developed to measure computer vision syndrome at the workplace. *J Clin Epidemiol*. 2015;68(6):662-673. https://doi.org/10.1016/j.jclinepi.2015.01.015

233. Çorbacıoğlu ŞK, Aksel G. Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. *Turkish journal of emergency medicine*. 2023;23(4):195-198. https://doi.org/10.4103/tjem.tjem_182_23
234. Lea JW, O'Driscoll JM, Hulbert S, Scales J, Wiles JD. Convergent validity of ratings of perceived exertion during resistance exercise in healthy participants: a systematic review and meta-analysis. *Sports medicine-open*. 2022;8(1):2. https://doi.org/10.1186/s40798-021-00386-8
235. Pincivero DM, Coelho AJ, Campy RM. Perceived exertion and maximal quadriceps femoris muscle strength during dynamic knee extension exercise in young adult males and females. *Eur J Appl Physiol*. 2003;89:150-156. https://doi.org/10.1007/s00421-002-0768-0
236. Scherr J, Wolfarth B, Christle JW, Pressler A, Wagenpfeil S, Halle M. Associations between Borg's rating of perceived exertion and physiological measures of exercise intensity. *Eur J Appl Physiol*. 2013;113:147-155. https://doi.org/10.1007/s00421-012-2421-x
237. Aucoin R, Ekström M, Li PZ, Bourbeau J, Tan WC, Jensen D. Response Bias for Modified Borg 0–10 Numerical Categories without Verbal Descriptors during Assessment of Exertional Symptoms. *Am J Respir Crit Care Med*. 2025;211(2):284-287. https://doi.org/10.1164/rccm.202406-1206RL
238. Day ML, McGuigan MR, Brice G, Foster C. Monitoring exercise intensity during resistance training using the session RPE scale. *The Journal of Strength & Conditioning Research*. 2004;18(2):353-358. https://doi.org/10.1519/R-13113.1
239. Gupta A, Stead TS, Ganti L. Determining a meaningful R-squared value in clinical medicine. *Academic Medicine & Surgery*. 2024;https://doi.org/10.62186/001c.125154
240. Abbiss CR, Peiffer JJ, Meeusen R, Skorski S. Role of ratings of perceived exertion during self-paced exercise: what are we actually measuring? *Sports Med*. 2015;45:1235-1243. https://doi.org/10.1007/s40279-015-0344-5
241. Von Leupoldt A, Ambruzsova R, Nordmeyer S, Jeske N, Dahme B. Sensory and affective aspects of dyspnea contribute differentially to the Borg scale's measurement of dyspnea. *Respiration*. 2006;73(6):762-768. https://doi.org/10.1159/000095910
242. Ponikowski P, Anker SD, AlHabib KF, et al. Heart failure: preventing disease and death worldwide. *ESC Heart Fail*. Sep 2014;1(1):4-25. https://doi.org/10.1002/ehf2.12005
243. Brahmbhatt DH, Cowie MR. Heart failure: classification and pathophysiology. *Medicine*. 2018;46(10):587-593. https://doi.org/10.1016/j.mpmed.2018.07.004

244. Lesyuk W, Kriza C, Kolominsky-Rabas P. Cost-of-illness studies in heart failure: a systematic review 2004-2016. *BMC Cardiovasc Disord*. May 2 2018;18(1):74. https://doi.org/10.1186/s12872-018-0815-3
245. Georgiopoulou VV, Kalogeropoulos AP, Chowdhury R, et al. Exercise Capacity, Heart Failure Risk, and Mortality in Older Adults: The Health ABC Study. *Am J Prev Med*. Feb 2017;52(2):144-153. https://doi.org/10.1016/j.amepre.2016.08.041
246. Wikner A, Sandstrom A, Rinnstrom D, et al. Impaired Exercise Capacity and Mortality Risk in Adults With Congenital Heart Disease. *JACC Adv*. Jul 2023;2(5):100422. https://doi.org/10.1016/j.jacadv.2023.100422
247. Pollentier B, Irons SL, Benedetto CM, et al. Examination of the six minute walk test to determine functional capacity in people with chronic heart failure: a systematic review. *Cardiopulm Phys Ther J*. Mar 2010;21(1):13-21.
248. Zavin A, Daniels K, Allsup K, et al. The six minute walk distance as an indicator of frailty and functional incapacity in elderly heart failure patients. *J Am Coll Cardiol*. 2012;59(13S):E934-E934. https://doi.org/10.1016/S0735-1097(12)60935-8
249. Enright PL, McBurnie MA, Bittner V, et al. The 6-min walk test: a quick measure of functional status in elderly adults. *Chest*. Feb 2003;123(2):387-98. https://doi.org/10.1378/chest.123.2.387
250. Zhang X, Kang Y, Luo Z, et al. Feasibility and safety of 1-min sit-to-stand test in acute decompensated heart failure confirmed by lung ultrasound. *Front Cardiovasc Med*. 2023;10:1103247. https://doi.org/10.3389/fcvm.2023.1103247
251. Tanriverdi A, Kahraman BO, Ozpelit E, Savci S. Test-Retest Reliability and Validity of 1-Minute Sit-to-Stand Test in Patients With Chronic Heart Failure. *Heart Lung Circ*. Apr 2023;32(4):518-524. https://doi.org/10.1016/j.hlc.2023.01.008
252. Crook S, Büsching G, Schultz K, et al. A multicentre validation of the 1-min sit-to-stand test in patients with COPD. *Eur Respir J*. 2017;49(3)https://doi.org/10.1183/13993003.01871-2016
253. Radtke T, Puhan MA, Hebestreit H, Kriemler S. The 1-min sit-to-stand test—A simple functional capacity test in cystic fibrosis? *Journal of Cystic Fibrosis*. 2016;15(2):223-226. https://doi.org/10.1016/j.jcf.2015.08.006
254. Reychler G, Boucard E, Peran L, et al. One minute sit-to-stand test is an alternative to 6MWT to measure functional exercise performance in COPD patients. *The clinical respiratory journal*. 2018;12(3):1247-1256. https://doi.org/10.1111/crj.12658
255. McDonagh TA, Metra M, Adamo M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) With the special contribution of the Heart

Failure Association (HFA) of the ESC. *Eur Heart J*. 2021;42(36):3599-3726. https://doi.org/10.1093/eurheartj/ehab368
256. Laboratories ACoPSfCPF. ATS statement: guidelines for the six-minute walk test. *Am J Respir Crit Care Med*. 2002;166:111-117. https://doi.org/10.1164/ajrccm.166.1.at1102
257. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health and quality of life outcomes*. 2006;4:1-5. https://doi.org/10.1186/1477-7525-4-54
258. Kelley K, Preacher KJ. On effect size. *Psychol Methods*. Jun 2012;17(2):137-52. https://doi.org/10.1037/a0028086
259. Chen H, Cohen P, Chen S. How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics—simulation and Computation®*. 2010;39(4):860-864. https://doi.org/10.1080/03610911003650383
260. Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF. A comparison of two time intervals for test-retest reliability of health status instruments. *J Clin Epidemiol*. 2003;56(8):730-735. https://doi.org/10.1016/s0895-4356(03)00084-2
261. Hazra A, Gogtay N. Biostatistics Series Module 6: Correlation and Linear Regression. *Indian J Dermatol*. Nov-Dec 2016;61(6):593-601. https://doi.org/10.4103/0019-5154.193662
262. Ousmen A, Touraine C, Deliu N, et al. Distribution- and anchor-based methods to determine the minimally important difference on patient-reported outcome questionnaires in oncology: a structured review. *Health Qual Life Outcomes*. Dec 11 2018;16(1):228. https://doi.org/10.1186/s12955-018-1055-z
263. Sevillano-Castaño A, Peroy-Badal R, Torres-Castro R, et al. Is there a learning effect on 1-min sit-to-stand test in post-COVID-19 patients? *ERJ Open Research*. 2022;8(3)https://doi.org/10.1183/23120541.00189-2022
264. Uszko-Lencer N, Mesquita R, Janssen E, et al. Reliability, construct validity and determinants of 6-minute walk test performance in patients with chronic heart failure. *Int J Cardiol*. Aug 1 2017;240:285-290. https://doi.org/10.1016/j.ijcard.2017.02.109
265. McArdle WD, Katch FI, Katch VL. *Exercise physiology: nutrition, energy, and human performance*. Lippincott Williams & Wilkins; 2010.
266. Zweerink A, van der Lingen A-LC, Handoko ML, van Rossum AC, Allaart CP. Chronotropic incompetence in chronic heart failure: a state-of-the-art review. *Circ Heart Fail*. 2018;11(8):e004969. https://doi.org/10.1161/CIRCHEARTFAILURE.118.004969

267. Bohannon RW. Sit-to-stand test for measuring performance of lower extremity muscles. *Percept Mot Skills*. Feb 1995;80(1):163-6. https://doi.org/10.2466/pms.1995.80.1.163
268. Ueno K, Kamiya K, Hamazaki N, et al. Leg strength and incidence of heart failure in patients with acute coronary syndrome. *Eur J Prev Cardiol*. May 11 2024;31(7):834-842. https://doi.org/10.1093/eurjpc/zwad406
269. Hülsmann M, Quittan M, Berger R, et al. Muscle strength as a predictor of long-term survival in severe congestive heart failure. *Eur J Heart Fail*. 2004;6(1):101-107. https://doi.org/10.1016/j.ejheart.2003.07.008
270. Bekfani T, Bekhite Elsaied M, Derlien S, et al. Skeletal muscle function, structure, and metabolism in patients with heart failure with reduced ejection fraction and heart failure with preserved ejection fraction. *Circ Heart Fail*. 2020;13(12):e007198. https://doi.org/10.1161/CIRCHEARTFAILURE.120.007198
271. Olusanya BO, Kancherla V, Shaheen A, Ogbo FA, Davis AC. Global and regional prevalence of disabilities among children and adolescents: Analysis of findings from global health databases. *Frontiers in Public Health*. 2022;10:3276. https://doi.org/10.3389/fpubh.2022.977453
272. Lee AC, Kozuki N, Blencowe H, et al. Intrapartum-related neonatal encephalopathy incidence and impairment at regional and global levels for 2010 with trends from 1990. *Pediatr Res*. 2013;74(1):50-72. https://doi.org/10.1038/pr.2013.206
273. Blencowe H, Lee AC, Cousens S, et al. Preterm birth–associated neurodevelopmental impairment estimates at regional and global levels for 2010. *Pediatr Res*. 2013;74(1):17-34. https://doi.org/10.1038/pr.2013.204
274. WorldBank. World Bank Indicators. Under-five mortality-Vietnam. . World Bank. Retrieved 2022 https://data.worldbank.org/indicator/SH.DYN.MORT
275. Khandaker G, Van Bang N, Dung TQ, et al. Protocol for hospital based-surveillance of cerebral palsy (CP) in Hanoi using the Paediatric Active Enhanced Disease Surveillance mechanism (PAEDS-Vietnam): a study towards developing hospital-based disease surveillance in Vietnam. *BMJ Open*. Nov 9 2017;7(11):e017742. https://doi.org/10.1136/bmjopen-2017-017742
276. Shahat ARS, Greco G. The Economic Costs of Childhood Disability: A Literature Review. *Int J Environ Res Public Health*. Mar 29 2021;18(7):3531. https://doi.org/10.3390/ijerph18073531
277. Guralnick MJ. Why Early Intervention Works: A Systems Perspective. *Infants Young Child*. Jan 1 2011;24(1):6-28. https://doi.org/10.1097/IYC.0b013e3182002cfe
278. Bolduc ME, Du Plessis AJ, Sullivan N, et al. Spectrum of neurodevelopmental disabilities in children with cerebellar malformations. *Dev*

Med Child Neurol. May 2011;53(5):409-16. https://doi.org/10.1111/j.1469-8749.2011.03929.x
279. Missiuna C, Rivard L, Bartlett D. Exploring assessment tools and the target of intervention for children with Developmental Coordination Disorder. *Phys Occup Ther Pediatr*. 2006;26(1-2):71-89.
280. Hanna SE, Russell DJ, Bartlett DJ, Kertoy M, Rosenbaum PL, Wynn K. Measurement practices in pediatric rehabilitation: a survey of physical therapists, occupational therapists, and speech-language pathologists in Ontario. *Phys Occup Ther Pediatr*. 2007;27(2):25-42.
281. WHO WHO. *International Classification of Functioning, Disability, and Health: Children & Youth Version: ICF-CY*. World Health Organization; 2007.
282. Haley S, Coster W, Ludlow L, Haltiwanger J, Andrellos P. Pediatric evaluation of disability inventory. *Assessing children's well-being: a handbook of measures*. 2003;11:13.
283. Haley SM, Coster W, Ludlow LH, Haltiwanger JT, Andrellos PJ. *Pediatric evaluation of disability inventory (PEDI)*. Health and Disability Research Institute Boston University; 1992.
284. Ogonowski J, Kronk R, Rice C, Feldman H. Inter-rater reliability in assigning ICF codes to children with disabilities. *Disabil Rehabil*. Mar 18 2004;26(6):353-61. https://doi.org/10.1080/09638280410001658658
285. Schulze C, Page J, Lilja M, Kottorp A. Cross-cultural validity of the German version of the Pediatric Evaluation of Disability Inventory (PEDI-G)—A Rasch model application. *Child Care Health Dev*. 2017;43(1):48-58. https://doi.org/10.1111/cch.12401
286. Mancini MC, Coster WJ, Amaral MF, Avelar BS, Freitas R, Sampaio RF. New version of the Pediatric Evaluation of Disability Inventory (PEDI-CAT): translation, cultural adaptation to Brazil and analyses of psychometric properties. *Braz J Phys Ther*. Nov-Dec 2016;20(6):561-570. https://doi.org/10.1590/bjpt-rbf.2014.0166
287. Vos-Vromans DC, Ketelaar M, Gorter JW. Responsiveness of evaluative measures for children with cerebral palsy: the Gross Motor Function Measure and the Pediatric Evaluation of Disability Inventory. *Disabil Rehabil*. Oct 30 2005;27(20):1245-52. https://doi.org/10.1080/09638280500076178
288. Chen K-L, Tseng M-H, Hu F-C, Koh C-L. Pediatric Evaluation of Disability Inventory: A cross-cultural comparison of daily function between Taiwanese and American children. *Res Dev Disabil*. 2010;31(6):1590-1600. https://doi.org/10.1016/j.ridd.2010.05.002
289. Wassenberg-Severijnen JE, Custers JW, Hox JJ, Vermeer A, Helders PJ. Reliability of the Dutch Pediatric Evaluation of Disability Inventory (PEDI). *Clin Rehabil*. Jul 2003;17(4):457-62. https://doi.org/10.1191/0269215503cr634oa

290. World Medical A. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. Nov 27 2013;310(20):2191-4. https://doi.org/10.1001/jama.2013.281053
291. Lakens D. Sample size justification. *Collabra: psychology*. 2022;8(1):33267. https://doi.org/10.1525/collabra.33267
292. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*. Dec 15 2000;25(24):3186-91. https://doi.org/10.1097/00007632-200012150-00014
293. Zamanzadeh V, Ghahramanian A, Rassouli M, Abbaszadeh A, Alavi-Majd H, Nikanfar AR. Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication. *J Caring Sci*. Jun 2015;4(2):165-78. https://doi.org/10.15171/jcs.2015.017
294. Tsang S, Royse CF, Terkawi AS. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi J Anaesth*. May 2017;11(Suppl 1):S80-S89. https://doi.org/10.4103/sja.SJA_203_17
295. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163.
296. Giavarina D. Understanding Bland Altman analysis. *Biochem Med (Zagreb)*. 2015;25(2):141-51. https://doi.org/10.11613/BM.2015.015
297. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20(1):37-46.
298. Mokkink LB, Terwee CB, Patrick 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*. Jul 2010;63(7):737-45. https://doi.org/10.1016/j.jclinepi.2010.02.006
299. McCarthy E, Guerin S. Family-centred care in early intervention: A systematic review of the processes and outcomes of family-centred care and impacting factors. *Child Care Health Dev*. 2022;48(1):1-32. https://doi.org/10.1111/cch.12901
300. Erki˙ n G, Elhan AlH, Aybay C, Si˙ rzai˙ H, Ozel S. Validity and reliability of the Turkish translation of the Pediatric Evaluation of Disability Inventory (PEDI). *Disabil Rehabil*. 2007;29(16):1271-1279. https://doi.org/10.1080/09638280600964307
301. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*. 1986;327(8476):307-310.
302. Kaur P, Stoltzfus JC. Bland–Altman plot: a brief overview. *International Journal of Academic Medicine*. 2017;3(1):110-111. https://doi.org/10.4103/IJAM.IJAM_54_17

303. Haley S, Coster W, Dumas H, et al. Pediatric Evaluation of Disability Inventory–Computer Adaptive Test. *Cre Care*. 2012;
304. Adolph KE, Hoch JE. Motor development: Embodied, embedded, enculturated, and enabling. *Annu Rev Psychol*. 2019;70(1):141-164. https://doi.org/10.1146/annurev-psych-010418-102836
305. Keller H. Culture and development: A systematic relationship. *Perspect Psychol Sci*. 2017;12(5):833-840. https://doi.org/10.1177/1745691617704097
306. Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. *J Rehabil Med*. 2003;35(3):105-115.
307. Vidotto LS, Carvalho CRFd, Harvey A, Jones M. Dysfunctional breathing: what do we know? *Jornal Brasileiro de Pneumologia*. 2019;45:e20170347. https://doi.org/10.1590/1806-3713/e20170347
308. Chenivesse C, Similowski T, Bautin N, et al. Severely impaired health-related quality of life in chronic hyperventilation patients: exploratory data. *Respir Med*. Mar 2014;108(3):517-23. https://doi.org/10.1016/j.rmed.2013.10.024
309. Boulding R, Stacey R, Niven R, Fowler SJ. Dysfunctional breathing: a review of the literature and proposal for classification. *Eur Respir Rev*. Sep 2016;25(141):287-94. https://doi.org/10.1183/16000617.0088-2015
310. Courtney R, Greenwood KM, Cohen M. Relationships between measures of dysfunctional breathing in a population with concerns about their breathing. *J Bodyw Mov Ther*. Jan 2011;15(1):24-34. https://doi.org/10.1016/j.jbmt.2010.06.004
311. Thomas M, McKinley RK, Freeman E, Foy C, Price D. The prevalence of dysfunctional breathing in adults in the community with and without asthma. *Primary Care Respiratory Journal*. 2005;14(2):78-82.
312. Taverne J, Salvator H, Leboulch C, et al. High incidence of hyperventilation syndrome after COVID-19. *J Thorac Dis*. Jun 2021;13(6):3918-3922. https://doi.org/10.21037/jtd-20-2753
313. Hornsveld H, Garssen B. The low specificity of the Hyperventilation Provocation Test. *J Psychosom Res*. Nov 1996;41(5):435-49. https://doi.org/10.1016/s0022-3999(96)00195-x
314. Lewis BI. The hyperventilation syndrome. *Ann Intern Med*. May 1953;38(5):918-27. https://doi.org/10.7326/0003-4819-38-5-918
315. Tiotiu A, Ioan I, Poussel M, Schweitzer C, Kafi SA. Comparative analysis between available challenge tests in the hyperventilation syndrome. *Respir Med*. 2021;179:106329.
316. Van Dixhoorn J, Folgering H. The Nijmegen Questionnaire and dysfunctional breathing. Eur Respiratory Soc; 2015.
317. Looha MA, Masaebi F, Abedi M, Mohseni N, Fakharian A. The Optimal Cut-off Score of the Nijmegen Questionnaire for Diagnosing Hyperventilation

Syndrome Using a Bayesian Model in the Absence of a Gold Standard. *Galen Medical Journal*. 2020;9:e1738.
318. van Dixhoorn J, Duivenvoorden HJ. Efficacy of Nijmegen Questionnaire in recognition of the hyperventilation syndrome. *J Psychosom Res*. 1985;29(2):199-206. https://doi.org/10.1016/0022-3999(85)90042-x
319. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. *J Man Manip Ther*. 2009;17(3):163-70. https://doi.org/10.1179/jmt.2009.17.3.163
320. Lynn MR. Determination and quantification of content validity. *Nurs Res*. Nov-Dec 1986;35(6):382-5.
321. Shrestha N. Factor Analysis as a Tool for Survey Analysis. *American Journal of Applied Mathematics and Statistics*. 2021;9(1):4-11. https://doi.org/10.12691/ajams-9-1-2
322. Hu Lt, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*. 1999;6(1):1-55. https://doi.org/10.1080/10705519909540118
323. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ*. Jun 27 2011;2:53-55. https://doi.org/10.5116/ijme.4dfb.8dfd
324. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. Jun 2016;15(2):155-63. https://doi.org/10.1016/j.jcm.2016.02.012
325. Ogilvie VL, Kersten P. A critical review of the psychometric properties of the Nijmegen Questionnaire for hyperventilation syndrome. *New Zealand Journal of Physiotherapy*. 2015;43(1):3–10-3–10.
326. von Leupoldt A, Sommer T, Kegat S, et al. The unpleasantness of perceived dyspnea is processed in the anterior insula and amygdala. *Am J Respir Crit Care Med*. May 1 2008;177(9):1026-32. https://doi.org/10.1164/rccm.200712-1821OC
327. Ravanbakhsh M, Nargesi M, Raji H, Haddadzadeh Shoushtari M. Reliability and Validity of the Iranian Version of Nijmegen Questionnaire in Iranians with Asthma. *Tanaffos*. 2015;14(2):121-7. https://doi.org/PMC4629426
328. Grammatopoulou EP, Skordilis EK, Georgoudis G, et al. Hyperventilation in asthma: a validation study of the Nijmegen Questionnaire--NQ. *J Asthma*. Oct 2014;51(8):839-46. https://doi.org/10.3109/02770903.2014.922190
329. Pakdee S, Piya-Amornphan N. The Thai version of the Nijmegen questionnaire. *Heliyon*. Dec 2022;8(12):e12296. https://doi.org/10.1016/j.heliyon.2022.e12296
330. Boynton PM, Greenhalgh T. Selecting, designing, and developing your questionnaire. *BMJ*. 2004;328(7451):1312-1315. https://doi.org/10.1136/bmj.328.7451.1312

331. Pomerantz EM, Wang Q. The role of parental control in children's development in Western and East Asian countries. *Curr Dir Psychol Sci*. 2009;18(5):285-289. https://doi.org/10.1111/j.1467-8721.2009.01653.x
332. Thomas EJ, Petersen LA. Measuring errors and adverse events in health care. *J Gen Intern Med*. 2003;18:61-67. https://doi.org/10.1046/j.1525-1497.2003.20147.x
333. Zhang Q, Lu H, Pan S, Lin Y, Zhou K, Wang L. 6MWT performance and its correlations with VO2 and handgrip strength in home-dwelling mid-aged and older Chinese. *Int J Environ Res Public Health*. 2017;14(5):473. https://doi.org/10.3390/ijerph14050473
334. Allado E, Poussel M, Albuisson E, et al. Physical Activity Capacity Assessment of Patients With Chronic Disease and the 1-Minute Sit to Stand Test: Is There an Interest? *Frontiers in Sports and Active Living*. 2022;4:839509. https://doi.org/10.3389/fspor.2022.839509