Percentile-based *versus* conventional thresholds in cardiopulmonary exercise testing: a retrospective cross-sectional study

Michaël Staes [1,2], Marieke Wuyts [3], Maarten De Vos [4,5], Thierry Troosters [3] and Wim Janssens [1,2]

[1]Laboratory of Respiratory Diseases and Thoracic Surgery (BREATHE), Department of Chronic Diseases and Metabolism, KU Leuven, Leuven, Belgium. [2]Clinical Department of Respiratory Diseases, University Hospitals Leuven, Leuven, Belgium. [3]Department of Rehabilitation Sciences, KU Leuven, Leuven, Belgium. [4]Woman and Child, Department of Development and Regeneration, KU Leuven, Leuven, Belgium. [5]STADIUS, Department of Electrical Engineering (ESAT), KU Leuven, Leuven, Belgium.

Corresponding author: Wim Janssens (wim.janssens@uzleuven.be)

Check for updates

Shareable abstract (@ERSpublications)
Conventional *versus* percentile-based thresholds to classify CPET results yield substantial differences in abnormality rates, especially in cardiorespiratory disease, with a potential impact on interpretation and clinical decision-making https://bit.ly/46azKLb

Cite this article as: Staes M, Wuyts M, De Vos M, *et al*. Percentile-based *versus* conventional thresholds in cardiopulmonary exercise testing: a retrospective cross-sectional study. *ERJ Open Res* 2026; 12: 01354-2025 [DOI: 10.1183/23120541.01354-2025].

Copyright ©The authors 2026

This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0. For commercial reproduction rights and permissions contact permissions@ersnet.org

This article has an editorial commentary: https://doi.org/10.1183/23120541.00432-2026

Received: 1 Oct 2025
Accepted: 20 Jan 2026

Abstract

Introduction Determining whether a value falls within the expected range is a central step in cardiopulmonary exercise testing (CPET) interpretation. Conventional thresholds based on absolute or percentage-predicted values are widely used but fail to capture population variability. This study evaluated the validity of percentile-based thresholds derived from the Study of Health in Pomerania (SHIP) for a Western population and compared their impact on classification with conventional thresholds in individuals without known cardiorespiratory disease and patients with cardiorespiratory disease.

Methods We retrospectively analysed 886 maximal CPETs from 294 individuals without known cardiorespiratory disease (control group) and 592 patients with COPD, interstitial lung disease, heart failure or pulmonary hypertension. Percentile-based thresholds were evaluated for suitability by comparing abnormality rates with the expected 5% in the control group for peak oxygen uptake (V'_{O_2peak}), ventilatory equivalent for carbon dioxide (V'_E/V'_{CO_2} nadir values), peak heart rate (HR_{peak}), V'_{O_2}/work rate (WR) slope and peak V'_{O_2}/HR. Agreement, discordance, and cumulative impact of percentile-based *versus* conventional thresholds were assessed.

Results Percentile-based thresholds classified 4–8% of CPETs as abnormal in the control group, consistent with the expected 5%. In patients, conventional thresholds classified up to 14% more V'_{O_2peak} values and 31% more HR_{peak} values as abnormal compared with percentile-based thresholds, whereas percentile-based thresholds identified more abnormal V'_E/V'_{CO_2} nadir values. Full agreement across all variables was observed in 82% of CPETs in the control group *versus* 46–58% in patients.

Conclusion The SHIP-derived percentile-based thresholds are valid for Western populations, and their application leads to substantial differences in abnormality classification of CPET results compared to conventional thresholds, especially in patients.

Introduction

Incremental cardiopulmonary exercise testing (CPET) is a maximal stress test that assesses the body's response to exercise. CPET measures multiple variables that reflect the function of the cardiovascular, respiratory and metabolic systems [1]. It serves as an important tool for evaluating exercise capacity and identifying the mechanisms underlying exercise limitation or exertional symptoms [2].

A central part of CPET interpretation is determining whether the obtained values fall within the expected physiological range [3]. Yet, due to the lack of standardisation, the definition of normality remains

Downloaded from https://publications.ersnet.org on September 29, 2026 at KU Leuven University Library. Please see licensing information on first page for reuse rights.

variable [4]. This variability can alter CPET interpretation and subsequent clinical management, as misclassification may lead to underdetection of physiological impairment and false reassurance on the one hand, or to overdiagnosis and unnecessary additional investigations on the other hand [1, 5, 6].

Current interpretative strategies typically rely on thresholds based on absolute values or fixed percentages of predicted values [1–3, 7, 8]. While percentages of predicted values partially adjust for characteristics such as age, sex and body size, they reflect only deviation from the mean and do not account for the full statistical variability within a population [9].

A more robust statistical approach to defining normality relies on lower and upper limits of normal (LLN and ULN), derived from the distribution of values observed in a healthy reference population. By convention, these limits correspond to the 5th and 95th percentiles, which represent the boundaries beyond which only 5% of healthy individuals would be expected to fall on either side. This is now standard in interpretation of pulmonary function tests (PFTs), as recommended by the American Thoracic Society and the European Respiratory Society [6], but its uptake in CPET has been limited. This is mainly because, unlike the Global Lung Function Initiative (GLI) for PFTs, most available CPET reference datasets provide only predicted mean values rather than full percentile distributions [10–12].

However, a notable exception is the Study of Health in Pomerania (SHIP), which published reference equations for several CPET variables, based on a large reference cohort of German adults. These equations include the 5th and 95th percentiles, which can be used as LLN and ULN. This way, they provide more individualised cut-offs for abnormality by adjusting for age, sex, height, weight, and even β-blocker use and smoking status, while also capturing the natural variability within the population [13–16]. They have been endorsed for use in Western populations by the Association for Respiratory Technology and Physiology (ARTP) in 2021 and the textbook “Wasserman & Whipp’s Principles of Exercise Testing and Interpretation”, but are not yet frequently used in clinical practice [17, 18].

The aim of this study is twofold: first, to validate SHIP-derived percentile thresholds for CPET in a Western population, and second, to examine how their application, compared with conventional thresholds, influences abnormality classification in individuals without known cardiorespiratory disease and patients with chronic cardiorespiratory disease. In addition, we assessed the cumulative impact of applying different classification methods across multiple variables.

Materials and methods

Study design and population

CPETs performed at the University Hospitals Leuven (Belgium) between 2010 and 2022 were retrospectively collected. Eligible were Caucasian individuals aged 40 to 80 years. The control group consisted of individuals without known cardiorespiratory disease, although undetected subclinical cardiorespiratory disease and comorbidities may have been present. The patients had a clinical diagnosis of COPD, fibrotic interstitial lung disease (ILD), heart failure (HF) (with preserved or reduced ejection fraction) or group 1 or group 4 pulmonary hypertension (PH). Patients with overlapping cardiorespiratory comorbidities were ruled out, just as submaximal CPETs or CPETs with an incremental phase <3 min. Submaximal tests were defined as those meeting none of the following criteria (adapted from Radtke *et al.* [7] and Kersten *et al.* [19]): peak respiratory exchange ratio >1.05, presence of a plateau in oxygen uptake (V'_{O_2}), ventilatory limitation (breathing reserve <20%) and post-exercise lactate $\geqslant 5$ mmol·L^{-1}, as well as those terminated by the test operator due to safety concerns (ECG changes or concerning symptoms). For individuals with multiple tests, only the first maximal test was included. The full selection criteria can be found in supplementary table S1. Ethical approval was obtained from the ethical committee of the University Hospitals Leuven (s67714).

Collected characteristics included age, sex, height, weight, body mass index, smoking status (current, former or never), pack-years, diagnosis and β-blocker use (binary yes or no).

CPET

Prior to each CPET, spirometry was conducted to assess forced expiratory volume in 1 s (FEV_1) and forced vital capacity (FVC). All CPETs were performed on an electronically braked cycle ergometer according to the European Respiratory Society (ERS) guidelines [7]. The cycle ergometer was calibrated annually, while flow/volume and gas analysers were calibrated daily. Following a resting phase and unloaded pedaling, individuals performed an individualised incremental protocol (starting at 20 Watts+10 or 20 Watts increase per minute) until volitional exhaustion.

Downloaded from https://publications.ersnet.org on September 29, 2026 at KU Leuven University Library. Please see licensing information on first page for reuse rights.

CPET data was measured breath-by-breath by the Vyntus or Oxycon Pro CPX metabolic cart and averaged over 10 s. Continuous ECG was monitored. Collected parameters included work rate, V'_{O_2}, carbon dioxide output (V'_{CO_2}), minute ventilation (V'_E), peripheral oxygen saturation (S_{pO_2}), breathing frequency, partial pressure of end-tidal CO2 (P_{ETCO_2}), partial pressure of end-tidal O2 (P_{ETO_2}) and heart rate (HR). Derived variables such as the ventilatory equivalent for CO2 (V'_E/V'_{CO_2}) and oxygen pulse (V'_{O_2}/HR) were calculated. When performed, capillary lactate was sampled from the fingertip 2 min after exercise. Modified Borg scores (from 0 to 10) for dyspnoea and leg fatigue at peak exercise were obtained after termination of the test.

Analysed CPET variables

We analysed five CPET variables representing distinct physiological domains: peak oxygen uptake (V'_{O_2peak}), reflecting overall aerobic exercise capacity; V'_E/V'_{CO_2} nadir, often used in the evaluation of pulmonary gas exchange; and peak heart rate (HR_{peak}), utilised to assess the cardiac response to exercise. Also two markers of oxygen delivery and/or utilisation, peak oxygen pulse (V'_{O_2}/HR) and the V'_{O_2}–work rate slope (V'_{O_2}/WR), were included. For each, abnormality was defined using conventional *versus* percentile-based thresholds (table 1).

Conventional thresholds were based on commonly used cut-offs in the literature [3, 4, 7, 17], with predicted values (50th percentile) calculated using SHIP reference equations. Percentile-based thresholds (LLN or ULN) were also derived from SHIP, but defined as the 5th or 95th percentile depending on the variable. For participants not taking β-blockers, the initial SHIP reference equations were applied, which account for age, height, weight and sex [13, 16]. For those taking β-blockers, adapted SHIP equations were used, incorporating additional adjustments for β-blocker use, smoking status and in some models FEV_1 [14, 15]. Because no SHIP reference equation for the 50th percentile of peak V'_{O_2}/HR in individuals taking β-blockers is available, these participants were excluded from the analysis of this variable.

Statistical analysis

All data were analysed using Python (version 3.11.5). Descriptive statistics were used to summarise the different groups of the cohort.

The proportion of abnormal values was calculated separately using conventional and percentile-based thresholds. To evaluate the suitability of the SHIP reference values in the study cohort, a two-sided binomial test was used to assess whether the abnormality rate deviated significantly from the expected 5% in the control group of the present study. The same test was performed using the conventional thresholds.

Agreement between classification methods was evaluated using Cohen's kappa statistic. To assess whether the proportion of abnormal classifications differed significantly between methods, McNemar's test was applied. The clinical impact of the two interpretive approaches was further examined by calculating discordance rate, defined as the proportion of CPETs in which the two classification methods yielded different results. 95% confidence intervals (95% CI) were calculated using the Wilson method, to quantify the precision of discordance estimates. To investigate the role of age, discordance rates were also stratified by 10-year age groups and a chi square test was performed.

The number of discordant test variables per individual was analysed across diagnostic groups, based on disagreement between conventional and percentile-based thresholds. A two-sided p-value <0.05 was considered statistically significant, and the Benjamini–Hochberg procedure was used for all analyses to account for multiple testing [20].

TABLE 1 Conventional *versus* percentile-based thresholds for five cardiopulmonary exercise test variables analysed

	Conventional threshold	Percentile-based threshold
Peak V'_{O_2}	<85% predicted	<5th percentile (LLN)
V'_E/V'_{CO_2} nadir	>34	>95th percentile (ULN)
Peak HR	<90% predicted	<5th percentile (LLN)
Peak V'_{O_2}/HR	<80% predicted	<5th percentile (LLN)
V'_{O_2}/WR	<8.4 $mL \cdot min^{-1} \cdot W^{-1}$	<5th percentile (LLN)

V'_{O_2}: oxygen uptake; LLN: lower limit of normal; V'_E: minute ventilation; ULN: upper limit of normal; V'_{CO_2}: carbon dioxide production; HR: heart rate; WR: work rate.

Downloaded from https://publications.ersnet.org on September 29, 2026 at KU Leuven University Library. Please see licensing information on first page for reuse rights.

TABLE 2 Baseline and cardiopulmonary exercise test (CPET) characteristics of the study population according to diagnostic group

	Control group	COPD	ILD	HF	PH
CPETs, n	294	203	163	123	103
Disease classification, n	NA	GOLD I: 7 GOLD II: 69 GOLD III: 106 GOLD IV: 21	IPF: 135 Non-IPF: 28	HFrEF: 84 HFmrEF: 18 HFpEF: 21	Group 1: 41 Group 4: 62
Baseline characteristics					
Age, years	56 (50–62.75)	62 (57–68)	66 (59–71.5)	60 (54–66.5)	61 (55–70)
Males, %	57	65	74	70	44
FEV_1 % predicted	105 (94–114)	45 (34–54.5)	83 (70.5–96)	93.5 (82.75–106.25)	83 (73.5–95.5)
BMI, $kg \cdot m^{-2}$	25.75 (23.14–28.34)	24.52 (21.19–29.18)	27.64 (25.2–30.86)	26.49 (23.62–29.26)	27.24 (23.89–29.76)
Ever-smoker, %	26	99	62	52	50
Pack-years	0 (0–1)	40 (30–49.5)	5.5 (0–20)	1 (0–20)	0.5 (0–17)
β-blocker use, %	13	22	25	93	30
Atrial fibrillation, %	0	1	0	7	1
Pacemaker, %	0	1	1	7	0
CPET characteristics					
Incremental phase duration, mm:ss	9:44 (8:01–11:48)	5:23 (4:16–6:34)	7:02 (5:32–8:32)	7:17 (5:58–8:48)	5:25 (4:11–6:38)
Peak resistance, W	185 (141.5–240)	70 (60–80)	100 (80–140)	100 (78.5–130)	80 (60–95)
V'_{O_2peak} % predicted	110.5 (96–129.75)	63 (54–71)	78 (68–90)	65 (53.5–76.5)	61 (52.5–69)
Peak RER	1.19 (1.15–1.26)	1.01 (0.94–1.11)	1.12 (1.07–1.18)	1.18 (1.12–1.24)	1.14 (1.07–1.22)
Peak ventilation, $L \cdot min^{-1}$	87 (68.25–106)	46 (37–55.5)	71 (57.5–84)	59 (46.5–73)	66 (54.5–79.5)
HR_{peak} bpm	160.5 (149–170)	123 (111.5–138)	134 (120–150)	113 (97–136.5)	132 (120–145.5)
V'_E/V'_{CO_2} nadir	26.25 (24.21–29.18)	36.8 (32.06–40.59)	36.3 (32.2–41.31)	32.19 (28.91–35.82)	44.78 (39.64–51.1)
V'_E/V'_{CO_2} slope	25.22 (22.76–27.94)	32.25 (27.63–38.81)	35.6 (30.92–42.71)	30.94 (26.72–37.09)	48.27 (39.71–60.33)

Data are presented as median (interquartile range), unless otherwise indicated. ILD: interstitial lung disease; HF: heart failure; PH: pulmonary hypertension (Group 1: pulmonary arterial hypertension; Group 4: chronic thromboembolic pulmonary hypertension); GOLD: Global Initiative for Chronic Obstructive Lung Disease; IPF: idiopathic pulmonary fibrosis; HFrEF: HF with reduced ejection fraction; HFmrEF: HF with mildly reduced ejection fraction; HFpEF: HF with preserved ejection fraction; FEV_1: forced expiratory volume in 1 s; BMI: body mass index; V'_{O_2peak}: peak oxygen uptake; RER: respiratory exchange ratio; HR_{peak}: peak heart rate; V'_E/V'_{CO_2}: ventilatory equivalent for carbon dioxide.

Results

Subject selection and characteristics

1281 eligible CPETs were identified, corresponding to 1073 unique individuals. After excluding submaximal (n=119) and then too short CPETs (n=68), 886 CPETs were included.

The cohort comprised 294 controls without known cardiorespiratory disease, 203 with COPD, 163 with fibrotic ILD, 123 with HF and 103 with PH. Median age ranged from 56 to 66 years across groups, and all except the PH group were predominantly male. All baseline and CPET characteristics of the included groups can be seen in table 2. Owing to the high prevalence of β-blocker use in HF, V'_{O_2}/HR analyses were not performed in this group.

Abnormality classification in the absence of cardiorespiratory disease

In the control group, percentile-based thresholds classified between 4% and 8% of tests as abnormal across CPET variables, aligning with the expected 5% (table 3). When stratified by 10-year age groups, abnormality rates remained statistically consistent with 5%, supporting the validity of the SHIP reference values in this population (supplementary table S2). In contrast, conventional thresholds resulted in higher abnormality rates for V'_{O_2peak}, HR_{peak} and V'_{O_2}/WR slope (~10% each; all $p \leqslant 0.01$).

Despite these differences, absolute discordance between the classification methods was modest (2–6%) (figure 1). Stratification by age revealed no significant variation (all Benjamini–Hochberg-adjusted $p>0.55$, supplementary table S3).

Abnormality classification in cardiorespiratory disease

The full results for all CPET variables using both classification methods, stratified by diagnostic group, are shown in figure 1.

Downloaded from https://publications.ersnet.org on September 29, 2026 at KU Leuven University Library. Please see licensing information on first page for reuse rights.

TABLE 3 Abnormality rates for different cardiopulmonary exercise test variables in the control group (individuals without known cardiorespiratory disease) using conventional thresholds and percentile-based thresholds

Variable	Subjects, n	Conventional thresholds		Percentile-based thresholds	
		Abnormal, %	BH-corrected p-value (*versus* 5%)	Abnormal, %	BH-corrected p-value (*versus* 5%)
Peak V'_{O_2}	294	10.5	<0.01	6.5	0.38
V'_E/V'_{CO_2} nadir	294	2.4	0.054	7.8	0.16
Peak HR	294	10.5	<0.01	4.1	0.59
V'_{O_2}/WR slope	294	9.5	0.01	6.8	0.38
Peak V'_{O_2}/HR	257	3.9	0.53	3.9	0.59

Adjusted p-values indicate deviation from the expected 5%, based on a two-sided binomial test. V'_{O_2}: oxygen uptake; V'_E/V'_{CO_2}: ventilatory equivalent for carbon dioxide; HR: heart rate; WR: work rate; BH: Benjamini–Hochberg.

Across all disease groups, V'_{O_2peak} showed a pattern where the conventional threshold classified significantly more tests as abnormal (7–14%) than the percentile-based threshold. The percentile-based threshold did not capture additional abnormality. Agreement between the two methods was moderate to substantial (κ 0.56–0.76).

For V'_E/V'_{CO_2} nadir, the opposite pattern was observed: the percentile-based threshold (>95th percentile) classified more abnormality (6–19%) than the conventional threshold of >34. The highest agreement was observed in ILD (κ=0.85), although 6% of tests were still uniquely flagged by the percentile threshold (p=0.015).

The conventional threshold of <90% predicted for HR_{peak} classified a substantially higher proportion of tests as abnormal, from an additional 20% in ILD to 31% in HF (all p-values=0.00). Kappa values ranged from 0.44 to 0.54, indicating only moderate agreement between the two methods.

Discordance between classification methods was modest across groups for V'_{O_2}/WR slope abnormality detection, though the conventional threshold (<8.4 $mL \cdot min^{-1} \cdot W^{-1}$) consistently classified more tests as abnormal. In the disease groups, the proportion of abnormal V'_{O_2}/WR slope values progressively increased across the COPD, ILD, HF and PH groups. Similarly, the additional percentage of CPETs classified as abnormal by the conventional threshold also rose in this order, with 3%, 7%, 7% and 11% additional cases identified, respectively.

For V'_{O_2}/HR peak, the conventional threshold classified 59% of COPD and 37% of ILD patients as abnormal, with 8–11% uniquely abnormal by this method. In PH, 61% of CPETs were abnormal by both thresholds, while 7% were flagged only by the conventional method and 4% only by the percentile-based method.

For all variables, differences in discordance between 10-year age groups were not significantly different (supplementary table S3).

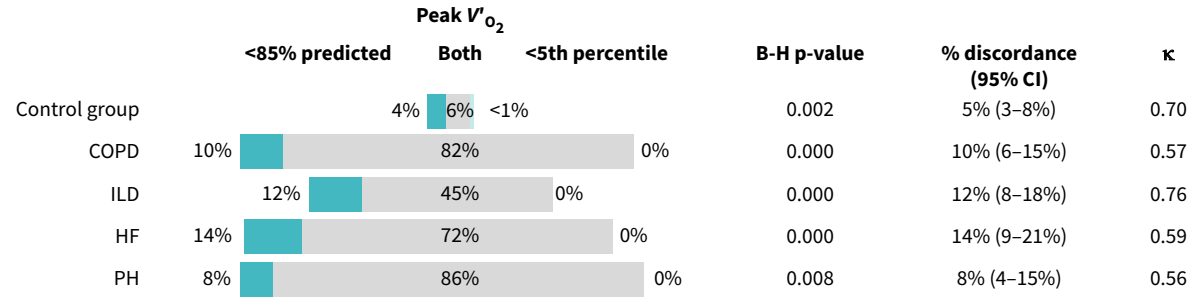
Cumulative effect of different classification methods

The cumulative effect of discordance across variables is shown in figure 2. In the control group, 82% of tests were concordant across all five variables, with 4% discordant in two or more variables. Full concordance was lower in disease groups (46–58%), with 7–14% of patients showing discordance in at least two variables.

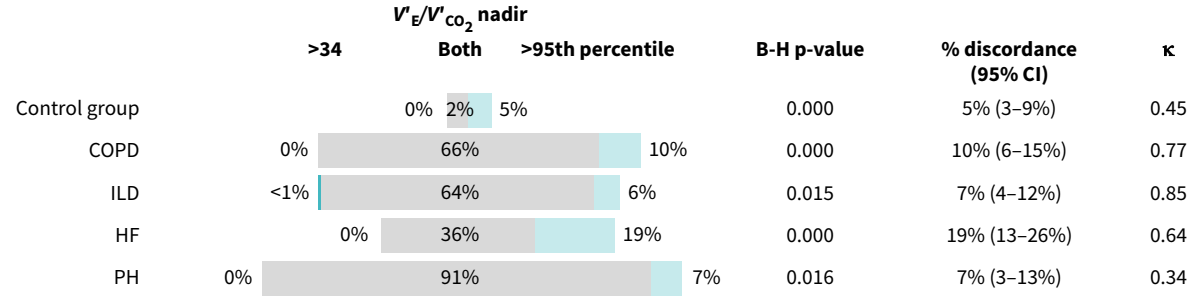
Discussion

Interpretation of the results of CPET requires knowledge of the normal responses, but wide variation exists in normality thresholds applied in clinical practice [17]. In this study, we examined the impact of using percentile-based thresholds instead of commonly used absolute or percentage-predicted cut-offs across five key CPET variables. In the control group of individuals without known cardiorespiratory disease, the SHIP-derived percentile-based thresholds yielded abnormality rates close to the expected 5%, even when stratified by age, which supports their validity for use in Western populations. In this control group, discordance between the two thresholding methods remained below 7% for each variable, while differences in normality classification between the two methods were substantially larger in patients with diagnosed cardiorespiratory disease.

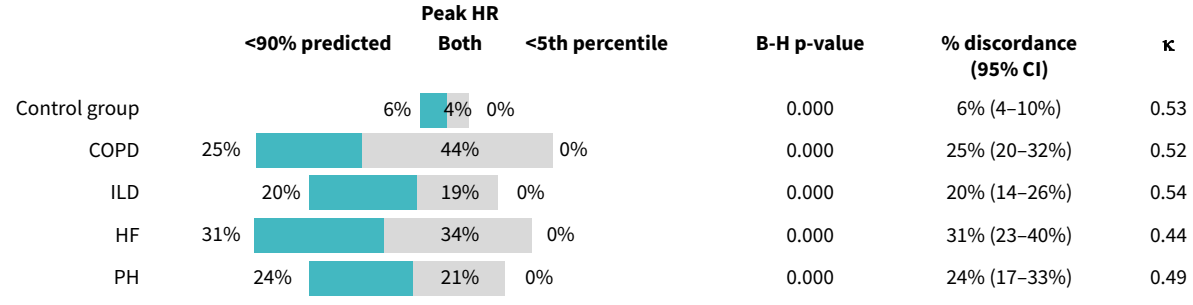
Downloaded from https://publications.ersnet.org on September 29, 2026 at KU Leuven University Library. Please see licensing information on first page for reuse rights.



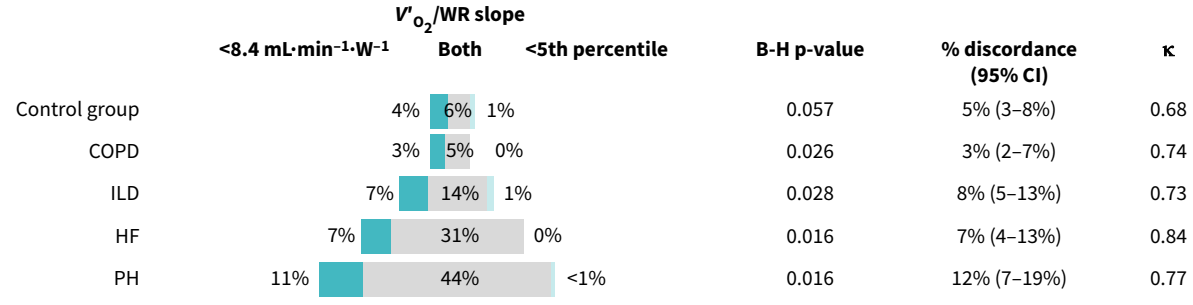
	B-H p-value	% discordance (95% CI)	κ
Control group	0.002	5% (3–8%)	0.70
COPD	0.000	10% (6–15%)	0.57
ILD	0.000	12% (8–18%)	0.76
HF	0.000	14% (9–21%)	0.59
PH	0.008	8% (4–15%)	0.56



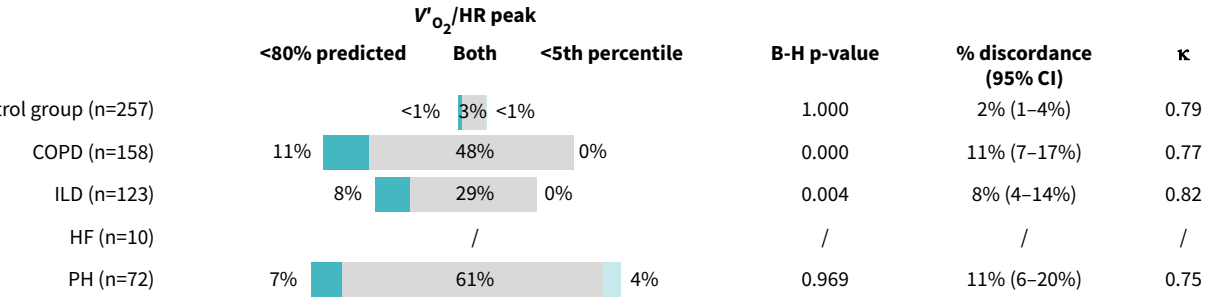
	B-H p-value	% discordance (95% CI)	κ
Control group	0.000	5% (3–9%)	0.45
COPD	0.000	10% (6–15%)	0.77
ILD	0.015	7% (4–12%)	0.85
HF	0.000	19% (13–26%)	0.64
PH	0.016	7% (3–13%)	0.34



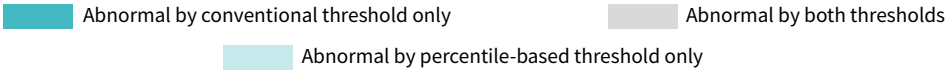
	B-H p-value	% discordance (95% CI)	κ
Control group	0.000	6% (4–10%)	0.53
COPD	0.000	25% (20–32%)	0.52
ILD	0.000	20% (14–26%)	0.54
HF	0.000	31% (23–40%)	0.44
PH	0.000	24% (17–33%)	0.49



	B-H p-value	% discordance (95% CI)	κ
Control group	0.057	5% (3–8%)	0.68
COPD	0.026	3% (2–7%)	0.74
ILD	0.028	8% (5–13%)	0.73
HF	0.016	7% (4–13%)	0.84
PH	0.016	12% (7–19%)	0.77



	B-H p-value	% discordance (95% CI)	κ
Control group (n=257)	1.000	2% (1–4%)	0.79
COPD (n=158)	0.000	11% (7–17%)	0.77
ILD (n=123)	0.004	8% (4–14%)	0.82
HF (n=10)	/	/	/
PH (n=72)	0.969	11% (6–20%)	0.75



Downloaded from https://publications.ersnet.org on September 29, 2026 at KU Leuven University Library. Please see licensing information on first page for reuse rights.

FIGURE 1 Comparison of conventional and percentile thresholds for five key cardiopulmonary exercise test variables. Each horizontal bar shows the proportion of patients classified as abnormal by the conventional threshold only, by both thresholds or by the percentile threshold only. For each diagnostic group, the Benjamini–Hochberg-adjusted p-value (B-H p-value) from McNemar's test, the percentage of discordant classifications and Cohen's κ are shown. Patients on β-blockers were excluded from the V'_{O_2}/HR peak analysis. No values are shown for V'_{O_2}/HR in heart failure due to lack of predicted values. V'_{O_2}: oxygen uptake; HR: heart rate; WR: work rate; V'_E: minute ventilation; V'_{CO_2}: carbon dioxide production; ILD: interstitial lung disease; HF: heart failure; PH: pulmonary hypertension.

To date, conventional thresholds remain the dominant interpretive approach and are widely incorporated into international CPET recommendations [3, 7]. Several commonly applied cut-offs to define normality originate from prognostic analyses. An example is the V'_E/V'_{CO_2} nadir or slope >34, which was mainly derived in cohorts with HF, COPD and PH, where values in the mid-30s demonstrated the highest discriminative ability to predict survival [21–23]. However, evidence indicates that the risk associated with increased V'_E/V'_{CO_2} rises progressively, with values below 34 already associated with worse outcomes in individuals with and without chronic cardiorespiratory disease [22, 24, 25]. Moreover, applying prognostic thresholds to other clinical contexts, where CPET is used to identify exercise-limiting factors, may be inappropriate, as supporting evidence in these settings is lacking.

Potentially even more problematic is that many thresholds used to define abnormality are based primarily on expert opinion rather than on distribution-based normative data. In this study, we analysed five frequently used conventional thresholds, but a wide range of alternative cut-offs exists in the literature, contributing to the existing heterogeneity in CPET interpretation [4, 26]. Percentile-based thresholds, in contrast, account for population variability and provide a statistically grounded method, thereby reducing interpretive variability. For this reason, Blanchard *et al.* [27] already advocated against the use of arbitrary thresholds, as they did not align with percentile-based definitions of normality in their paediatric cohort.

Although no studies have directly evaluated this issue, misclassification due to inappropriate normality thresholds may have important clinical consequences, as CPET results frequently guide further medical management. In our cohort, applying the conventional V'_E/V'_{CO_2} threshold underestimated the prevalence of abnormal ventilatory inefficiency compared with percentile-based limits, potentially overlooking subtle gas-exchange abnormalities or early pulmonary vascular disease [28]. Conversely, the conventional HR_{peak} threshold identified substantially more abnormal responses than percentile-based thresholds, which could lead to misinterpretation of cardiac limitation, the incorrect labelling of chronotropic incompetence and even unnecessary cardiologic investigations or medication adjustments. Importantly, because CPET interpretation is integrative and pattern-based rather than reliant on any single parameter, the effect of threshold choice across multiple variables is cumulative. In our control group, nearly one in five tests contained at least one discordant variable out of five, and in disease groups the full concordance rate across all variables dropped to <60%.

Nevertheless, a major barrier to wider use of percentile-based thresholds remains the limited availability of robust reference equations. As highlighted by Takken *et al.* [29], no existing CPET reference sets are entirely free from methodological flaws and many studies report only mean or median values, preventing the derivation of LLN and ULN. While the SHIP equations are a notable exception that offer percentile-based thresholds for several CPET variables, they also have shortcomings. Their variable coverage is restricted, as percentile-based reference values are lacking for several key CPET parameters, including breathing reserve. Moreover, the sample size is modest compared with large-scale initiatives such as the GLI, which limits the statistical robustness of the derived percentiles. Finally, the SHIP cohort consisted predominantly of Caucasian individuals from Northern Germany, which reduces the generalisability of these reference values to populations with different ethnic backgrounds. As a result, until more comprehensive population-based reference sets are developed, conventional thresholds will remain necessary in clinical practice, at least for some CPET variables. These findings underscore the need for international efforts, similar to the GLI, to establish robust, population-based reference values for CPET. Attempting to fit this research gap, the ERS founded a Global Lung Initiative task force to develop reference equations for CPET, including Z-scores [30].

The availability of Z-scores would also address the inherently binary nature of current interpretative methods. By quantifying the magnitude of deviation from expected values, Z-scores allow more nuanced interpretation than simple dichotomous thresholds [8]. However, their application relies on the assumption of approximately normal distributions, which may not hold for all CPET variables, such as HR_{peak} in older adults [31]. While Z-scores are preferable for monitoring longitudinal change, having access to full

Downloaded from https://publications.ersnet.org on September 29, 2026 at KU Leuven University Library. Please see licensing information on first page for reuse rights.

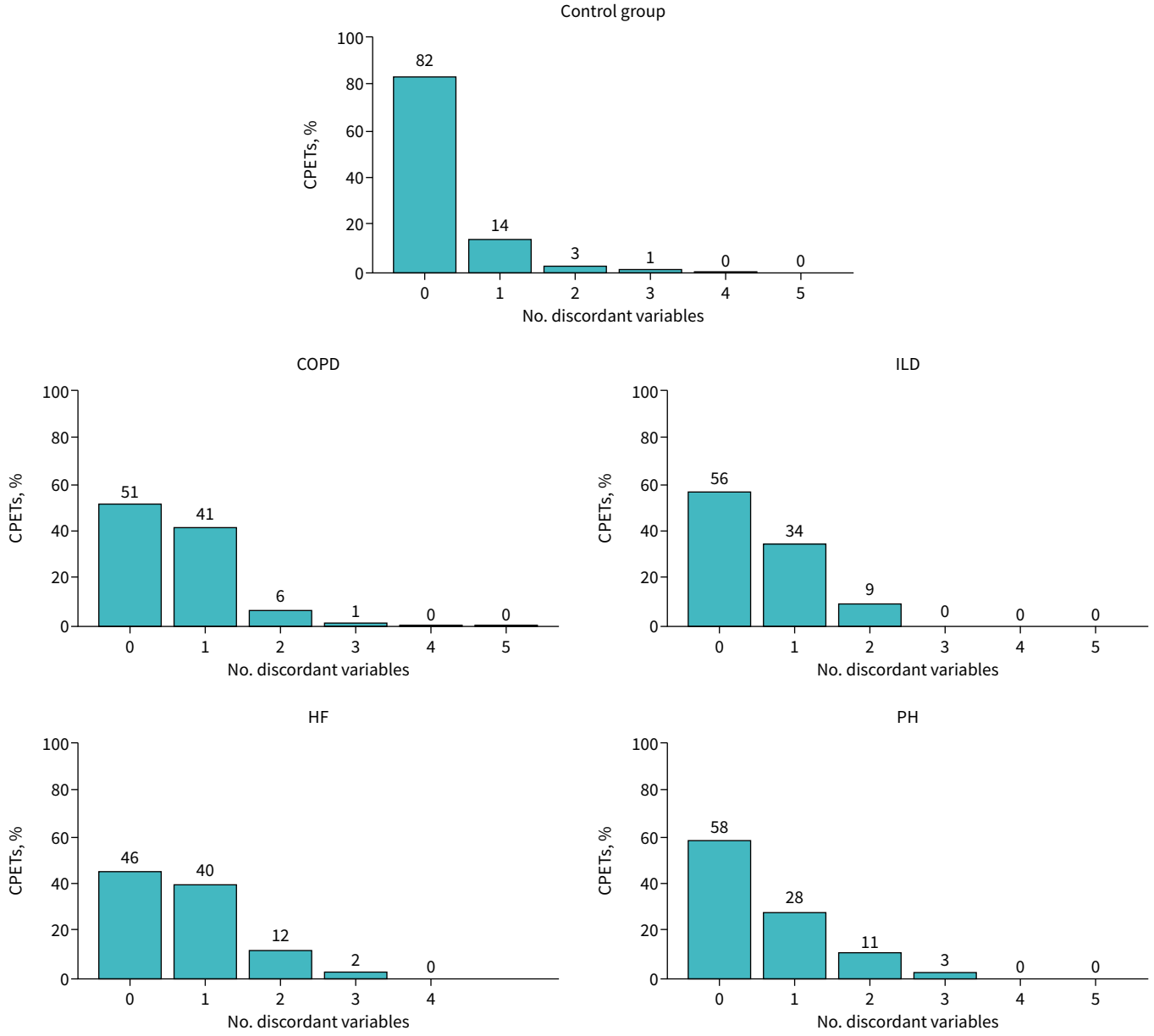


FIGURE 2 Distribution of the number of discordant variables per cardiopulmonary exercise testing (CPET), reflecting disagreement in classification between conventional and percentile-based thresholds. For heart failure (HF), only four variables were analysed because V'_{O_2}/HR was excluded due to the small sample size after removing patients taking β-blockers. V'_{O_2}: oxygen uptake; HR: heart rate; ILD: interstitial lung disease; PH: pulmonary hypertension.

percentile curves could offer an alternative for variables with non-normal distributions, similarly enabling continuous interpretation without requiring parametric assumptions [32].

This is the first study to evaluate the impact of CPET threshold selection in a large cohort of individuals with cardiorespiratory disease. However, our study has several limitations. First, its retrospective and single-centre design may limit generalisability. As a consequence, data obtained from two different metabolic carts were included, which may have introduced minor variability. However, daily calibration makes possible device-related differences likely minimal and nonsystematic. Additionally, differences in baseline disease severity and the use of medication beyond β-blockers may have influenced abnormality rates. Second, analysis of peak oxygen pulse in HF was not possible because SHIP reference equations for the 50th percentile are lacking for individuals on β-blockers. Third, the control group may have included individuals with unrecognised comorbidities, and no sensitivity analysis excluding borderline spirometry

Downloaded from https://publications.ersnet.org on September 29, 2026 at KU Leuven University Library. Please see licensing information on first page for reuse rights.

was performed. Nonetheless, this group represents a pragmatic sample of the general population rather than an idealised disease-free cohort. Finally, some key CPET variables, such as breathing reserve, were not included because percentile-based reference values are not yet available. Despite these limitations, our results expose an important issue in contemporary CPET interpretation.

Conclusion

The SHIP-derived percentile-based thresholds proved valid for a Western population and their application led to substantial differences in the classification of CPET results compared with conventional thresholds, which can alter CPET interpretation and clinical management. Although percentile-based thresholds provide a more statistically robust approach, their broader use remains limited by the incomplete coverage of CPET variables. Our findings underscore the urgent need for robust population-based reference values across ethnicities, ideally including Z-scores, to standardise CPET interpretation and enhance its clinical utility.

Provenance: Submitted article, peer reviewed.

Ethics statement: This study was approved by the Ethics Committee Research UZ/KU Leuven, approval number s67714.

Conflict of interest: M. Staes and M. Wuyts have nothing to disclose. M. De Vos reports support for the present manuscript from FWO project G0C9623N. T. Troosters reports consultancy fees from Roche; and payment or honoraria for lectures, presentations, speakers' bureaus, manuscript writing or educational events from AZ and Duomed, Belgium. W. Janssens reports grants from Chiesi and GSK; consultancy fees from AstraZeneca, Roche, Sanofi and GSK; and payment or honoraria for lectures, presentations, speakers' bureaus, manuscript writing or educational events from AstraZeneca, Roche, Sanofi and GSK.

References

1 Glaab T, Taube C. Practical guide to cardiopulmonary exercise testing in adults. *Respir Res* 2022; 23: 9.
2 Laveneziana P, Di Paolo M, Palange P. The clinical value of cardiopulmonary exercise testing in the modern era. *Eur Respir Rev* 2021; 30: 1–9.
3 American Thoracic Society, American College of Chest Physicians. ATS/ACCP statement on cardiopulmonary exercise testing. *Am J Respir Crit Care Med* 2003; 167: 211–277.
4 Staes M, Gyselinck I, Goetschalckx K, *et al.* Identifying limitations to exercise with incremental cardiopulmonary exercise testing: a scoping review. *Eur Respir Rev* 2024; 33: 240010.
5 Datta D, Normandin E, Zuwallack R. Cardiopulmonary exercise testing in the assessment of exertional dyspnea. *Ann Thorac Med* 2015; 10: 77–86.
6 Shreffler J, Huecker MR. Type I and type II errors and statistical power. *In*: StatPearls. Treasure Island, StatPearls Publishing, 2023.
7 Radtke T, Crook S, Kaltsakas G, *et al.* ERS statement on standardisation of cardiopulmonary exercise testing in chronic lung diseases. *Eur Respir Rev* 2019; 28: 180101.
8 Palange P, Laveneziana P, Neder JA, *et al*., eds. Clinical Exercise Testing (ERS Monograph). Sheffield, European Respiratory Society, 2018.
9 Stanojevic S, Kaminsky DA, Miller MR, *et al.* ERS/ATS technical standard on interpretive strategies for routine lung function tests. *Eur Respir J* 2022; 60: 2101499.
10 Quanjer PH, Stanojevic S, Cole TJ, *et al.* Multi-ethnic reference values for spirometry for the 3–95-yr age range: the Global Lung Function 2012 equations. *Eur Respir J* 2012; 40: 1324–1343.
11 Stanojevic S, Graham BL, Cooper BG, *et al.* Official ERS technical standards: Global Lung Function Initiative reference values for the carbon monoxide transfer factor for caucasians. *Eur Respir J* 2017; 50: 1700010.
12 Hall GL, Filipow N, Ruppel G, *et al.* Official ERS technical standard: Global Lung Function Initiative reference values for static lung volumes in individuals of European ancestry. *Eur Respir J* 2021; 57: 2000289.
13 Koch B, Schäper C, Ittermann T, *et al.* Reference values for cardiopulmonary exercise testing in healthy volunteers: the SHIP study. *Eur Respir J* 2009; 33: 389–397.
14 Sven G, Koch B, Ittermann T, *et al.* Influence of age, sex, body size, smoking, and β blockade on key gas exchange exercise parameters in an adult population. *Eur J Prev Cardiol* 2010; 17: 469–476.
15 Barron AJ, Dhutia NM, Gläser S, *et al.* Physiology of oxygen uptake kinetics: insights from incremental cardiopulmonary exercise testing in the Study of Health in Pomerania. *IJC Metab Endocr* 2015; 7: 3–9.
16 Gläser S, Ittermann T, Schäper C, *et al.* The Study of Health in Pomerania (SHIP) reference values for cardiopulmonary exercise testing. *Pneumologie* 2013; 67: 58–63.
17 Sietsema KE, Sue DY, Stringer WW, *et al.* Wasserman & Whipp's Principles of Exercise Testing and Interpretation. 6th Edn. Philadelphia, Wolters Kluwer Health, 2021.

Downloaded from https://publications.ersnet.org on September 29, 2026 at KU Leuven University Library. Please see licensing information on first page for reuse rights.

18 Pritchard A, Burns P, Correia J, *et al.* ARTP statement on cardiopulmonary exercise testing 2021. *BMJ Open Respir Res* 2021; 8: e001121.
19 Kersten J, Hoyo L, Wolf A, *et al.* Cardiopulmonary exercise testing distinguishes between post-COVID-19 as a dysfunctional syndrome and organ pathologies. *Int J Environ Res Public Health* 2022; 19: 11421.
20 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 1995; 57: 289–300.
21 Neder JA, Alharbi A, Berton DC, *et al.* Exercise ventilatory inefficiency adds to lung function in predicting mortality in COPD. *COPD* 2016; 13: 416–424.
22 Balady GJ, Arena R, Sietsema K, *et al.* Clinician's guide to cardiopulmonary exercise testing in adults. *Circulation* 2010; 122: 191–225.
23 Dardi F, Boucly A, Benza R, *et al.* Risk stratification and treatment goals in pulmonary arterial hypertension. *Eur Respir J* 2024; 64: 2401323.
24 Nayor M, Xanthakis V, Tanguay M, *et al.* Clinical and hemodynamic associations and prognostic implications of ventilatory efficiency in patients with preserved left ventricular systolic function. *Circ Heart Fail* 2020; 13: e006729.
25 Zannoni J, Guazzi M, Milani V, *et al.* Prognostic value of cardiopulmonary exercise testing in a European cohort with cardiovascular risk factors absent of a cardiovascular disease diagnosis. *Int J Cardiol* 2023; 370: 402–404.
26 Neder JA. Cardiopulmonary exercise testing applied to respiratory medicine: myths and facts. *Respir Med* 2023; 214: 107249
27 Blanchard J, Blais S, Chetaille P, *et al.* New reference values for cardiopulmonary exercise testing in children. *Med Sci Sports Exerc* 2018; 50: 1125–1133.
28 Humbert M, Kovacs G, Hoeper MM, *et al.* 2022 ESC/ERS guidelines for the diagnosis and treatment of pulmonary hypertension. *Eur Heart J* 2022; 43: 3618–3731.
29 Takken T, Mylius CF, Paap D, *et al.* Reference values for cardiopulmonary exercise testing in healthy subjects: an updated systematic review. *Expert Rev Cardiovasc Ther* 2019; 17: 413–426.
30 European Respiratory Society. Ongoing task forces. Date last accessed: 16 September 2025. www.ersnet.org/science-and-research/ongoing-task-forces/
31 Nes BM, Janszky I, Wisløff U, *et al.* Age-predicted maximal heart rate in healthy subjects: The HUNT Fitness Study. *Scand J Med Sci Sports* 2013; 23: 697–704.
32 Wang Y, Chen HJ. Use of percentiles and z-scores in anthropometry. *In:* Preedy VR, ed. Handbook of Anthropometry. New York, Springer New York, 2012; pp. 29–48.

Downloaded from https://publications.ersnet.org on September 29, 2026 at KU Leuven University Library. Please see licensing information on first page for reuse rights.