International Journal of Cardiology Congenital Heart Disease 18 (2024) 100540

Contents lists available at ScienceDirect

International Journal of Cardiology Congenital Heart Disease

journal homepage: www.journals.elsevier.com/international-journal-of-cardiology-congenital-heart-disease

Exercise testing in clinical context: Reference ranges for interpreting anaerobic threshold as an outcome for congenital heart disease patients

Katherine Hansen [a,b,*], Tracy Curran [a], Lindsey Reynolds [a], Catherine Cameron [a], Jennifer Pymm [a], Julie Ann O'Neill [a], Rachel Losi [a], Cara Sherman [a], Elise Ackermans [a], Suellen Yin [a], Tajinder Singh [a], Mark E. Alexander [a], Kimberlee Gauvreau [a], Naomi Gauthier [a]

[a] Department of Cardiology, Boston Children's Hospital, Boston, MA, USA
[b] Division of Pediatric Cardiology, UT Southwestern, Children's Medical Center, Dallas, TX, USA

ARTICLE INFO

Keywords:
Anaerobic threshold
Inter-rater variability
Reference range
Cardiopulmonary exercise test
Congenital heart disease

ABSTRACT

Background: Change in the oxygen consumption (VO_2) at the ventilatory anaerobic threshold (VAT) is an important outcome in research studies of children with congenital heart disease (CHD). The range of values reported by different raters for any given VAT is needed to contextualize a change in VAT in intervention studies.
Methods: Sixty maximal cardiopulmonary exercise tests (CPET) for CHD patients 8–21 years old were independently reviewed by six exercise physiologists and four pediatric cardiologists. For each of the unique rater pairs for the 60 CPETs, the absolute difference in VAT was calculated and displayed on a histogram to demonstrate the distribution of inter-rater variability. This method was repeated for subgroups of test modality (cycle/treadmill), patient factors (diagnoses, exercise capacity), and rater factors (cardiologist/physiologist, years of experience).
Results: Rater agreement was good with an intraclass correlation coefficient of 0.79–0.91 but the distribution of differences was broad. The median difference was 2.7 % predicted peak VO_2 (60 mL/min, 1.0 mL/kg/min), the 75th percentile was 6.4 % (140 mL/min, 2.5 mL/kg/min), and the 95th percentile was 16.3 % (421 mL/min, 6.5 mL/kg/min). Distributions were similar for CPET modality and years of rater experience, but differed for other factors.
Conclusions: The baseline distribution of reported VAT is relatively broad, varied by units, and was not explained by differences in rater experience or test modality, but varies by patient factors. When evaluating clinical relevance, a change in the VO_2 at VAT in response to an intervention of <6.5 % predicted falls within the majority (75th percentile) of expected variability and should be interpreted with caution.

1. Statements and declarations

The authors did not receive support from any organization for the submitted work. The authors have no relevant financial or non-financial interests to disclose.

The cardiopulmonary exercise test (CPET) is a robust assessment that can be used to assess functional status pre- and post-intervention in patients with congenital heart disease (CHD) ([1]). Peak oxygen consumption (VO_2) has been utilized as the primary outcome marker for decades, but requires a maximal CPET to assess which can sometimes be challenging to obtain ([2,3]). Submaximal fitness markers are also utilized and may be more commonly attainable and more pertinent clinically as most physical activity in daily life occurs at submaximal levels. Submaximal variables have been shown to be associated with health outcomes including all-cause mortality in an adult population ([4]), short-term adverse cardiac-related events in an adult CHD population ([5]), and fitness and quality of life in children with CHD ([6]).

The anaerobic threshold is one such submaximal marker of fitness. Studies assessing outcomes in CHD report the change in oxygen consumption at the ventilatory anaerobic threshold (VO_2 at VAT) in response to medication or exercise intervention as a clinically important marker of fitness ([7–12]). Understanding the variation that occurs in measurements is an essential consideration for any measurement tool. While inter-rater variability of the VO_2 at VAT has been evaluated previously both in adult ([13–18]) and pediatric ([19,20]) populations, variability has been primarily described using correlative statistics. These provide limited insight into the interpretation of magnitude of change or clinical importance of the VO_2 at VAT as an outcome metric. Multiple statistical methods have been developed and applied to

* Corresponding author. 1935 Medical District Drive B3440, Dallas, TX, USA 75235.
E-mail address: katherine.hansen@utsouthwestern.edu (K. Hansen).

https://doi.org/10.1016/j.ijcchd.2024.100540
Received 7 July 2024; Received in revised form 24 August 2024; Accepted 25 August 2024
Available online 27 August 2024

Abbreviations

VAT	ventilatory anaerobic threshold
CHD	congenital and pediatric acquired heart disease
CPET	cardiopulmonary exercise test
ICC	intraclass correlation coefficient
VCO_2	carbon dioxide production
VE/VCO_2	minute ventilation to oxygen consumption ratio
VO_2	oxygen consumption
VO_2 at VAT	oxygen consumption at the anaerobic threshold
$\%VO_2$ at VAT	oxygen consumption at the anaerobic threshold represented as % predicted peak oxygen consumption

quantify variability in measurements. Popovic et al. ([21]) argue that the standard error of the measure is preferred, but this is not able to be determined without a known "true" VO_2 at VAT. We aimed to describe the variability of the VO_2 at VAT using several descriptive analyses to provide context for interpreting changes in the VO_2 at VAT in response to an intervention as an outcome metric in pediatric heart disease, and evaluate if test-, rater-, or patient-specific factors contribute to variability of the VO_2 at VAT.

2. Methods

The Boston Children's Hospital database was queried for previously performed cardiopulmonary exercise tests. CPET inclusion criteria included testing done on patients who had CHD, were 8–21 years old, and were considered maximal as defined by a peak respiratory exchange ratio $\geq$1.09. CPETs that had sustained arrhythmia during the test or in recovery were excluded. Starting from January 1, 2022, CPETs that met inclusion and did not meet exclusion criteria were selected sequentially until 60 tests were chosen with testing modality equally represented (50 % treadmill, 50 % cycle ergometer). Every 10th CPET was selected from the chronological list of 60 total CPETs to yield 6 tests to be used as duplicates for intra-rater variability evaluation. All CPETs that were chosen from the database had been performed as part of the patient's clinical care either on a cycle ergometer using a ramp protocol or a treadmill using the Bruce protocol. Metabolic data were measured using an Ultima CPX™ metabolic stress testing system (MGC Diagnostics, St Paul, MN). Peak VO_2 was determined by the highest reliable value obtained during exercise, and prediction equations were used to determine percent predicted peak oxygen consumption per modality was done using standard clinical practices ([22]). Raters included six Master's trained exercise physiologists and four exercise cardiologists. Methods for reporting the AT have been extensively described in the literature. One method is the V-slope method: on a plot of VCO_2 and VO_2, the inflection point at which VCO_2 increases out of proportion to the VO_2 increase marks the VO_2 at VAT. The anaerobic threshold by the ventilatory equivalent method is identified at point at which the VE/VCO_2 (ratio of minute ventilation to carbon dioxide production) begins to increase while the VE/VO_2 (ratio of minute ventilation to oxygen consumption) remains flat. A third method utilizes the end-tidal pO_2; the anaerobic threshold can be marked as the point at which the pO_2 rises as a result of increased minute ventilation. CPET data is classically represented on the "nine-panel plot", and each of the above methods can be applied to the nine-panel plot to determine the VO_2 at VAT. In our exercise lab, these three modalities are utilized clinically to determine the VO_2 at VAT. Visschers et al. compared these three methods with the addition of respiratory exchange ratio, which is the ratio of VCO_2 to VO_2, in children with congenital heart or lung disease and found that three of the four methods, V-slope, ventilatory equivalent method, and the end-tidal O_2 method, had comparable agreement across a range of conditions and raters using the intraclass correlation coefficient (ICC) method [19].

The patient and test data were provided to the raters and each rater determined the VO_2 at VAT using the method(s) that best represented their typical practice. Cross checking across methods was allowed, as is commonly employed clinically, and no one method was pre-determined or chosen as the standard. Exercise physiologists individually selected the VO_2 at VAT digitally on the Ultima CPX™ system, adjusting the automated computer-generated VO_2 at VAT at their discretion using the method(s) above and recording the value on a study data collection sheet. Exercise cardiologists, who do not have ready access to the digital values on Ultima CPX™ system as part of usual workflow, were provided with printed copies of the same exercise test raw data as in clinical practice, and asked to choose the VO_2 at VAT based on the nine-panel plot and document the value on a study data collection sheet. Each rater evaluated every CPET, and all were blinded to the values chosen by the other raters. Six tests were duplicated to assess intra-rater variability, yielding 66 tests for each rater, and 10 raters in total. If the VO_2 at VAT was designated as indeterminate by any rater for any CPET, it was excluded from the analyses.

There were ten raters, and each rater's VAT value was compared to every other rater's values for each patient. For each pair, inter-rater variability was assessed by calculating absolute differences between raters for three measures: 1) VO_2 at VAT, expressed in mL/min; 2) weight-indexed VO_2 at VAT, expressed in mL/kg/min; and 3) $\%VO_2$ at VAT. The absolute differences for all possible combinations of raters were then compiled and used to assess the distribution of inter-rater variability for each measure using histograms. Summary measures calculated for each distribution of differences included the median (50th percentile) and every 5th percentile up to the 95th percentile, plus the 99th percentile. Intra-rater variability was quantified using the same technique, where all possible absolute differences were calculated within the same rater.

In addition to evaluating the overall distribution of inter-rater variability, separate distributions were generated stratified by CPET modality, rater type and experience, and patient diagnosis and peak VO_2. CPET modality was defined as either cycle ergometer or treadmill, rater type as either exercise physiologist or exercise cardiologist, rater experience as either $<$5 years or $\geq$5 years in practice, patient diagnosis as either simple or non-simple CHD (Appendix 1), and patient peak VO_2 as normal/mildly depressed or moderately/severely depressed (defined as percent predicted peak $VO_2 \geq$ 70 % or $<$70 % predicted, respectively). Separate histograms were created for differences measured for CPETs within each subgroup to display distributions of inter-rater variability. The intent was to describe these distributions; no hypotheses were tested.

To evaluate computer-generated versus final study AT, the $\%VO_2$ at VAT that was reported in the final report of the CPET, as determined by the exercise physiologist and exercise cardiologist who performed the final clinical interpretation of the test, was recorded. The software was then re-run using the software's proprietary algorithm to auto-detect the AT. The $\%VO_2$ at VAT of the final study was averaged across all 66 studies and was compared to the average of the computer-generated % VO_2 at VAT.

The study received Institutional Ethics approval waiving the need for consent as it utilized anonymized, clinically obtained data.

3. Results

Ten raters each evaluated all study CPETs. There were three individual instances of indeterminate VO_2 at VAT across all raters and CPETs which were excluded from the analyses. Six raters had 5 or more years of experience interpreting CPETs, and four had less than 5 years of experience. The median age of patients represented by the CPETs was 15 years, with a range of 8–21 years. Baseline rater, patient, and CPET characteristics are summarized in Table 1. Twenty-one (35 %) patients included in the sample had simple CHD, and 39 (65 %) had non-simple

Table 1
Baseline Characteristics. Median (range), or n (%).

Patient age (years)	15 (8–21)
Patient Sex	
Female	22 (37 %)
Male	38 (63 %)
Height (cm)	167 (132–189)
Weight (kg)	60 (28–166)
BSA (m^2)	1.67 (1.15–11.6)
BMI (kg/m^2)	22.1 (15.0–57.5)
Patient CHD Class	
Simple	21 (35 %)
Non-Simple	39 (65 %)
Complex	23 (38 %)
Single Ventricle	7 (12 %)
Cardiomyopathy/Transplant	9 (15 %)
Patient Peak VO_2 Class	
≥70 % predicted	40 (67 %)
<70 % predicted	20 (33 %)
Test Type	
Cycle Ergometer	30 (50 %)
Treadmill	30 (50 %)
Rater type	
Exercise Physiologists	6 (60 %)
Exercise Cardiologist	4 (40 %)
Rater Years of Experience	
<5	4 (40 %)
≥5	6 (60 %)

CHD (38 % had complex CHD, 15 % had cardiomyopathy or cardiac transplant and 12 % had single ventricle). Forty (67 %) patients had peak VO_2 ≥ 70 % predicted while 20 (33 %) had peak VO_2 <70 % predicted. The average VO_2 at VAT across all raters and CPETs was 1201 mL/min, 18.9 mL/kg/min when indexed by weight, or 48.4 % predicted.

There were 2673 pairs for analysis of inter-rater variability for the group as a whole. The median difference between raters in the VO_2 at VAT measurement was 60 mL/min, 1.0 mL/kg/min, and 2.7 % predicted. The range of differences was broad; the 75th percentile was 140 mL/min, 2.5 mL/kg/min, and 6.4 % predicted (Table 2). The intraclass correlation coefficient (ICC), a measure used to quantify agreement, is shown alongside the percentiles for additional context. The ICC suggested good agreement between raters, ranging from 0.79 to 0.91. The variability in reported VO_2 at VAT between raters is plotted in Fig. 1.

The variability by subgroups is available in Appendix 2. When evaluating by test modality, the median difference in %VO_2 at VAT was 2.5 % predicted for cycle ergometer vs 2.8 % predicted for treadmill (Fig. 2) with an ICC of 0.83 for cycle ergometer and 0.71 for treadmill. When evaluating for rater characteristics (profession, experience), the median difference was 0.9 % predicted for cardiologists (ICC = 0.85) versus 3.1 % predicted for exercise physiologists (ICC = 0.75) and 2.4 % predicted for less than 5 years experience and 2.7 % predicted for greater than 5 years experience. When evaluating for patient factors (CHD type, cardiorespiratory fitness level) the median difference was

Table 2
Percentiles and values for the differences in reported VAT.

Percentile	VO_2 at VAT (mL/min)	VO_2 at VAT by weight (mL/kg/min)	%VO_2 at VAT (%)
50th	60	1.0	2.7
55th	74	1.2	3.1
60th	88	1.5	3.7
65th	102	1.7	4.4
70th	118	2.1	5.4
75th	140	2.5	6.4
80th	185	2.9	7.7
85th	238	3.6	9.5
90th	301	4.5	11.2
95th	421	6.5	16.3
99th	744	10.6	24.7
Intra-class correlation coefficient	0.91	0.87	0.79

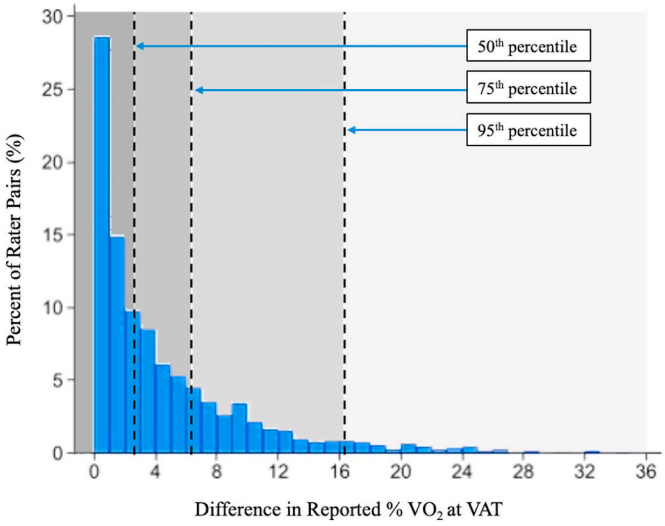

Fig. 1. Distribution of the range of reported differences in percent of predicted VO_2 at VAT.

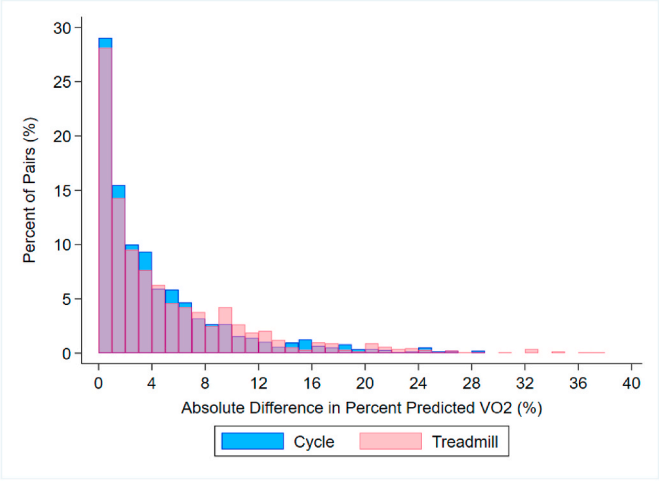

Fig. 2. Superimposed histograms for CPET modality.

3.5 % predicted for simple CHD and 2.3 % predicted for non-simple CHD, and 3.1 % predicted for normal or mildly depressed peak VO_2 vs 1.9 % predicted for moderately or severely depressed peak VO_2 (Appendix 2).

For intra-rater variability analyses, there were 60 possible pairs as there were 10 raters per CPET and each rater had 6 pairs. There were 50 % treadmill and 50 % cycle ergometer, representing patients that were 83 % male, a median body mass index of 19.9, 33 % of whom had simple CHD and 67 % had non-simple CHD. All studies were of patients with peak VO_2 >70 % predicted. The median intra-rater difference in %VO_2 at VAT was 1.6 % predicted (Appendix 3), lower than the inter-rater median difference of 2.7 % predicted.

The %VO_2 at VAT determined by the computer-generated algorithm was compared to the %VO_2 at VAT determined by exercise physiologists and exercise cardiologists. The computer-generated VO_2 at VAT was lower than the VO_2 at VAT chosen by the raters, with a mean difference of −4.8 % predicted and a median difference of −3.0 % predicted (Appendix 4).

4. Discussion

The present study demonstrates the distribution of variability for the

VO_2 at VAT, providing context for interpreting the change in VO_2 at VAT that may be reported in research studies in response to an intervention for children with CHD. These findings provide insight into this metric that correlative statistics, such as the intraclass correlation coefficient (ICC), do not provide.

Previous studies have evaluated the variability of the VO_2 at VAT and have found varying degrees of agreement. The intraclass correlation coefficient (ICC) is a commonly used measure of reproducibility with a value above 0.90 considered "excellent," 0.75–0.90 "good," 0.5–0.75 "moderate," and <0.5 "poor" ([23]). Kaczmarek et al. used the ICC to evaluate variability in the VO_2 at VAT determination by experienced physicians and medical assistants in asymptomatic volunteers. They found that, between physicians, the ICC was 0.901, which improved after a training period to an ICC of 0.95. In comparing physicians with trained medical assistants, the variability was greater, with an ICC of 0.759–0.762 [13]. These findings were consistent with our results; in our study the ICC was 0.79–0.91 across all raters, and 0.85–0.95 for physicians. Our data shows that, despite the ICC indicating "good" to "excellent" agreement between raters, the baseline distribution of reported VAT was relatively broad. This variability was not explained by rater experience or differences in exercise test modality, but did vary by patient factors. These findings suggest that when evaluating clinical relevance, a change in the VO_2 at VAT in response to intervention of <6.5 % (140 mL/min, 2.5 mL/kg/min) would fall within the majority (75th percentile) of expected variability and should be interpreted with caution.

There have been few investigations into whether or not the VO_2 at VAT variability increases depending on CPET protocol, patient, or rater factors. Studies often used CPETs from mixed modalities (cycle ergometer and treadmill) as predicted values for oxygen consumption take the modalities into account. For the VO_2 at VAT, however, it is not known whether the variability in reporting would be different between cycle ergometer and treadmill testing. In our data, the median inter-rater differences were similar for treadmill (2.8 % predicted) and for cycle ergometer (2.5 % predicted), and the histograms of each modality overlap substantially, as demonstrated in Fig. 2. This provides some reassurance in the reliability of using the change in VO_2 at VAT in studies with mixed modalities.

When examining characteristics of the raters, the cardiologists had less variability than the exercise physiologists, with a median difference of 0.9 percent VO_2 at VAT for cardiologists and 3.1 percent VO_2 at VAT for exercise physiologists. Interestingly, greater experience in reading CPETs did not factor into the observed variability in the VO_2 at VAT determination.

The VO_2 at VAT variability by patient factors was also analyzed and showed that the variability of the VO_2 at VAT was generally greater for the patients with simple disease and normal-range peak VO_2. This could potentially be due to higher overall peak VO_2 measurement in more fit individuals, which means potentially greater variability in determining the VAT. Our findings are reinforced in that our peak VO_2 and diagnosis-based subgroup results are similar; it seems that greater variability occurs in healthier patients as compared to less well patients, possibly due to the greater range of data points for patients achieving a higher peak VO_2. The peak VO_2 subgroups are represented in Appendix 2 to provide a peak VO_2-specific distribution which can be used to contextualize an individual's VO_2 at VAT within the distribution of inter-rater variability.

The distribution of values listed in Table 2 and found on the histograms in Fig. 1 can be utilized as references to evaluate whether a reported change in VO_2 at VAT from a treatment or intervention in a clinical research study lies close to the median of expected variation (50th percentile) or if that change lies beyond the expected distribution of values. For instance, a study from 2005 reported a mean improvement in VO_2 at VAT of 4 mL/kg/min or 10.8 % predicted after cardiac rehabilitation in pediatric patients with congenital heart disease ([12]). When compared to our reference ranges, these were greater than the 85th percentile of VO_2 at VAT measurement differences in our study, supporting this reported change as likely due to the treatment rather than potentially related to variation in the VO_2 at VAT measurement. The same group showed that improvements were sustained over time; the VO_2 at VAT had increased a mean of 3.6 mL/kg/min, or 7.2 % predicted ([24]), which are at the 85th percentile and between the 75th and 80th percentiles, respectively, on the VO_2 at VAT difference distribution histogram. In a study evaluating the effect of an exercise prescription on adolescent Fontan patients ([10]), there was a statistically significant improvement in their VO_2 at VAT from 18 ± 3.5 to 20 ± 4.8 mL/kg/min. However, an increase of 2 mL/kg/min falls between the 65th and 70th percentile on the distribution. In this case, the change in the VO_2 at VAT may be related to the variability of the metric and not the intervention itself.

Changes in the VO_2 at VAT have also been reported in response to medications. The FUEL Trial reported a statistically significant improvement in the VO_2 at VAT in Fontan patients treated with Udenafil of 33 mL/min in the treatment group ([8]). An increase of 33 mL/min in the VO_2 at VAT falls within the median range of our data (see Table 2) and thus may not reflect clinical change. In the follow up Fontan patient study, serial exercise testing demonstrated a decline in percent predicted VO_2 at VAT of 0.8 ± 2.6 % which, while statistically significant, is again within expected variation ([7]). Certainly, an increase in response to treatment or decrease over time in VO_2 at VAT should not be discounted, but may require additional evaluation for variability of VAT values pre- and post-assessment to establish typical ranges.

Individual pediatric CPET laboratories may have differences in reporting procedures ([25]), including how the VAT is reported. Variability of the measurement may differ substantially across labs. Quality improvement and educational initiatives can improve the variability, as demonstrated by Prusi et al. ([26]). In their initiative, baseline interrater reliability between exercise physiologists and the interpreting cardiologist was 20 %, indicating that 80 % of tests required correction of the AT by the interpreting cardiologist. After an educational intervention, reliability improved to >80 %. Prusi et al. noted that incorrect automated AT determination by the vendor platform was likely a contributor to the low baseline reliability. In our study, there was substantial difference between the computer-generated VO_2 at VAT and that determined by the exercise physiologists and cardiologists. The computer software generally reported the VO_2 at VAT as lower than the manually chosen VO_2 at VAT. Accepting the computer-generated VO_2 at VAT without critical evaluation of the data by an expert may result in inaccuracies and contribute to greater variability.

There are several limitations to this study. While an analysis of variability of the anaerobic threshold for 10 raters and 60 CPETs at a single institution is greater than most VAT variability studies, these findings may not be generalizable to other institutions. There is significant variability across pediatric CPET labs nationally related to quality control, workflow, physician presence, reporting practices, and software ([25]). Certainly, other institutions may have less or greater variability in their VAT determination depending in the fidelity of the data and methods used. Second, the tests used for intra-rater correlations were skewed toward patients with normal-range peak VO_2, although the intra-rater results were less relevant to the main study question. Finally, the study design introduces the possibility of the Hawthorne effect; raters may unwittingly modify their practice and thus the true clinical VAT variability at our institution may differ from the results of this study.

These results demonstrate that, despite good inter- and intra-rater agreement, there is a broad distribution in differences in reporting of the VO_2 at VAT. While prior studies have demonstrated that there is variability of the measurement, they generally conclude that the variability is acceptable based on correlative statistics. This does not provide clarity on whether a specified magnitude of change VO_2 at VAT is clinically relevant, however. The data presented in the present study provides useful context for interpretation of the measurement within the expected range of differences between raters. Future studies evaluating

the VO_2 at VAT as an outcome metric should consider the variability of the VAT determination when evaluating the clinical meaningfulness of their findings. Quality improvement and educational initiatives within pediatric CPET laboratories may improve variability across raters.

CRediT authorship contribution statement

Katherine Hansen: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Tracy Curran:** Writing – review & editing, Resources, Methodology. **Lindsey Reynolds:** Writing – review & editing. **Catherine Cameron:** Writing – review & editing, Resources. **Jennifer Pymm:** Writing – review & editing, Resources, Conceptualization. **Julie Ann O'Neill:** Writing – review & editing, Resources. **Rachel Losi:** Writing – review & editing, Resources. **Cara Sherman:** Writing – review & editing, Resources. **Elise Ackermans:** Writing – review & editing, Resources. **Suellen Yin:** Writing – review & editing, Resources. **Tajinder Singh:** Writing – review & editing, Resources. **Mark E. Alexander:** Writing – review & editing, Resources, Conceptualization. **Kimberlee Gauvreau:** Writing – review & editing, Resources, Methodology, Formal analysis, Data curation, Conceptualization. **Naomi Gauthier:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijcchd.2024.100540.

References

[1] Rhodes J, Alexander ME, Opotowsky AR, editors. Exercise physiology for the pediatric and congenital cardiologist. Springer International Publishing; 2019. https://doi.org/10.1007/978-3-030-16818-6.

[2] Rowland TW. Does peak VO2 reflect VO2max in children?: evidence from supramaximal testing. Med Sci Sports Exerc 1993;25(6):689–93.

[3] Armstrong N, Welsman J, Winsley R. Is peak VO2 a maximal index of Children's aerobic fitness? Int J Sports Med 1996;17(5):356–9. https://doi.org/10.1055/s-2007-972860.

[4] Laukkanen JA, Kunutsor SK, Yates T, Willeit P, Kujala UM, Khan H, et al. Prognostic relevance of cardiorespiratory fitness as assessed by submaximal exercise testing for all-cause mortality: a UK biobank prospective study. Mayo Clin Proc 2020;95(5):867–78. https://doi.org/10.1016/j.mayocp.2019.12.030.

[5] Das BB, Godoy A, Kadish T, Niu J. Maximal versus sub-maximal effort during cardiopulmonary exercise testing in adults with congenital heart disease: outcome analysis of short-term cardiac-related events. Cardiol Young 2021;31(1):91–6. https://doi.org/10.1017/S104795112000339X.

[6] Amedro P, Picot MC, Moniotte S, Dorka R, Bertet H, Guillaumont S, et al. Correlation between cardio-pulmonary exercise test variables and health-related quality of life among children with congenital heart diseases. Int J Cardiol 2016; 203:1052–60. https://doi.org/10.1016/j.ijcard.2015.11.028.

[7] Goldberg DJ, Zak V, McCrindle BW, Ni H, Gongwer R, Rhodes J, et al. Exercise capacity and predictors of performance after Fontan: results from the pediatric heart network Fontan 3 study. Pediatr Cardiol 2021;42:158–68. 10.1007/s00246-020-02465-1.

[8] Goldberg DJ, Zak V, Goldstein BH, Schumacher KR, Rhodes J, Penny DJ, et al. Results of the FUEL trial. Circulation 2020;141(8):641–51. https://doi.org/10.1161/CIRCULATIONAHA.119.044352.

[9] Seckeler MD, Barber BJ, Colombo JN, Bernardi AM, Hoyer AW, Andrews JG, et al. Exercise performance in adolescents with fontan physiology (from the pediatric heart network fontan public data set). AJC 2021;149:119–25.

[10] Pyykkönen H, Rahkonen O, Ratia N, Lähteenmäki S, Tikkanen H, Piirilä P, et al. Exercise prescription enhances maximal oxygen uptake and anaerobic threshold in young single ventricle patients with Fontan circulation. Pediatr Cardiol 2022;43 (5):969–76.

[11] Lin KL, Liou IH, Chen GB, Sun SF, Weng KP, Li CH, et al. Serial exercise testing and echocardiography findings of patients with kawasaki disease. Front Pediatr 2022; 10:847343.

[12] Rhodes J, Curran TJ, Camil L, Rabideau N, Fulton DR, Gauthier NS, et al. Impact of cardiac rehabilitation on the exercise function of children with serious congenital heart disease. Pediatrics 2005;116(6):1339–45. https://doi.org/10.1542/peds.2004-2697.

[13] Kaczmarek S, Habedank D, Obst A, Dörr M, Völzke H, Gläser S, et al. Interobserver variability of ventilatory anaerobic threshold in asymptomatic volunteers. Multidiscip Respir Med 2019;14(1):20. https://doi.org/10.1186/s40248-019-0183-6.

[14] Myers J, Goldsmith RL, Keteyian SJ, Brawner CA, Brazil DA, Aldred H, et al. The ventilatory anaerobic threshold in heart failure: a multicenter evaluation of reliability. J Card Fail 2010;16(1):76–83. https://doi.org/10.1016/j.cardfail.2009.08.009.

[15] Shimizu M, Myers J, Buchanan N, Walsh D, Kraemer M, McAuley P, et al. The ventilatory threshold: method, protocol, and evaluator agreement. Am Heart J 1991;122(2):509–16. https://doi.org/10.1016/0002-8703(91)91009-C.

[16] Pereira DAG, Vieira DSR, Samora GAR, Lopes FL, Alencar MCN, Lage SM, et al. Reproducibility of the determination of anaerobic threshold in patients with heart failure. Arq Bras Cardiol 2010;94:771–8. https://doi.org/10.1590/S0066-782X2010005000044.

[17] Sinclair RCF, Danjoux GR, Goodridge V, Batterham AM. Determination of the anaerobic threshold in the pre-operative assessment clinic: inter-observer measurement error. Anaesthesia 2009;64(11):1192–5.

[18] Ohuchi H, Nakajima T, Kawade M, Matsuda M, Kamiya T. Measurement and validity of the ventilatory threshold in patients with congenital heart disease. Pediatr Cardiol 1996;17(1):7–14.

[19] NaomiCA Visschers, ErikH Hulzebos, Marco van Brussel, Tim Takken. Comparing four non-invasive methods to determine the ventilatory anaerobic threshold during cardiopulmonary exercise testing in children with congenital heart or lung disease. Clin Physiol Funct Imag 2015;35(6):451–9. https://doi.org/10.1111/cpf.12183.

[20] Takken T, van der Net J, Helders PJ. The reliability of an aerobic and an anaerobic exercise tolerance test in patients with juvenile onset dermatomyositis. J Rheumatol 2005;32(4):734–9. https://doi.org/10.3899/jrheum.32.4.734.

[21] Popović ZB, Thomas JD. Assessing observer variability: a user's guide. Cardiovasc Diagn Ther 2017;7(3):317–24. https://doi.org/10.21037/cdt.2017.03.12.

[22] Rhodes J, Ubeda Tikkanen A, Jenkins KJ. Exercise testing and training in children with congenital heart disease. Circulation 2010;122(19):1957–67. https://doi.org/10.1161/CIRCULATIONAHA.110.958025.

[23] Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15(2):155–63. https://doi.org/10.1016/j.jcm.2016.02.012.

[24] Rhodes J, Curran TJ, Camil L, Rabideau N, Fulton DR, Gauthier NS, Gauvreau K, Jenkins KJ. Sustained effects of cardiac rehabilitation in children with serious congenital heart disease. Pediatrics 2006 Sep 1;118(3):e586–93. https://doi.org/10.1542/peds.2005-1891.

[25] Chang RKR, Gurvitz M, Rodriguez S, Hong E, Klitzner TS. Current practice of exercise stress testing among pediatric cardiology and pulmonology centers in the United States. Pediatr Cardiol 2006;27(1):110–6. https://doi.org/10.1007/s00246-005-1046-9.

[26] Prusi M, Stewart C, Yono S, Garofolo L, Largent B, Hansen J. Improving anaerobic threshold determination: a quality improvement study in a pediatric clinical exercise laboratory. In: Poster presented at: cardiology 2022: 25th annual update on pediatric and congenital cardiovascular disease; September 3, 2022. Huntington Beach, CA.