Journal of the American Heart Association

ORIGINAL RESEARCH

Percentage Predicted Peak Oxygen Consumption in People With Fontan Circulation: A Rapid Systematic Scoping Review and Validation Study

Curtis A. Wadey, PhD; Owen W. Tomlinson, PhD; Alan R. Barker, PhD; A. Graham Stuart, MBChB, PGcert, MSc; Derek L. Tran, PhD; Karina Laohachai, MBBS; Julian Ayer, BSc (Med), MBBS, PhD; Robert G. Weintraub, MBBS (Hons); Rachael Cordina, MBBS, PhD; Craig A. Williams, PhD

BACKGROUND: Peak oxygen consumption (peak $\dot{V}O_2$) is routinely measured in people who have congenital heart disease and is reported as a percentage of predicted value, based upon age- and sex-matched normative reference values (NRVs). This study aimed to identify which NRVs are being used, assess whether NRVs are being applied appropriately, and evaluate if recommended NRVs are valid when applied to people with congenital heart disease.

METHODS AND RESULTS: A systematic scoping review identified studies that reported peak $\dot{V}O_2$ percentage of predicted value in people with congenital heart disease. A modified risk of bias tool evaluated the included studies. Forty-five studies reported peak $\dot{V}O_2$ percentage of predicted value, and only 21 (47%) studies described or provided a reference on how their percentage of predicted value was calculated. The most cited NRVs were from Wasserman (n=12) and Cooper and Weiler-Ravell (n=7). Risk of bias analysis judged 63% of studies as having some concerns. The NRVs recommended by the American Heart Association were applied to participants with a Fontan circulation (n=70; aged 26.5±6.4 years; 59% women) to examine validity. Predicted peak $\dot{V}O_2$ values from the Wasserman NRV was not significantly associated to measured peak $\dot{V}O_2$ values (men: b=0.31, R^2≤0.01; women: b=0.07, R^2=0.02).

CONCLUSIONS: Numerous NRVs have been applied to individuals with congenital heart disease and are often poorly reported and inappropriately matched to participants. The Wasserman NRV was the most cited but showed poor validity when applied to a Fontan cohort.

Key Words: cardiorespiratory fitness ■ congenital heart disease ■ CPET

Peak oxygen consumption (peak $\dot{V}O_2$) measured using gold-standard cardiopulmonary exercise testing (CPET) is an objective measure of cardiorespiratory fitness,[1] and is commonly lower in people with congenital heart disease (ConHD).[2] Peak $\dot{V}O_2$ is a key parameter used in the management of people with ConHD, because it has been shown to be predictive of future major adverse cardiovascular events, independent of traditional risk factors.[3] Peak $\dot{V}O_2$ is directly measured in liters per minute (L/min^{-1}), although is typically expressed per kilogram of body mass (mL/kg^{-1} per min^{-1}), in an attempt to create a size-independent expression.[4] However, expressing peak $\dot{V}O_2$ in ratio to body mass often fails to appropriately normalize for

Correspondence to: Craig A. Williams, PhD, Children's Health & Exercise Research Centre (CHERC), University of Exeter, Faculty of Life and Health Sciences, St. Luke's Campus, Heavitree Road, Exeter EX1 2LU, Devon, United Kingdom. Email: c.a.williams@exeter.ac.uk

This article was sent to Sakima A. Smith, MD, MPH, Associate Editor, for review by expert referees, editorial decision, and final disposition.

Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.123.034035

For Sources of Funding and Disclosures, see page 9.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- Peak oxygen consumption is commonly reported as a percentage of predicted value based on a normative reference value.
- The Wasserman normative reference value was not valid when applied to people with complex congenital heart disease.

What Are the Clinical Implications?

- Interpret percentage of predicted values with caution; if using normative reference values in clinical practice, ensure they are matched to the patient's age, sex, ethnicity, and exercise modality.

Nonstandard Abbreviations and Acronyms

$\%_{Pred}$	percentage of predicted
AHA	American Heart Association
ConHD	congenital heart disease
CPET	cardiopulmonary exercise testing
NRV	normative reference value
Peak $\dot{V}O_2$	peak oxygen consumption
RoB	risk of bias

body size and composition due to known statistical artifacts, which unfairly penalize heavier individuals.[4–7] Therefore, for ease of interpretation by clinicians and patients alike, peak $\dot{V}O_2$ (L/min^{-1} or mL/kg^{-1} per min^{-1}) can also be expressed as a percentage of predicted ($\%_{Pred}$) value, accounting for an individual's age, sex, and body size by comparing their measured peak $\dot{V}O_2$ to published normative reference values (NRVs).

Systematic reviews have identified ≈60 NRVs that have been derived from healthy populations, 29 of which were published between 2014 and 2019.[8,9] Published NRVs have been reported as having questionable methodological rigor and predictive performance.[8,9] Specific issues include a lack of quality assurance (ie, equipment and procedures), statistical power, and validation.[9] The rapid growth of available NRVs has increased the uncertainty on which one to choose for a given participant or patient population.

There are joint scientific statements from the American Thoracic Society and the American College of Chest Physicians,[10] and from the European Association for Cardiovascular Prevention and Rehabilitation and the American Heart Association (AHA),[11,12] which recommend the NRV produced by Wasserman et al,[13] despite no data to validate its use in people with ConHD. Moreover, no single NRV is consistently recommended by other medical societies or training organizations, even within specific medical disciplines. For example, the Association for Respiratory Technology and Physiology[14] and the European Respiratory Society,[15,16] both of which are leading respiratory medicine organizations, recommend a range of alternative NRVs[17–19] and provide little to no evidence-based justification for their recommendation.

A recent case report has highlighted the impact of poor standardization in the application of NRVs to clinical populations, which has been detrimental to clinical decision making and patient outcomes.[20] Furthermore, the risk of applying inappropriately matched or poorly performing NRVs to clinical populations may further result in poor peak $\dot{V}O_2$ predictions, confounding peak $\dot{V}O_2$ expressed as $\%_{Pred}$. Additionally, in research, the use of different NRVs to interpret the same raw participant data has been shown to produce different conclusions (ie, some NRVs show significant improvements in fitness, whereas others show no improvement), which may produce misinterpretations of data and invalid conclusions.[21]

Practitioners should be cautious and select NRVs that are appropriate for their cohort's demographics (ie, age, sex, body size, ethnicity) and related to the specific chosen exercise modality (ie, running, cycling). For example, the application of an NRV derived from a treadmill exercise test should not be used to interpret data collected from a cycle-based exercise test, and vice versa, due to the known difference in the peak $\dot{V}O_2$ between modalities.[22] Similarly, it would be inappropriate to apply an NRV derived from adult populations to pediatric populations, because absolute $\dot{V}O_2$ (L/min^{-1}) is closely related to body size.[6] Finally, some NRVs produce considerable error and thus may have limited practical or clinical usefulness in apparently healthy individuals.[23,24]

The error in the prediction at an individual level from an NRV is especially important to consider for clinical groups including those with complex ConHD, where a routine CPET is recommended in their long-term follow-up and aids clinical decision making.[25] An inaccurate prediction can potentially influence clinical decision making and alter patient health outcomes. However, the extent to which NRVs are used in ConHD is unknown, as is the prospective error associated with using NRVs for interpreting peak $\dot{V}O_2$ in this group.

Therefore, the overall aim of this study was to assess the application of NRVs to people with ConHD. Specifically, this study sought to (1) identify which NRVs are being applied to individuals with ConHD; (2) assess whether NRVs are being appropriately matched for age, sex, ethnicity, and exercise modality in study populations; and (3) evaluate, for the first time, NRVs

recommended by the AHA to people who have complex ConHD.

METHODS

Study Design

To address aims 1 and 2, a rapid scoping review (reported in accordance with the Preferred Reporting Items for Systematic reviews and Meta-analyses guidelines; Table S1) was performed to identify NRVs used within the ConHD literature and to assess if they are applied appropriately using a novel risk of bias (RoB) approach. For aim 3, the validity of the Wasserman NRV[13] (as recommended by the European Association for Cardiovascular Prevention and Rehabilitation/AHA statement[11]) was assessed by applying the NRVs to adults who have a Fontan circulation (n=70). For this, participant data were used from the Australian and New Zealand Fontan Registry.[26–28] Ethics approval for the data collection was granted by the relevant ethics review committee at each participating center where subjects gave informed consent, and the Australian and New Zealand Fontan Registry Steering Committee approved data transfer agreements in 2022. The data that support the findings of this study are available from Dr Cordina upon reasonable scientific request.

Scoping Review (Aims 1 and 2)

Studies that performed a CPET with the aim of predicting clinically important outcomes (eg, death, transplantation, clinical worsening) in a population of people who have structural ConHD were included. Studies with the aim of predicting clinical events were included because they were likely to report outcomes important to patients that may suggest changes in patient management based on peak $\dot{V}O_2$ for risk stratification and to narrow down the potential pool of studies. Studies were included regardless of center location, publication year, candidate prognostic factor, exercise modality, and exercise protocol. However, studies that performed a CPET without pulmonary gas analysis or did not report peak $\dot{V}O_2$ as a $\%_{Pred}$ were excluded.

The search strategy was adopted from a previously published systematic review on the prognostic role of CPET in people with ConHD (Data S1).[3] The following electronic databases were originally searched on April 30, 2020, and for the purpose of this rapid review, an updated search was undertaken on April 24, 2023: Allied and Complementary Medicine Database (EBSCO), CINAHL Complete (EBSCO), SPORTDiscus (EBSCO), Medline (Ovid), Embase (Ovid), Web of Science (Thomson Reuters), and Cochrane Central Register of Controlled Trials. The search terms included prognosis (mortality, morbidity, event-free survival) with ConHD (Fontan, tetralogy of Fallot), and CPET parameters (peak $\dot{V}O_2$, exercise test) (Data S1). The first author performed the initial title and abstract and full-text screening for both the original and updated search, and subsequently extracted study-level (eg, study location, CPET protocols) and participant-level data (eg, age, sex, ethnicity), using piloted forms and then synthesized study data.

Risk of Bias Assessment

The RoB analysis was performed only on those studies that cited an NRV equation, using a modified RoB tool that has been used previously.[29] The RoB assessment was performed by the first author and ascertained whether the NRV selected by a study was appropriate for that study's population (age, sex, ethnicity) and exercise modality.

To achieve a low RoB, studies had to achieve an exact match (eg, study participants were exercised on a cycle ergometer, and the NRV was derived from a population tested using cycle ergometry); some concerns were if either the study or the NRV did not report sufficient information or only matched for 1 aspect (eg, the study has both cycle ergometry and treadmill ergometry and only used a cycle ergometry-based NRV) and a high RoB if the study cohort did not explicitly match the NRV (eg, the study performed cycle ergometry and used an NRV based from treadmill ergometry). The highest RoB score across the 4 domains was retained as that study's overall RoB (eg, the study was judged as a high RoB in the domain of modality, and thus the study was judged as having a high RoB overall).

A randomized subsample (via internet random number generator) of 10% of included publications were independently extracted and had an RoB assessed by a second author (O.W.T.). There were no discrepancies between these 2 authors for this 10% subsample (ie, 100% agreement).

Validation (Aim 3)

Validation of NRV recommended by the AHA was performed in a cohort of adult Fontan patients (aged ≥18 years), whose data were extracted from the Australian and New Zealand Fontan Registry.[26–28] All patients, in contributing data to this registry, performed a CPET on an electronically braked cycle ergometer using a ramp-incremental protocol.[26] The protocol was individualized with the aim of the participant reaching volitional exhaustion in approximately 8 to 12 minutes. This protocol has been demonstrated to have excellent test–retest reliability for the outcome of peak $\dot{V}O_2$ (intraclass correlation coefficient=0.95 [95% CI, 0.94–0.97]) in participants with cardiac and respiratory disease.[30] Participant's dyspnea and ratings of perceived exertion were recorded using a modified Borg scale.

The peak $\dot{V}O_2$ was measured breath by breath and exported in 20 seconds, and peak $\dot{V}O_2$ was defined as the highest 20-second average.[26,31]

The Wasserman NRV[13] (Table) were selected for validation due to its inclusion within the current guidelines, being recommended by both the European Association for Cardiovascular Prevention and Rehabilitation and AHA.[11,12] The NRV was applied to 70 adult Fontan participants (aged 26.5±6.4 years; 59% women; body mass, 68.9±14.6 kg; body mass index, 24.7±4.8 kg/m^2; aortopulmonary n=12; lateral tunnel n=23; extracardiac conduit n=35).

The predicted peak $\dot{V}O_2$ (mL/min^{-1}) was calculated by using the full process documented in the Table, for subsequent comparison against measured peak $\dot{V}O_2$. At first, the participant's measured body weight was compared with their predicted body weight, and then predicted peak $\dot{V}O_2$ was calculated depending on their weight status (ie, underweight, at weight±1 kg, or overweight; Table). As a sensitivity analysis, predicted peak $\dot{V}O_2$ was also calculated using the participants measured mass, assuming their body weight status was normal. This was to assess the influence of correcting for the participants predicted body weight in this myopenic population.[26,32]

Statistical Analysis

Data are presented as mean±SD or as a count and proportion (percent). The validity of applying the Wasserman NRV to the Fontan cohort were assessed by plotting the relationship between the predicted and the measured peak $\dot{V}O_2$ values using a line of identity plot and performing linear regression analyses. Regression analyses reported the intercept, slope, and the coefficient of determination for measured and predicted peak $\dot{V}O_2$, for men and women, respectively. Microsoft Excel (version 2208; Microsoft, Redmond, WA), IBM SPSS Statistics (version 28.0; IBM, Armonk, NY), and GraphPad (version 9; GraphPad Software, San Diego, CA) were used for all analyses, and the α level was 0.05.[33]

RESULTS

Aims 1 and 2: Scoping Review

Study Selection and Characteristics of Included Studies

One hundred twenty articles were assessed for full-text inclusion, and 45 (38%) were included in this rapid scoping review (Figure 1, Table S2). These 45 studies encompass 15 806 participants, ≈43% were women, and the median age was 27.6 years (range, 9–49 years). The studies were published between 2009 and 2023 and included both pediatric and adult ConHD populations. The research originated from Europe (44%), North America (38%), Asia (11%), multicenter/continent collaboration (4%), and Australia (2%).

Cardiopulmonary Exercise Tests

Of the 45 studies, 17 (38%) were performed on a treadmill, 14 (31%) were performed on a cycle ergometer, 10 (22%) combined cycle ergometry and treadmill ergometry data, and 4 (9%) did not report the exercise modality. The 3 most reported exercise protocols were the ramp protocol (n=16, 33%), the Naughton protocol (n=3, 6%) for cycle ergometry, and the Bruce/modified Bruce protocol for treadmill ergometry (n=13, 27%). A large proportion of studies did not report the protocol (n=9, 18%) or were not explicit (n=8, 16%), for example stating symptom-limited protocol.

The criteria for verifying a maximal test were inconsistent. Twenty-three (51%) studies did not report if they confirmed a maximal test, 13 (29%) studies used a range of respiratory exchange ratio cut points (ie, >1.00, >1.05, >1.10, and >1.15), 6 (13%) studies used a combination of respiratory exchange ratio cut points and 1 other criteria (ie, percent of predicted peak heart

Table. Process to Calculate Predicted Peak $\dot{V}O_2$ (mL/min^{-1})[13]

Equation	Men	Women
Predicted mass =	0.79×height−60.7	0.65×height−42.8
Cycle factor =	50.72−0.372×age	22.78−0.17×age
If measured mass = predicted mass (±1 kg):	Peak $\dot{V}O_2$=measured mass×cycle factor	Peak $\dot{V}O_2$=(measured mass+43)×cycle factor
If measured mass < predicted mass:	Peak $\dot{V}O_2$=[(predicted mass+measured mass)/2] ×cycle factor	Peak $\dot{V}O_2$=[(predicted mass+measured mass+86)/2] ×cycle factor
If measured mass > predicted mass:	Peak $\dot{V}O_2$=(predicted mass×cycle factor)+6×(measured mass−predicted mass)	Peak $\dot{V}O_2$=[(predicted mass+43) ×cycle factor]+6×(measured mass−predicted mass)
Using measured mass only*	Peak $\dot{V}O_2$=measured mass×cycle factor	Peak $\dot{V}O_2$=(measured mass+43) ×cycle factor

Peak $\dot{V}O_2$ indicates peak oxygen consumption.
*Sensitivity analysis.

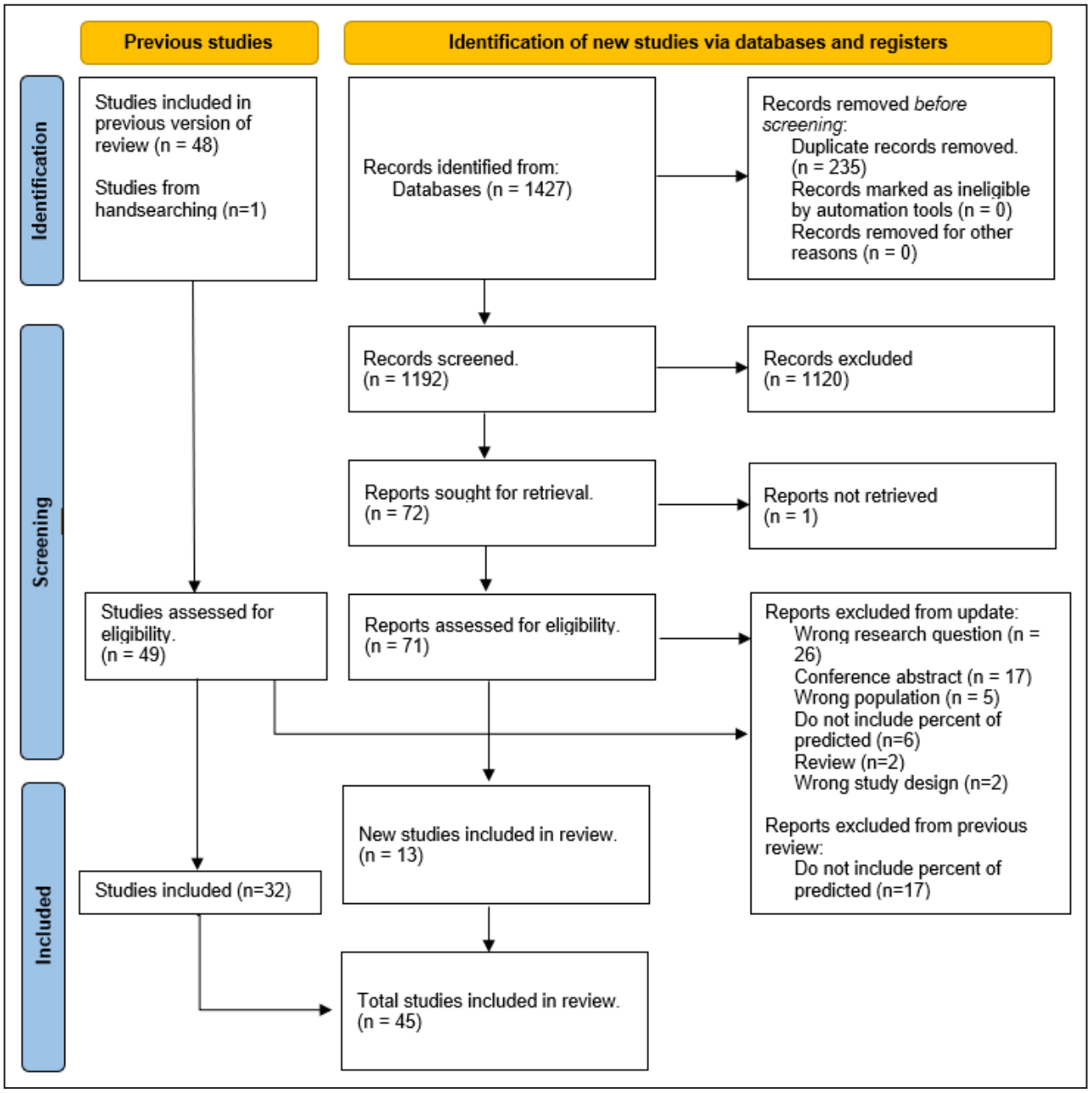

Figure 1. Study selection flow diagram.

rate or ratings of perceived exertion), and 2 studies (7%) terminated the CPET once a participant reached a specific respiratory exchange ratio cutoff or a rating of perceived exertion ≥7 (scale 1–10).

Percentage of Predicted Equations

Of the 45 studies, 21 (47%) described how their $\%_{Pred}$ value was calculated, either by providing the equation or a reference. Thus, 53% of studies did not report the NRVs used. There were 12 different NRVs applied across the 21 studies (Table S3). The top 3 most used NRVs were produced by Wasserman,[13] Cooper and Weiler-Ravell,[34] and Jones,[35] with 12, 7, and 4 citations, respectively.

Some studies used >1 NRV, meaning there were 30 occasions across 21 studies where NRVs had been applied to ConHD populations (ie, use of both a pediatric NRV and an adult NRV in a combined adult–pediatric study). The RoB analysis revealed 7 (23%) instances where the application of NRVs to study populations were judged to have a low RoB, 19 (63%) instances where there were some concerns, and 4 (13%), instances where it was judged to have a high

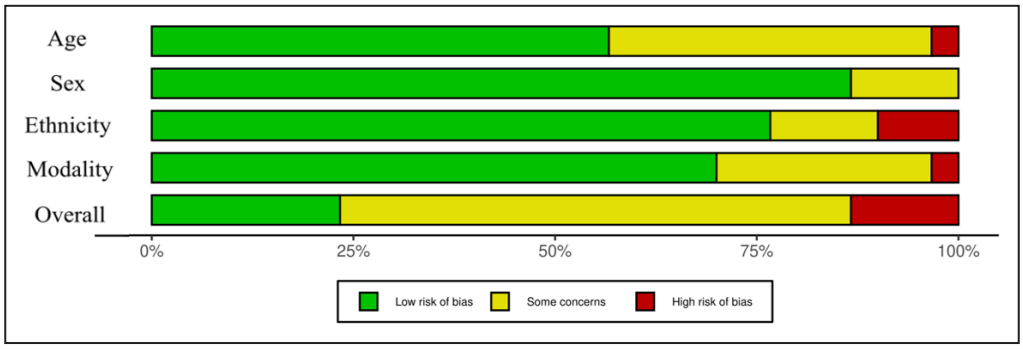

Figure 2. RoB across domains of age, sex, ethnicity, modality, and overall RoB.
RoB was calculated for studies that explicitly state the NRV used and is presented as a percentage. Color scheme: red=inappropriate matching/high RoB; yellow=partial or unclear matching/moderate RoB; Green=appropriate matching/low RoB. NRV indicates normative reference value; and RoB, risk of bias.

RoB (Figure 2). Most studies were judged as having some concerns due to partial matching of the study population and modality to that of the NRVs. An additional contribution to the high proportion of studies that achieved some concerns was that some NRVs do not describe the population or the modality that the NRV was derived from. The RoB results for each individual study are illustrated in Figure S1.

NRVs Applied to Fontan Populations

Twenty-seven studies included people with a Fontan circulation and expressed peak $\dot{V}O_2$ as a $\%_{Pred}$, of which 9 (33%) studies did not report which NRVs they used. Of the studies that did report the NRV (n=18; 66%), there were 26 citations of NRVs. There were 10 (38%) citations of Wasserman (adult),[13] 5 (19%) citations of Cooper and Weiler-Ravell (pediatric),[34] 4 citations of Jones (adolescent–adult; 1997, n=2; 1989, n=2; 15%),[36,37] and 7 citations of other individual NRVs cited once each.[38–44]

Aim 3: Validation

The Wasserman NRV was applied to 70 adults with a Fontan circulation. The NRV was partially matched for age and fully matched for sex and modality. Directly measured absolute peak $\dot{V}O_2$ in the Fontan participants was 1547±527 mL/min^{-1} (Table S4), which corresponded to 47.6±14.9 $\%_{Pred}$ (3290±609 mL/min^{-1}). Measured and predicted peak $\dot{V}O_2$ were not significantly associated for men (n=29; b=0.31; R^2<0.01; P=0.75) or women (n=41; b=0.07; R^2=0.02; P=0.39) (Figure 3). Individual and group level agreement between

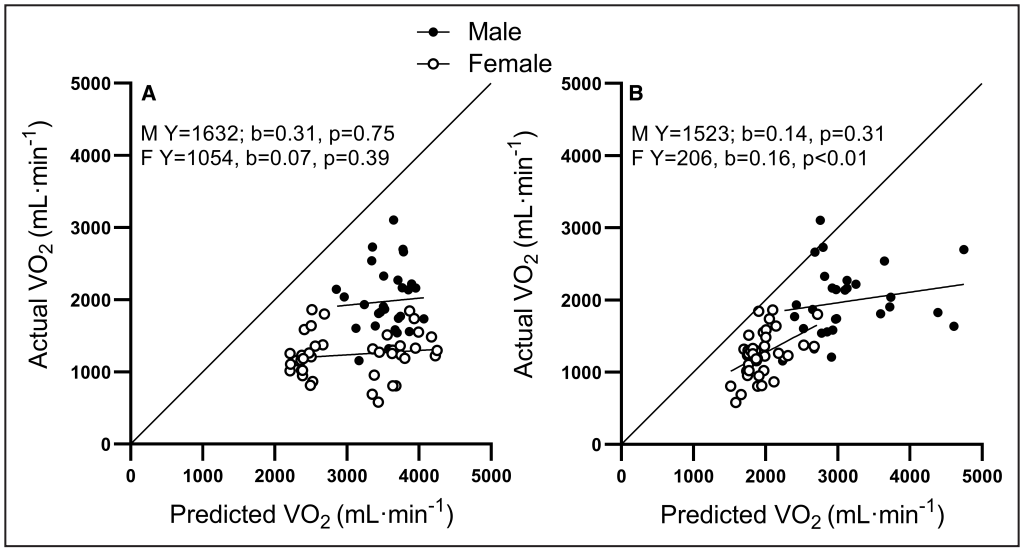

Figure 3. Validation of Wasserman and Hansen NRVs against measured peak $\dot{V}O_2$ in participants with a Fontan circulation.
Solid lines in (**A**) and (**B**) are the lines of identity. **A**, Predicted peak $\dot{V}O_2$ was calculated using all proposed steps (Table). **B**, Predicted peak $\dot{V}O_2$ was calculated using measured mass. F indicates female; M, male; NRV, normative reference value; and peak $\dot{V}O_2$, peak oxygen consumption.

measured and predicted peak $\dot{V}O_2$ are reported in Data S2, Figure S2, and Table S5.

Sensitivity Analysis

A sensitivity analysis using participants measured mass was performed. Participants peak $\dot{V}O_2$ was 64.4±15.5 $\%_{Pred}$ (2428±735 mL/min^{-1}). This is an increase of 16.7±14.9% $_{Pred}$, and by 9.2±12.3 $\%_{Pred}$ and 22.1±14.4 $\%_{Pred}$ for men and women, respectively, compared with the full method (Table). The sensitivity analysis reported no significant association between measured and predicted peak $\dot{V}O_2$ for men, but there was a significantly positive association for women (n=41; b=0.16; R^2=0.21; P<0.01) (Figure 3).

DISCUSSION

The current research reports for the first time that a range of NRVs (n=12) are used across the prognostic ConHD literature, although the most cited NRVs were from Wasserman[13] and Cooper and Weiler-Ravell,[34] being cited 12 and 7 times each for adult and pediatric populations, respectively. However, most concerningly, 53% of studies in the ConHD literature did not state the NRVs they used to normalize peak $\dot{V}O_2$ as a $\%_{Pred}$. This insufficient reporting limits the reader's ability to interpret results with confidence or replicate study findings. It was also found that NRVs are often applied inappropriately, due to either no or partial matching between the NRV derivation cohort and study participants and protocols. The acceptance and uncritical use of potentially inappropriate NRVs (ie, NRVs that do not match study participants or protocols) may result in inaccurate predictions that negatively influence clinical decision making. Finally, when the recommended Wasserman NRV was applied to a cohort of Fontan participants, the NRVs showed poor validity due to a lack of association between measured and predicted values. These data seriously question current clinical practice due to poor reporting, the uncritical use of NRVs, and the lack of concordance between recommended NRVs and clinical populations.

Aims 1 and 2: Scoping Review and Risk of Bias

The Wasserman NRV was the most cited adult equation, likely due to its recommendations by several scientific organizations including the AHA.[11,12] Performing backward citation searches reveals this particular NRV was itself amalgamated from the work of Bruce et al in 1973[45] and Hansen et al in 1984,[46] and is not a novel NRV itself. Moreover, Bruce et al's NRV was treadmill based in a mixed sex cohort (n=295; 48.6±11.1 years), whereas Hansen et al's NRV was cycle ergometry based in men only (n=77; 54.3±9.2 years). There were no data on women undergoing cycle ergometry in these equations, yet they are purported to be used in both sexes regardless of exercise modality. Therefore, it could be argued that using this NRV for women may lead to misleading results and therefore constitutes substandard clinical practice. Furthermore, no standard error statistics for this NRV can be identified, either from the original work or from textbooks recommending its use,[13,45–47] meaning readers are unable to confidently evaluate this NRV's performance. Previous systematic reviews have identified poor reporting of statistical procedures and performance (ie assumption checking, calibration, validation) across the NRV literature,[8,9] and this current scoping review reinforces this paucity of statistically robust data.

The second most cited NRV was produced by Cooper and Weiler-Ravell,[34] having been cited 7 times. This NRV was derived in 1984 from a pediatric cohort of 109 children (≈47% girls), aged 12±3 years (range, 6–17 years), performing cycle ergometry. The NRV predicts peak $\dot{V}O_2$ in mL/min^{-1} from stature and sex, the standard error of the estimate of the boy's equation is ~0.4 L/min^{-1} and ≈0.3 for girls. Therefore, if a boy has a stature of 160 cm, he would be predicted to have a peak $\dot{V}O_2$ of 2.4 L/min^{-1} with a 95% CI of 1.7 to 3.2 L/min^{-1}, a prediction interval of 1.5 L/min^{-1} (≈62% error), which is an inappropriate level of error. The median year of publication for studies that cited the Cooper and Weiler-Ravell NRV was 2016 (range, 2013–2018), and it is unclear why this NRV is commonly used in the recent literature. There are NRVs that have been developed more recently, with larger sample sizes and smaller standard error,[9,48] and therefore clinicians and researchers should be encouraged to use NRVs that are methodologically and statistically robust and justify their adopted approach.

Interestingly, a recent multicenter and multicontinent collaboration (France, Germany, and the United States) produced a rigorously derived NRV from 909 healthy children and adolescents (including underweight and obese participants), which has been reported to be statistically superior to the Cooper and Weiler-Ravell NRV.[49] Furthermore, when applied to young people with ConHD, the new (Z score based) NRV was able to discriminate individuals with ConHD versus healthy controls significantly better compared with the Cooper and Weiler-Ravell height- and weight-based NRV.[50] The methodological robustness and statistical reliability and validity of this new NRV is encouraging,[50] and its use should certainly be considered in future studies.

Studies included in the current review were often unable to apply NRVs appropriately. For example, the RoB analysis reported instances where study authors were applying NRVs derived from cycle ergometry to

treadmill data, adult NRVs to pediatric data, and NRVs derived from White to Asian populations, all of which will likely result in erroneous predictions potentially influencing patient care. Moreover, dated NRVs with poorly reported methods are being applied to clinical populations for which they were not intended to be used in. The current authors are not criticizing the original NRVs per se, nor are they discrediting the findings made by those studies that have used potentially inappropriate NRVs, but there is clearly an issue in the literature with the misapplication of NRVs.

There have been serious concerns raised by clinicians over the application of such NRVs to current-day clinical cases,[20] where different NRVs lead to differences in patient management. From a ConHD disease perspective (a perspective likely shared by other pediatric chronic diseases), pediatric patients are transitioned into adult cardiology services, and often NRVs are developed in distinct pediatric,[34,48,49] adolescent to adult,[19] and adult[13,45,46] cohorts. Therefore, 1 NRV is rarely used to monitor the changes in fitness over time, which reduces the confidence in whether any change to the $\%_{Pred}$ value is due to a change in NRV (due to transitioning into adult care) or a true decline in fitness. Thus, a potential research gap is to produce NRVs that cover the lifespan, akin to the global lung function equations for spirometry.[51]

Aim 3: Validation

The current study sought to apply the commonly recommended and cited Wasserman NRV[13] to an appropriately matched (ie, age, sex, exercise modality) cohort of adult Fontan patients. There are several iterations of this NRV, and the most recommended[10,12] and most cited version was chosen (Table). The application of the Wasserman NRV to patients with a Fontan circulation appears invalid, because predicted and measured peak $\dot{V}O_2$ were not significantly associated for men or women (Figure 3). The β coefficients reported were ≈0.3 for men and ≈0.1 for women, meaning for every milliliter increase in predicted peak $\dot{V}O_2$, measured peak $\dot{V}O_2$ failed to increase at the same rate. Thus, the error in the prediction increases as the predicted peak $\dot{V}O_2$ increases. A sensitivity analysis using measured weight did not improve the NRV's performance for men but did for women; however, further validation is required in larger and more diverse samples.

These data are concerning and potentially dangerous, because an erroneous peak $\dot{V}O_2$ value from a prediction equation can influence patient management and clinical decisions. Recently, the European Respiratory Society (TF-2021–09) has set up a task force[52] to create a global NRV for peak $\dot{V}O_2$, which once published should improve the reporting and application of NRVs to clinical populations. This project will likely have several challenges (eg, standardization of exercise protocols, geographical limitations, data sharing) that may delay such endeavors. Therefore, in the interim, there are several options to improve current practices: (1) A plethora of NRVs based from healthy populations are available,[8,9,49] and at a minimum, practitioners should select NRVs that match patient demographics (ie, age, sex), exercise modalities, and have a small error of the estimate. (2) Practitioners/researchers should externally validate existing NRVs to check their suitability for their patient cohorts.[50] (3) Disease-specific NRVs should be developed through collaboration; however, with the clinical heterogeneity of ConHD and relatively small numbers, these will be challenging to produce. (4) Where possible, longitudinally test individuals and use them as their own control. (5) If possible, cease the use of NRVs and report peak $\dot{V}O_2$ scaled to body mass or preferably fat-free mass.[5,7] (6) Meticulously report interpretation practices. To aid future reporting practices, the authors of the current study have produced a checklist (Table S6), which is anticipated to improve the transparency and reproducibility of future CPET research.

Study Considerations

There are several strengths to the current research, including simultaneously combining a rapid systematic scoping review to identify current research practice in congenital cardiology, and then assesses current practice using a large cohort of people who have complex ConHD (Fontan circulation). A further strength of this research is that it uses data from the Australian and New Zealand Fontan Registry, which has a large volume of Fontan exercise data that has been robustly collected. However, it is also acknowledged that this review was focused on articles that use fitness data to predict outcomes in ConHD; thus, a limitation will mean not every research article that measured peak $\dot{V}O_2$ in people with ConHD would have been included within the analyses. Another consideration is the finding of this study on the applicability of the Wasserman NRV to people with a Fontan circulation may not apply to other subsets of ConHD. Furthermore, although the Wasserman NRV may not be valid for predicting peak $\dot{V}O_2$ in complex ConHD, it does have prognostic usefulness in people with heart failure or a Fontan circulation.[53,54]

CONCLUSIONS

Fifty-three percent of studies that report peak $\dot{V}O_2$ $\%_{Pred}$ fail to cite the NRV that the authors used in the ConHD prognosis literature. There are concerns over the applicability of poorly reported and developed

NRVs being applied in clinical practice. Furthermore, applying the recommended and most-cited NRV to people with a Fontan circulation resulted in poor validity, which when used clinically could potentially impact clinical decision making. Thus, the use of NRVs should either be ceased, or if used, meticulously selected and reported.

Recommendations for Practice and Research

- If using NRVs in a clinical or research setting, explicitly report the NRV used and ensure it is matched to participants demographics and exercise modality. Furthermore, ensure NRVs used have been rigorously and robustly developed and validated.
- If no such NRVs are available, then do not use $\%_{pred}$ and retain the use of traditional expressions of peak $\dot{V}O_2$ (ie, mL/min^{-1}, mL/kg^{-1} per min^{-1})
- As recommended by the European Society of Cardiology, people with a Fontan circulation should have their cardiorespiratory fitness monitored longitudinally, preferably using peak $\dot{V}O_2$ expressed in absolute terms and relative to fat-free mass.[5,25]
- Future research should aim to validate robustly developed NRVs in clinical populations and develop disease-specific NRVs.
- A checklist has been provided to aid the reporting of NRVs in CPET testing.

ARTICLE INFORMATION

Received January 18, 2024; accepted June 5, 2024.

Affiliations

Department of Public Health and Sport Science, Children's Health & Exercise Research Centre (CHERC), Faculty of Health and Life Sciences (Curtis A. Wadey, O.W.T., A.R.B., Craig A. Williams) and Department of Clinical and Biomedical Sciences, Faculty of Health and Life Sciences (O.W.T.), University of Exeter, United Kingdom, Bristol Congenital Heart Centre, The Bristol Heart Institute, University Hospitals Bristol NHS Foundation Trust, Bristol, United Kingdom (A.G.S.); Central Clinical School, University of Sydney, Camperdown, New South Wales, Australia (D.L.T., K.L., J.A., R.C.); Department of Cardiology, Royal Prince Alfred Hospital, Camperdown, New South Wales, Australia (D.L.T., K.L., J.A., R.C.); Heart Research Institute, Charles Perkins Centre, The University of Sydney, Camperdown, New South Wales, Australia (D.L.T., K.L., J.A., R.C.); Department of Cardiology, Royal Children's Hospital, Melbourne, Australia (R.G.W.); and Department of Paediatrics, Melbourne University, Melbourne, Australia (R.G.W.).

Sources of Funding

C.A. Wadey was funded by an industrial PhD scholarship by Canon Medical Systems UK Ltd. and the University of Exeter. D.L.T. is funded by the Medical Research Future Fund–Cardiovascular Health Mission–Congenital Heart Disease Grant (ARGCHDG000016).

Disclosures

None.

Supplemental Material

Data S1–S2
Tables S1–S6
Figures S1–S2

REFERENCES

1. Taylor H, Buskirk E, Henschel A. Maximal oxygen intake as an objective measure of cardio-respiratory performance. *J Appl Physiol*. 1955;8:73–80. doi: 10.1152/jappl.1955.8.1.73
2. Villaseca-Rojas Y, Varela-Melo J, Torres-Castro R, Vasconcello-Castillo L, Mazzucco G, Vilaró J, Blanco I. Exercise capacity in children and adolescents with congenital heart disease: a systematic review and meta-analysis. *Front Cardiovasc Med*. 2022;9:1–16. doi: 10.3389/fcvm.2022.874700
3. Wadey C, Weston M, Dorobantu D, Pieles G, Stuart G, Barker A, Taylor R, Williams C. The role of cardiopulmonary exercise testing in predicting mortality and morbidity in people with congenital heart disease: a systematic review and meta-analysis. *Eur J Prev Cardiol*. 2021;29:513–533. doi: 10.1093/eurjpc/zwab125
4. Welsman J, Armstrong N. Statistical techniques for interpreting body size-related exercise performance during growth. *Pediatr Exerc Sci*. 2000;12:513–533. doi: 10.1123/pes.12.2.112
5. Wadey CA, Barker AR, Stuart G, Tran DL, Laohachai K, Ayer J, Cordina R, Williams CA. Scaling peak oxygen consumption for body size and composition in people with a Fontan circulation. *J Am Heart Assoc*. 2022;11:1–8. doi: 10.1161/JAHA.122.026181
6. Lolli L, Batterham AM, Weston KL, Atkinson G. Size exponents for scaling maximal oxygen uptake in over 6500 humans: a systematic review and meta-analysis. *Sports Med*. 2017;47:1405–1419. doi: 10.1007/s40279-016-0655-1
7. Loftin M, Sothern M, Abe T, Bonis M. Expression of VO2peak in children and youth, with special reference to allometric scaling. *Sports Med*. 2016;46:1451–1460. doi: 10.1007/s40279-016-0536-7
8. Blais S, Berbari J, Counil FP, Dallaire F. A systematic review of reference values in pediatric cardiopulmonary exercise testing. *Pediatr Cardiol*. 2015;36:1553–1564. doi: 10.1007/s00246-015-1205-6
9. Takken T, Mylius CF, Paap D, Broeders W, Hulzebos HJ, Van Brussel M, Bongers BC. Reference values for cardiopulmonary exercise testing in healthy subjects–an updated systematic review. *Expert Rev Cardiovasc Ther*. 2019;17:413–426. doi: 10.1080/14779072.2019.1627874
10. Weisman IM, Weisman IM, Marciniuk D, Martinez FJ, Sciurba F, Sue D, Myers J, Casaburi R, Marciniuk D, Beck K, et al. ATS/ACCP statement on cardiopulmonary exercise testing. *Am J Respir Crit Care Med*. 2003;167:211–277. doi: 10.1164/ajrccm.167.10.952
11. Guazzi M, Adams V, Conraads V, Halle M, Mezzani A, Vanhees L, Arena R, Fletcher GF, Forman DE, Kitzman DW, et al. Clinical recommendations for cardiopulmonary exercise testing data assessment in specific patient populations. *Circulation*. 2012;126:2261–2274. doi: 10.1161/CIR.0b013e31826fb946
12. Guazzi M, Adams V, Conraads V, Halle M, Mezzani A, Vanhees L, Arena R, Fletcher GF, Forman DE, Kitzman DW, et al. EACPR, AHA clinical recommendations for cardiopulmonary exercise testing data assessment in specific patient populations. *Eur Heart J*. 2012;33:2917–2927. doi: 10.1093/eurheartj/ehs221
13. Wasserman K, Hansen J, Sue D, Stringer W, Whipp B. Normal values. *Principles of Exercise Testing and Interpretation*. 4th ed. Lippincott Williams & Wilkin; 2005:160–182.
14. Pritchard A, Burns P, Correia J, Jamieson P, Moxon P, Purvis J, Thomas M, Tighe H, Sylvester KP. ARTP statement on cardiopulmonary exercise testing 2021. *BMJ Open Respir Res*. 2021;8:1–26. doi: 10.1136/bmjresp-2021-001121
15. Radtke T, Crook S, Kaltsakas G, Louvaris Z, Berton D, Urquhart DS, Kampouras A, Rabinovich RA, Verges S, Kontopidis D, et al. ERS statement on standardisation of cardiopulmonary exercise testing in chronic lung diseases. *Eur Respir Rev*. 2019;28:1–25. doi: 10.1183/16000617.0101-2018
16. Hebestreit H, Arets HGM, Aurora P, Boas S, Cerny F, Hulzebos EHJ, Karila C, Lands LC, Lowman JD, Swisher A, et al. Statement on exercise testing in cystic fibrosis. *Respiration*. 2015;90:332–351. doi: 10.1159/000439057
17. Orenstein D. Assessment of exercise pulmonary function. In: Rowland T, ed *Pediatric Laboratory Exercise Testing Clinical Guidelines*. Champaign: Human Kinetics; 1993:141–163.
18. Werkman M, Hulzebos E, Helders P, Arets B, Takken T. Estimating peak oxygen uptake in adolescents with cystic fibrosis. *Arch Dis Child*. 2014;99:21–25. doi: 10.1136/archdischild-2012-303439
19. Jones NL, Makrides L, Hitchcock C, Chypchar T, McCartney N. Normal standards for an incremental progressive cycle ergometer test. *Am Rev Respir Dis*. 1985;131:700–708. doi: 10.1164/arrd.1985.131.5.700

20. Waterfall JL, Burns P, Shackell D, Pepke-Zaba J, Oates KE, Sylvester KP. The risks of applying normative values in paediatric cardiopulmonary exercise testing: a case report. *ERJ Open Res*. 2020;6:1–3. doi: 10.1183/23120541.00333-2020
21. Hütler M, Schnabel D, Staab D, Tacke A, Wahn U, Böning D, Beneke R. Effect of growth hormone on exercise tolerance in children with cystic fibrosis. *Med Sci Sports Exerc*. 2002;34:567–572. doi: 10.1249/00005768-200204000-00001
22. Armstrong N, Welsman J. Development of peak oxygen uptake from 11–16 years determined using both treadmill and cycle ergometry. *Eur J Appl Physiol*. 2019;119:801–812. doi: 10.1007/s00421-019-04071-3
23. Peterman JE, Harber MP, Imboden MT, Whaley MH, Fleenor BS, Myers J, Arena R, Finch WH, Kaminsky LA. Accuracy of nonexercise prediction equations for assessing longitudinal changes to cardiorespiratory fitness in apparently healthy adults: ball st cohort. *J Am Heart Assoc*. 2020;9:1–22. doi: 10.1161/JAHA.119.015117
24. Malek MH, Berger DE, Housh TJ, Coburn JW, Beck TW. Validity of $\dot{V}O_{2max}$ equations for aerobically trained males and females. *Med Sci Sports Exerc*. 2004;36:1427–1432. doi: 10.1249/01.MSS.0000135795.60449.CE
25. Baumgartner H, De Backer J, Babu-Narayan SV, Budts W, Chessa M, Diller G-P, Lung B, Kluin J, Lang IM, Meijboom F, et al. 2020 ESC guidelines for the management of adult congenital heart disease. *Eur Heart J*. 2021;42:563–646. doi: 10.1093/eurheartj/ehaa554
26. Tran D, D'Ambrosio P, Verrall CE, Attard C, Briody J, D'Souza M, Fiatarone Singh M, Ayer J, D'Udekem Y, Twigg S, et al. Body composition in young adults living with a Fontan circulation: the Myopenic profile. *J Am Heart Assoc*. 2020;9:e015639. doi: 10.1161/JAHA.119.015639
27. Iyengar AJ, Winlaw DS, Galati JC, Gentles TL, Weintraub RG, Justo RN, Wheaton GR, Bullock A, Celermajer DS, d'Udekem Y. The Australia and New Zealand fontan registry: description and initial results from the first population-based fontan registry. *Intern Med J*. 2014;44:148–155. doi: 10.1111/imj.12318
28. Attard C, Monagle P, D'Udekem Y, Mackay M, Briody J, Cordina R, Hassan E, Simm P, Rice K, Ignjatovic V, et al. Long-term outcomes of warfarin versus aspirin after Fontan surgery. *J Thorac Cardiovasc Surg*. 2021;162:1218–1228. doi: 10.1016/j.jtcvs.2020.12.102
29. Tomlinson OW, Wadey CA, Williams CA. Normal reference values for aerobic fitness in cystic fibrosis: a scoping review. *BMJ Open Sport Exerc Med*. 2022;8:e001490. doi: 10.1136/bmjsem-2022-001490
30. Barron A, Dhutia N, Mayet J, Hughes AD, Francis DP, Wensel R. Test-retest repeatability of cardiopulmonary exercise test variables in patients with cardiac or respiratory disease. *Eur J Prev Cardiol*. 2014;21:445–453. doi: 10.1177/2047487313518474
31. Cordina R, OMeagher S, Karmali A, Rae CL, Liess C, Kemp GJ, Puranik R, Singh N, Celermajer DS. Resistance training improves cardiac output, exercise capacity and tolerance to positive airway pressure in Fontan physiology. *Int J Cardiol*. 2013;168:780–788. doi: 10.1016/j.ijcard.2012.10.012
32. Powell AW, Wittekind SG, Alsaied T, Lubert AM, Chin C, Veldtman GR, Cordina R, Katz DA, Mays WA, Knecht SK, et al. Body composition and exercise performance in youth with a fontan circulation: a bioimpedance based study. *J Am Heart Assoc*. 2020;9:1–10. doi: 10.1161/JAHA.120.018345
33. Cohen J. A power primer. *Psychol Bull*. 1992;112:155–159. doi: 10.1037/0033-2909.112.1.155
34. Cooper D, Weiler Ravell D. Gas exchange response to exercise in children. *Am Rev Respir Dis*. 1984;129:S47–S48. doi: 10.1164/arrd.1984.129.2p2.s47
35. Jones NL. *Clinical Exercise Testing*. 3rd ed.; 1988. Philadelphia: W B Saunders Co Ltd
36. Jones NL, Summers E, Killian KJ. Influence of age and stature on exercise capacity during incremental cycle ergometry in men and women. *Am Rev Respir Dis*. 1989;140:1373–1380. doi: 10.1164/ajrccm/140.5.1373
37. Jones NL. *Clinical Exercise Testing*. 4th ed. Saunders; 1997.
38. Gläser S, Koch B, Ittermann T, Schäper C, Dörr M, Felix S, Völzke H, Ewert R, Hansen JE. 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. doi: 10.1097/HJR.0b013e328336a124
39. Rhodes J. Exercise testing. In: Keane J, Lock J, Fyler D, eds *Nadas' Pediatric Cardiology*. 2nd ed. Saunders; 2006:275–287. doi: 10.1016/B978-1-4160-2390-6.50021-0
40. Ten HADJ, Takken T, Van Osch-Gevers M, Helbing W a. Normal values for cardiopulmonary exercise testing in children. *Eur J Prev Cardiol*. 2011;18:48–54. doi: 10.1097/HJR.0b013e32833cca4d
41. Bongers BC, Hulzebos EHJ, van Brussel M, Takken T. Pediatric norms for cardiopulmonary exercise testing. *Relation to Gender and Age*. 1st ed. 's-Herogenbosch, Netherlands: Uitgeverij BOXPress; 2014.
42. Ohuchi H. Exercise testing in pediatric and adult patients with and without congenital heart disease. *Pediatr Cardiol Card Surg*. 2014;30:635–645. doi: 10.9794/jspccs.30.635
43. Kaminsky LA, Arena R, Myers J. Reference standards for cardiorespiratory fitness measured with cardiopulmonary exercise testing data from the fitness registry and the importance of exercise national database. *Mayo Clin Proc*. 2015;90:1515–1523. doi: 10.1016/j.mayocp.2015.07.026
44. Kaminsky LA, Imboden MT, Arena R, Myers J. Reference standards for cardiorespiratory fitness measured with cardiopulmonary exercise testing using cycle ergometry: data from the fitness registry and the importance of exercise National Database (FRIEND) registry. *Mayo Clin Proc*. 2017;92:228–233. doi: 10.1016/j.mayocp.2016.10.003
45. Bruce R, Kusumi F, Hosmer D. Fundamentals of clinical cardiology and recognition. *Fundam Clin Cardiol*. 1974;88:372–379.
46. Hansen JE, Sue DY, Wasserman K. Predicted values for clinical exercise testing. *Am Rev Respir Dis*. 1984;129:S49–S55. doi: 10.1164/arrd.1984.129.2p2.s49
47. Wasserman K, Hansen J, Sue D, Stringer W, Sietsema K, Sun X, Whipp B. Normal values. In: Hansen J, Sue D, Stringer W, Sietsema K, Sun X, Whipp B, eds *Principles of Exercise Testing and Interpretation Including Pathophysiology and Clinical Applications*. Vol 24. 5th ed. Lippincott Williams & Wilkins; 2012:311. doi: 10.3109/02770908709070958
48. Blanchard J, Blais S, Chetaille P, Bisson M, Counil FP, Huard-Girard T, Berbari J, Boulay P, Dallaire F. New reference values for cardiopulmonary exercise testing in children. *Med Sci Sports Exerc*. 2018;50:1125–1133. doi: 10.1249/MSS.0000000000001559
49. Gavotto A, Mura T, Rhodes J, Yin S, Hager A, Hock J, Guillaumont S, Vincenti M, De La Villeon G, Requirand A, et al. Reference values of aerobic fitness in the contemporary paediatric population. *Eur J Prev Cardiol*. 2023;30:820–829. doi: 10.1093/eurjpc/zwad054
50. Amedro P, Mura T, Matecki S, Guillaumont S, Requirand A, Jeandel C, Kollen L, Gavotto A. Use of new paediatric VO_{2max} reference equations to evaluate aerobic fitness in overweight or obese children with congenital heart disease. *Eur J Prev Cardiol*. 2023;30:1–9.
51. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, Enright PL, Hankinson JL, Ip MSM, Zheng J, 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. doi: 10.1183/09031936.00080312
52. Sylvester K, Hebestreit H. Developing Global Lung Function Initiative (GLI) reference equations for cardiopulmonary exercise testing (CPET) – an ERS Technical Standard. Accessed February 16, 2023. https://www.ersnet.org/science-and-research/ongoing-task-forces/.
53. Diller GP, Giardini A, Dimopoulos K, Gargiulo G, Müller J, Derrick G, Giannakoulas G, Khambadkone S, Lammers AE, Picchio FM, et al. Predictors of morbidity and mortality in contemporary Fontan patients: results from a multicenter study including cardiopulmonary exercise testing in 321 patients. *Eur Heart J*. 2010;31:3073–3083. doi: 10.1093/eurheartj/ehq356
54. Myers J, De Souza E, Silva CG, Arena R, Kaminsky L, Christle JW, Busque V, Ashley E, Moneghetti K. Comparison of the FRIEND and Wasserman-Hansen equations in predicting outcomes in heart failure. *J Am Heart Assoc*. 2021;10:1–12. doi: 10.1161/JAHA.121.021246