

Confirming the attainment of maximal oxygen uptake within special and clinical groups: a systematic review and meta-analysis of cardiopulmonary exercise test and verification phase protocols

Short title: Verification phase testing in special and clinical groups

Victor A. B. Costa^{1,2}, Adrian W. Midgley^{3*}, Julia K. Baumgart⁴, Sean Carroll⁵, Todd A. Astorino⁶, Gustavo Z. Schaun⁷, Guilherme F. Fonseca^{1,2}, and Felipe A. Cunha^{1,2}

- ¹ Graduate Program in Exercise Science and Sports, University of Rio de Janeiro State, Rio de Janeiro, Brazil.
- ² Laboratory of Physical Activity and Health Promotion, University of Rio de Janeiro State, Rio de Janeiro, Brazil.
- ³ Department of Sport and Physical Activity, Edge Hill University, Ormskirk, England.
- ⁴ Centre for Elite Sports Research, Department of Neuromedicine and Movement Science, Norway, University of Science and Technology, Trondheim, Norway.
- ⁵ School of Sport, Exercise and Rehabilitation Sciences, University of Hull, Hull, England.
- ⁶ Department of Kinesiology, California State University, San Marcos, CA, USA.
- ⁷ Centre for Sport Science and University Sports, University of Vienna, Vienna, Austria.

* Corresponding author

E-mail: Midglead@edgehill.ac.uk (AWM)

Trial registration

PROSPERO (CRD42021247658). <https://www.crd.york.ac.uk/prospero>

Abstract

Background and aim

A plateau in oxygen uptake ($\dot{V}O_2$) during an incremental cardiopulmonary exercise test (CPET) to volitional exhaustion appears less likely to occur in special and clinical populations. Secondary maximal oxygen uptake ($\dot{V}O_{2max}$) criteria have been shown to commonly underestimate the actual $\dot{V}O_{2max}$. The verification phase protocol might determine the occurrence of 'true' $\dot{V}O_{2max}$ in these populations. The primary aim of the current study was to systematically review and provide a meta-analysis on the suitability of the verification phase for confirming 'true' $\dot{V}O_{2max}$ in special and clinical groups. Secondary aims were to explore the applicability of the verification phase according to specific participant characteristics and investigate which test protocols and procedures minimise the differences between the highest $\dot{V}O_2$ values attained in the CPET and verification phase.

Methods

Electronic databases (PubMed, Web of Science, SPORTDiscus, Scopus, and EMBASE) were searched using specific search strategies and relevant data were extracted from primary studies. Studies meeting inclusion criteria were systematically reviewed. Meta-analysis techniques were applied to quantify weighted mean differences (standard deviations) in peak $\dot{V}O_2$ from a CPET and a verification phase within study groups using random-effects models. Subgroup analyses investigated the differences in $\dot{V}O_{2max}$ according to individual characteristics and test protocols. The methodological quality of the included primary studies was assessed using a modified Downs and Black checklist to obtain a level of evidence. Participant-level $\dot{V}O_2$ data were analysed according to the threshold criteria reported by the studies or the inherent measurement error of the metabolic analysers and displayed as Bland-Altman plots.

Results

Forty-three studies were included in the systematic review, whilst 30 presented quantitative information for meta-analysis. Within the 30 studies, the highest mean $\dot{V}O_2$ values attained in the CPET and verification phase protocols were similar (mean difference = -0.00 [95% confidence intervals, CI = -0.03 to 0.03] L·min⁻¹, $p = 0.87$; level of evidence, LoE: strong). The specific clinical groups with sufficient primary studies to be meta-analysed showed a similar $\dot{V}O_{2max}$ between the CPET and verification phase ($p > 0.05$, LoE: limited to strong). Across all 30 studies, $\dot{V}O_{2max}$ was not affected by differences in test protocols ($p > 0.05$; LoE: moderate to strong). Only 23 (53.5%) of the 43 reviewed studies reported how many participants achieved a lower, equal, or higher $\dot{V}O_2$ value in the verification phase versus the CPET or reported or supplied participant-level $\dot{V}O_2$ data for this information to be obtained. The percentage of participants that achieved a lower, equal, or higher $\dot{V}O_2$ value in the verification phase was highly variable across studies (e.g. the percentage that achieved a higher $\dot{V}O_2$ in the verification phase ranged from 0% to 88.9%).

Conclusion

Group-level verification phase data appear useful for confirming a specific CPET protocol likely elicited $\dot{V}O_{2max}$, or a reproducible $\dot{V}O_{2peak}$, for a given special or clinical group. Participant-level data might be useful for confirming whether specific participants have likely elicited $\dot{V}O_{2max}$, or a reproducible $\dot{V}O_{2peak}$, however, more research reporting participant-level data is required before evidence-based guidelines can be given.

Introduction

Maximal oxygen uptake ($\dot{V}O_{2\max}$) represents the upper physiological limit of utilising oxygen to produce energy during volitional exercise to exhaustion [1]. The original concept emerged in the 1920's in the seminal works of Hill and colleagues [2, 3]. These authors described this phenomenon as the “ceiling” of oxygen uptake ($\dot{V}O_2$) during a discontinuous step-incremented exercise test, beyond which no additional increase in $\dot{V}O_2$ is observed despite an increase in work rate. In special groups such as apparently healthy children and older adults, and clinical groups such as people with chronic respiratory and metabolic conditions, $\dot{V}O_{2\max}$ testing is increasingly recommended and typically determined using an incremental cardiopulmonary exercise test (CPET) with concurrent recording of electrocardiography, blood pressure, and oxyhemoglobin saturation [4]. Applications of CPET include facilitating the diagnosis and evaluating the physiological impact of cardiopulmonary disease [5], evaluating fitness for major surgery [4], deciding on the appropriateness of cardiac transplantation [6], investigating unexplained dyspnea and exercise intolerance [4], prescribing exercise to special and clinical groups [7, 8], and evaluating the chronic effects of exercise training programs and other interventions to promote health [9]. Over 40 outcome variables can be derived from a CPET [10], however, $\dot{V}O_{2\max}$ is widely regarded as the most important, since it is considered the gold-standard measure of cardiorespiratory fitness [4]. The importance of cardiorespiratory fitness for health and longevity is highlighted by the scientific statement from the American Heart Association, which calls for cardiorespiratory fitness to be recognised as a “clinical vital sign” [11].

Different ergometers, test protocols, procedures, and criteria have been developed and applied to promote accurate $\dot{V}O_{2\max}$ determination, which is significant considering the relationship between $\dot{V}O_{2\max}$ and health status. Potential reasons for not obtaining valid $\dot{V}O_{2\max}$ values include inappropriate test protocols and poor effort from participants [12]. However, there is no consensus on the best approach to establish whether a ‘true’ $\dot{V}O_{2\max}$ has been attained. The primary criterion for verifying that $\dot{V}O_{2\max}$ has been attained has been based on establishing a $\dot{V}O_2$ plateau at volitional exhaustion,

first proposed in 1955 by Taylor, Buskirk and Henschel [13]. They defined the plateau occurrence as an increase in $\dot{V}O_2$ of less than $150 \text{ mL}\cdot\text{min}^{-1}$ (or $\leq 2.1 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$) during a discontinuous step-incremented protocol incorporating 3-5 visits to the laboratory. With advances in technology (specifically, gas analysers that record breath-by-breath pulmonary gas exchange data and electronically programmable treadmills and cycle ergometers), discontinuous step-incremented tests protocols have largely been replaced by more time-efficient continuous ramp and pseudo-ramp protocols [14-17]. Ramp protocols typically elicit an accelerated or linear $\dot{V}O_2$ response during the final portion of the test, which reduces plateau incidence, especially in low fit individuals [18], although a high level of variability has been observed between studies [19-21]. This variability can be explained by the use of different $\dot{V}O_2$ plateau criterion thresholds [22], $\dot{V}O_2$ sampling intervals [23], test protocols [24], and variation in participant characteristics [25]. This inherent dependency on the test protocol and procedures, and participant characteristics, reduces the robustness of the $\dot{V}O_2$ plateau concept for confirming $\dot{V}O_{2\text{max}}$ attainment.

Numerous investigators have not applied a $\dot{V}O_2$ plateau criterion or have relied upon so-called secondary criteria to assess whether $\dot{V}O_{2\text{max}}$ has been elicited if a $\dot{V}O_2$ plateau threshold has not been met. Examples of secondary $\dot{V}O_{2\text{max}}$ criteria include attainment of 90% age-predicted maximal heart rate, maximal respiratory exchange ratio ≥ 1.10 , post-exercise blood lactate concentration $\geq 8 \text{ mmol}\cdot\text{L}^{-1}$, and ratings of perceived exertion ≥ 18 [12, 26]. However, secondary $\dot{V}O_{2\text{max}}$ criteria may result in an underestimation of $\dot{V}O_{2\text{max}}$, given several studies have observed that participants often satisfy the thresholds at submaximal exercise intensities [20, 22, 27]. Moreover, like the $\dot{V}O_2$ plateau, secondary $\dot{V}O_{2\text{max}}$ criteria are dependent on the test protocol and procedures and participant characteristics [12]. Some researchers have therefore encouraged the abandonment of traditional primary and secondary $\dot{V}O_{2\text{max}}$ criteria due to their lack of sensitivity and specificity in establishing whether a 'true' $\dot{V}O_{2\text{max}}$ has been attained [20, 27].

The verification phase is an increasingly recognised procedure for confirming 'true' $\dot{V}O_{2max}$ and typically involves performing a subsequent constant work rate (square wave), sub- or supra- peak bout of exercise after the CPET has been voluntarily terminated (incorporating a short or more extended rest period in between the tests) [20, 27]. The supra peak verification phase is conceptually like the discontinuous $\dot{V}O_{2max}$ tests that were used in developing $\dot{V}O_2$ plateau criteria, but has the advantage of requiring only one visit to the laboratory [28]. The verification phase has emerged as a potentially valid alternative for establishing whether a 'true' $\dot{V}O_{2max}$ has been attained [28, 29]. A recent meta-analysis of studies recruiting apparently healthy participants reported that unlike traditional $\dot{V}O_{2max}$ criteria, the verification phase is not affected by the $\dot{V}O_{2max}$ test protocol or procedures, or participant characteristics such as sex and level of cardiorespiratory fitness [29]. Considering the $\dot{V}O_2$ plateau is less likely to occur in unfit participants, and that secondary criteria commonly underestimate $\dot{V}O_{2max}$ in special and clinical groups [30-32], the verification phase might be particularly useful to establish the occurrence of 'true' $\dot{V}O_{2max}$ or, when $\dot{V}O_{2max}$ has not been elicited, the highest possible attainable $\dot{V}O_2$ in these groups. Notably, participant and test protocol characteristics may not allow the attainment of $\dot{V}O_{2max}$ during the CPET or verification phase. This has direct clinical applications in establishing functional capacity and the effectiveness of exercise training, and the subsequent evaluation of health risks in clinical populations. However, no systematic review and meta-analysis has investigated the utility of the verification phase across diverse special and clinical groups according to different CPET and verification phase protocols and procedures. The effect of these factors on the utility of the verification phase in special and clinical groups is therefore unclear.

The primary aim of the current study was to systematically review and provide a meta-analysis on the suitability of the verification phase for confirming 'true' $\dot{V}O_{2max}$ in special and clinical groups. Secondary aims were to explore the applicability of the verification phase according to specific participant characteristics and investigate which test protocols and procedures minimise the differences between the highest $\dot{V}O_2$ values attained in the CPET and verification phase.

Methods

Protocol and registration

The systematic review and meta-analysis were performed and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [33]. A full PRISMA checklist is shown in S1 Fig. The protocol was registered at <https://www.crd.york.ac.uk/prospero> (CRD42021247658).

Search strategy

MEDLINE (accessed through PubMed), Web of Science, SPORTDiscus, Scopus, and EMBASE were searched for peer-reviewed literature. The search strategy included terms relating to cardiopulmonary exercise test, verification phase, $\dot{V}O_{2max}$ test, and oxygen uptake, using a combination of entry terms and synonyms. Medical subject heading (MeSH) descriptors were also included in the PubMed search. All studies published from the inception of the databases until the search date (14th October 2023) were sought. All references from electronic search results were imported into Endnote bibliographic software (version X9, Bld 12062, Clarivate) and duplicates were removed. The electronic searches were re-run before the final analysis and further studies were retrieved for inclusion. A list containing the full search strategy for each database is available (see S2 Fig). Backward searching for additional relevant studies was conducted by scrutinising the reference lists of the full-text articles of the initial included studies. Forward searching for additional relevant studies was conducted within electronic databases by scrutinising studies that have cited the initial included studies since their publication.

Eligibility criteria for inclusion were studies published in English or Portuguese language involving 1) *Population* – individuals affected by any disease, disability or clinical condition, apparently healthy children and adolescents (<18 years of age), and older adults (\geq 65 years of age) according to American College of Sports Medicine and American Heart Association definitions [34]; 2) *Type of study* – any research design that included at least one CPET and at least one verification phase carried out on a

cycle ergometer, while walking or running or wheelchair propulsion on a treadmill, or on a ski-/wheelchair-/arm-ergometer; and 3) *Outcome* – $\dot{V}O_{2max}$ determined using expired gas analysis during the maximal CPET (control) and verification phase (comparator). Studies were excluded if: 1) they involved secondary analysis of previously included studies; and 2) they investigated older adults, and it could not be ascertained whether any of the participants were below 65 years old. Two blinded researchers performed the searches and screening procedures.

Study selection

Studies were screened for inclusion using a three-step approach: 1) titles and abstracts were initially screened for potentially eligible articles; 2) the full texts of all potentially eligible articles were obtained; and 3) all inclusion and exclusion criteria were applied to full-text articles for the final decision on eligibility. Two of the authors independently determined whether each study met the eligibility criteria. Disagreements were resolved by discussion. Fig 1 shows the screening and selection phases.

Fig 1. Flowchart for the systematic review and meta-analysis according to PRISMA guidelines.

Abbreviations: $\dot{V}O_{2max}$: maximal oxygen uptake.

Data extraction and management

The following data were systematically extracted to a Microsoft Excel (2013) spreadsheet: 1) total sample size; 2) characteristics of study participants (special population or clinical condition, sex, age, body mass index, and cardiorespiratory fitness); 3) exercise modality; 4) type of CPET protocol; 5) verification phase protocol and procedures (work rate, type of recovery, timing in relation to whether performed on the same day or different day to the CPET, and whether or not a verification phase threshold criterion was used); and 6) outcome measures (mean \pm standard deviation [SD] test duration and absolute $\dot{V}O_{2max}$ for the CPET and verification phase). Authors of the original articles were

contacted to request data when these were not reported. Non-responses from authors were followed up with a second email.

Quality assessment

The quality of the included studies was independently assessed by two of the authors using a modified version of the Downs and Black checklist [35] (see S3 Fig). Modified versions of this checklist have been employed in reviews in the sport and exercise sciences, which also mainly used cross-sectional studies for their data retrieval [36, 37]. The original checklist comprises 27 items, which are distributed over five sub-scales: reporting (items 1–10), external validity (items 11–13), bias (items 14–20), confounding (items 21–26), and power (item 27) [35]. The Downs and Black checklist was originally designed for intervention studies. Since the present review does not focus on intervention studies, items 8, 9, 12–16, 19, and 22–26 were excluded, and the remaining 14 items included. Furthermore, an additional item was added on whether the included studies provided information on the sampling method. The term “patient” was replaced by “participant”, the term “principal confounders” by “participant characteristics”, and, where applicable, the term “treatment” was interpreted in the context of “testing” [36, 37]. All items, except item numbers 4 and 6, were rated as “Yes” (1 point), “No” (0 points), or “Unknown” (0 points).

For item 4, both the CPET and verification phase needed to be described in sufficient detail, i.e., 1) the duration and magnitude of increments in the CPET; 2) total duration of the CPET; 3) whether the verification phase used a sub- or supra-peak work rate/speed during the CPET; 4) total duration of the verification phase; 5) exercise modality used for the CPET and verification phase; and 6) type and duration of recovery between the CPET and verification phase. Two points were given if all six items were described, and one point was given if four or five of the six items were described. For item 6, both simple outcome data for the major findings of the study, as well as values for $\dot{V}O_2$ and its unit of measurement needed to be sufficiently described for this item to be scored a “Yes” with 2 points. Only providing the percentage change or absolute difference in $\dot{V}O_2$ between the CPET to exhaustion and

verification phase was not sufficient. If only one was sufficiently described, it was rated as a "Yes" with 1 point. A third author helped reach consensus where there were disagreements between the primary reviewers. Quality assessment cut-off points were decided on retrospectively and studies were regarded as low (0–8 points), moderate (9–14 points), or good (15–17 points) methodological quality, based on the total score achieved on the modified Downs and Black checklist. The level of evidence (LoE) for the results of the main and subgroup analyses was categorised from very limited to strong by combining the quality scores of each of the studies included (Table 1). A figure showing the results of the quality assessments was constructed using the R studio programme (R Core Team. R: A language and environment for statistical computing. R foundation for statistical computing Vienna, Austria; 2013).

Table 1. Criteria for determining the level of evidence based on the quality of studies included in the main analysis and each of the subgroup analyses (adjusted from the criteria provided by Van Tulder, Furlan [38]).

Level	Criteria
Strong	Data provided in multiple studies of good methodological quality OR in one study of good methodological quality and multiple studies of moderate methodological quality.
Moderate	Data provided in multiple studies of moderate methodological quality OR in one study of good methodological quality.
Limited	Data provided in one study of moderate methodological quality.
Very limited	Data provided in one study of low quality.

Statistical analysis

All meta-analyses were performed using the Review Manager (RevMan) software version 5.3 (Copenhagen, The Nordic Cochrane Centre, The Cochrane Collaboration, 2014). Data are presented as the mean \pm SD, following Cochrane Handbook guidelines [39]. These guidelines state that to perform a meta-analysis of continuous data, authors should utilise the mean value, standard deviation, and

number of participants for whom the outcome was measured in each intervention group or test protocol. The outcome was the mean difference (95% confidence interval [CI]) between the CPET and verification phase for the highest absolute $\dot{V}O_2$ value in $L \cdot \text{min}^{-1}$. Given that absolute $\dot{V}O_2$ are continuous data, the weighted mean difference method was used for combining study effect size estimates. With the weighted mean difference method, the pooled effect estimate represents a weighted mean of all included study group comparisons. The weighting assigned to each individual study group (i.e., the comparison of CPET and verification phase results) in the analysis was inversely proportional to the variance of the absolute $\dot{V}O_2$. This method typically assigns more weight in the meta-analysis to studies with higher precision (inverse variance) and larger sample sizes. The weighted mean differences were calculated using random-effects models given the study group differences in participants' characteristics, CPET modalities and protocols, types of recovery, and verification phase protocols. These differences in both participants and protocols characteristics allow that the effect size could vary from study to study. A standardised mean difference method with pooled effect estimate represents a weighted standardised mean of all included study group comparisons and was included as a sensitivity check given the differences in absolute $\dot{V}O_2$ observed across different clinical groups.

Heterogeneity of net group changes in absolute $\dot{V}O_{2\text{max}}$ was examined using the Q statistic. Cochran's Q statistic is computed by summing the squared deviations of each trial's estimate from the overall meta-analytic estimate and weighting each trial's contribution in the same manner as in the meta-analysis. The *p*-values were obtained by comparing the statistic with a χ^2 distribution with *k*-1 degrees of freedom (where *k* is the number of trials). A *p*-value of < 0.10 was adopted since the Q statistic tends to suffer from low differential power [40]. The formal Q statistic was used in conjunction with the methods for assessing heterogeneity. The I^2 statistic was used to measure the extent of inconsistency among the results of the primary study groups, interpreted approximately as the proportion of total variation in point estimates due to heterogeneity rather than sampling error. Effect

sizes with a corresponding I^2 value of $\leq 50\%$ were considered to have low heterogeneity. Potential publication bias or studies with outlier data were assessed using a funnel plot.

Subgroup analyses were defined a priori to investigate the magnitude of differences between CPET and verification phases due to variations in group characteristics, exercise modality, CPET protocol design, or how the verification phase was performed. The following subgroups based on medical conditions and participant characteristics were considered: paediatric (obese and non-obese under 18 yrs), geriatric (≥ 65 yrs), wheelchair (elite wheelchair athletes and individuals in a wheelchair without spinal cord injury or spina bifida), respiratory (cystic fibrosis, chronic asthma/airway disorders, and bronchiectasis), metabolic (overweight and obese adults with and without metabolic syndrome or hypertension), oncological, and cardiological. Forest plots were constructed to display values at the 95% confidence level. Effect sizes were calculated by subtracting the highest mean values for absolute $\dot{V}O_2$ observed in the CPET from the verification phase values, based on grouping studies with selected verification phase characteristics for work rate (i.e., sub vs. supra peak work rate) and type of recovery between the CPET and verification phase (i.e., active vs. passive). The studies were also classified according to whether a criterion threshold for $\dot{V}O_{2max}$ was used for the verification phase, involving an absolute or relative differences in $\dot{V}O_2$ between the CPET and verification phase (i.e., yes vs. no), typically characterised as a percentage difference between tests. In addition to the meta-analytical approach, a participant-level analysis was conducted with data directly reported in the reviewed studies or those supplied by the corresponding authors. The analysis of the differences between the highest $\dot{V}O_2$ values elicited in the CPET and verification phase was based on the threshold criteria utilised by the studies or the error of measurement of the metabolic analysers that were used.

Results

The literature search identified 2108 potential studies, of which 2082 were obtained from electronic databases and 26 from manual searches through a wider inspection of reference lists and citations of

these articles. Forty-three articles published between 1993 and 2023 met the eligibility criteria and were included in the systematic review, whilst 30 presented relevant quantitative information to be considered for meta-analysis (see Fig 1).

Tables 2 and 3 show, respectively, the sample characteristics for the reviewed studies and the exercise testing protocols used to measure the $\dot{V}O_{2max}$ in all participants. Twenty-two studies (51%) used continuous step-incremented protocols, 19 (44%) used ramp-incremented protocols, one used a discontinuous protocol (2%), and one included both continuous and discontinuous step-incremented protocols (2%). Twenty-three studies (53%) used one or more traditional $\dot{V}O_{2max}$ criteria, of which 17 (39%) used a $\dot{V}O_2$ plateau, 16 (37%) used a heart rate plateau or criteria based on the age-predicted maximal heart rate, 17 (39%) used maximal respiratory exchange ratio, 5 (12%) used post-exercise blood lactate concentration, and 6 (14%) used ratings of perceived exertion cut-off values.

Regarding respiratory expired gas analysis procedures, smoothing of pulmonary gas exchange data is required during exercise testing for the determination of $\dot{V}O_{2max}$, especially for data collected from participants on a breath-by-breath basis. The most common approach was based on time averages. Thirty-two studies (74%) reported using time averages of between 10 and 30 s, two (5%) used moving time averages, two (5%) applied 12-breath rolling averages, and three (7%) did not describe which $\dot{V}O_2$ data processing method was applied. Amongst more traditional expired gas collection techniques, four studies (9%) used Douglas bag collections of between 30 and 60 s. No study addressed the effect of different $\dot{V}O_2$ sampling intervals on the difference between the peak $\dot{V}O_2$ values attained in the CPET and verification phase.

Regarding the type of recovery between CPET termination and the start of the verification phase, 10 studies (23%) used active recovery, 11 (26%) used passive recovery, 8 (19%) adopted a combination of passive and active recovery, and 11 (26%) did not report the type of recovery. The verification phase was carried out on a different day as the CPET in three studies (7%). When the verification phase was

performed on the same day as the CPET, the recovery period varied from 4 to 25 min. Ten studies (23%) used a 10-min recovery, which was the most common. Two articles (5%) did not state the duration of the recovery period used in the studies.

Table 2. Sample characteristics of the reviewed studies (N= 43).

Study	Year	Population	Sex M/ F	N	Mean values		
					Age years	BMI kg/m ²	VO _{2max} mL·kg ⁻¹ ·min ⁻¹
Armstrong, Welsman and Winsley [41]	1996	Boys	M	17	9.9	16	64
		Girls	F	18		20.7	52
Astorino, Bediamol [42]	2019	Spinal cord injury	M/F	9/1	33.3	22.6	17.4
Astorino, De La Rosa [43]	2020	Inactive with obesity	F	17	37	39	19.5
Barker, Williams [30]	2011	Children	M/F	8/5	9-10	17.9	50
Baumgart, Moes [44]	2018	Elite para ice hockey players	M/F	14/1	27.1	24.6	36.0
Bhammar, Stickford [45]	2017	Non-obese children	M/F	5/4	11	18.2	43.4
		Obese children	M/F	6/3	10.9	29.9	28.2
Bhammar, Adams-Huet and Babb [46]	2019	Non-obese children	M/F	20/7	11.3	18	39.8
		Obese children	M/F	13/13	11.7	28.6	26.5
Bhammar and Chien [47]	2021	Adults with prehypertension	M	7	21.7	25.2	34.6
			F	4	28.3	23.6	26.6
Blazquez, Guillermo [48]	2011	Chronic fatigue syndrome	F	32	40.3	24.9	15.6
Bowen, Cannon [49]	2012	Chronic heart failure	M	24	64	29.7	14.6
Causer, Shute [50]	2018	Cystic fibrosis (paediatric)	M/F	12/5	12.9	19.5	40.1
		Cystic fibrosis (adults)	M/F	18/10	31.3	22.4	32.0
Cockcroft, Bond [51]	2019	Adolescents	M	7	14.3	21.5	40.7
de Groot, Takken [52]	2009	Spina bifida (normal ambulation)	M/F	10	9.9	18.9	39.4
		Spina bifida (community ambulation)	M/F	10	11.1	21.9	28.7
Goosey-Tolfrey, Paulson [53]	2014	Elite wheelchair athletes (tetraplegics)		9	30	NR	21.2
		Elite wheelchair athletes (paraplegics)	M	9	29		39
		Elite wheelchair athletes (without SCI)		8	27		44.2
Lambrick, Bertelsen [54]	2016	Healthy children	M/F	21/29	9.4	20.6	55.2
Lambrick, Jakeman [55]	2017	Healthy children	M/F	11/10	9.6	20.8	55.3
Leicht, Tolfrey [56]	2013	Elite wheelchair athletes (tetraplegics)		8	28.1	NR	23.1
		Elite wheelchair athletes (paraplegics)	M	8	31.7		37.1
		Elite wheelchair athletes (without SCI)		8	24		39.9
Mahoney, Baughman [57]	2019	Obese adults	M	9	24	31.8	35.1
Manresa-Rocamora, Fuertes-Kenneally [58]	2023	Heart failure and reduced ejected fraction	M/F	13/8	64	26	15.0
McCreery, Mackintosh [59]	2021	Bronchiectasis patients	M/F	7/3	64.5	28.3	11.1
Michalski, Ferreira [60]	2022	Stroke patients	M/F	4/3	58	26.5	21.0
Moreno-Cabañas, Ortega [61]	2020	Metabolic syndrome with obesity	M/F	66/34	57.2	32.1	24.6
Moreno-Cabañas, Ortega [62]	2020	Metabolic syndrome with obesity	M/F	28/16	58	31.8	26.2
Murias, Kowalchuk and Paterson [63]	2010	Older adults	F	6	69	27	23.9
Murias, Kowalchuk and Paterson [64]	2010	Older adults	M	8	68	26	28.3

Table 2. (Continued)

Study	Year	Population	Sex M/F	N	Mean values		
					Age years	BMI kg/m ²	$\dot{V}O_{2max}$ mL·kg ⁻¹ ·min ⁻¹
Murias, Pogliaghi and Paterson [65]	2018	Older adults	M	31	68	25.8	33.0
Oliveira, Barker [66]	2019	Adolescents	M	13	14	18.6	50.9
Pal, Schneider [67]	2021	Prostate cancer survivors	M	10	59.9	27.3	23.1
		Breast cancer survivors	F	11			
Paulson, Goosey-Tolfrey [68]	2013	Elite wheelchair athletes (tetraplegics)		8	31		21.4
		Elite wheelchair athletes (paraplegics)	M	10	30	NR	39.4
		Elite wheelchair athletes (without SCI)		8	27		44.2
Robben, Poole and Harms [69]	2013	Children with expiratory flow limitation	M	12	9.6	18	37.2
Rowland [70]	1993	Healthy children	M/F	6/3	11.4	16.6	55.0
Sansum, Weston [71]	2019	Healthy children	M	76	13.3	20.4	46.9
			F	52	13.9	21.2	36.2
Sawyer, Tucker [72]	2015	Sedentary with obesity	M	10	33.4	37.1	22.0
			F	9	38.4	34.5	
Saynor, Barker [31]	2013	Cystic fibrosis	M/F	10/4	13.1	24.7	33.0
Saynor, Barker [73]	2013	Cystic fibrosis	M/F	9/4	12.8	21.7	34.8
Schaun, Alberton [32]	2021	Hypertensive older adults	M/F	9/24	67.1	32.2	23.9
Schaun, Alberton [74]	2022	Hypertensive older adults	M/F	4/8	67.3	31.2	25.9
Schneider, Schlüter [75]	2020	Prostate cancer survivors	M	32	66.4	27.5	21.7
		Breast cancer survivors	F	43	57.5	25.8	21.3
Tomlinson, Barker [76]	2018	Cystic fibrosis	M/F	21/15	13.4	20.7	37.7
		Healthy children and adolescents	M/F	21/15	13.2	20.2	39.9
Villanueva, Campbell [77]	2021	Older adults	M	9	69	26	29.8
			F	13	65	26.6	24.2
Werkman, Hulzebos [78]	2011	Cystic fibrosis	M/F	8/8	14.6	18.1	38.9
Woloschuk, Hodges [79]	2020	Healthy boys	M	12	9.7	16.7	46.9
Wood, Hills [80]	2010	Overweight	M	32	36.7	28.4	42.4
			F	36	36.9	28.1	32.2
		Obese	M	35	37.7	33.6	37.7
			F	32	37.1	32.9	27.2

Abbreviations: *BMI* = body mass index; *F* = females; *M* = males; *NR* = not reported; *SCI* = spinal cord injury; $\dot{V}O_{2max}$ = maximal oxygen uptake. *Note:* Authors were contacted to provide unpublished data.

Table 3. Characteristics of the cardiopulmonary exercise test (CPET) and verification phase protocols used in the reviewed studies (N = 43)

Study	$\dot{V}O_2$ sampling method	Traditional $\dot{V}O_{2max}$ criteria adopted	Ergometer	CPET Protocol	Recovery Phase	Verification Phase (VP) Protocol	Verification Criteria Threshold
Armstrong, Welsman and Winsley [41]	NS	$\dot{V}O_2$ plateau of 150 mL·min ⁻¹ or 2.1 mL·kg ⁻¹ ·min ⁻¹ during the final minute of the penultimate and final stages; HR ≥ 200 bpm; RER _{max} ≥ 1.00	TR	DiscSI (grade elevation of 2.5% after each 3-min stage, with constant speed)	Different day	↑ slope to 2.5% and 5% in relation to the first test	NS
Astorino, Beditamol [42]	2 x 15-s	$\dot{V}O_2$ plateau - individual approach based on the expected O ₂ cost	CYC	Ramp (3 W·min ⁻¹ for tetraplegics and 13 W·min ⁻¹ for paraplegics)	10-min active	105% WR _{max}	NS
Astorino, De La Rosa [43]	2 x 15-s	HR _{max} ≤ 2 bpm	CYC	Ramp (20 W·min ⁻¹)	10-min active	105% WR _{max}	Typical error score equal to 0.06 L·min ⁻¹ in $\dot{V}O_{2max}$ between protocols
Barker, Williams [30]	15-s average	$\dot{V}O_2$ plateau (negative linear regression residuals during last 60 s of test); RER _{max} ≥ 1.00; ≥ 85% APMHR and HR _{max} 195 bpm; La _{max} ≥ 6 mmol·L ⁻¹	CYC	Ramp (10 W·min ⁻¹)	10-min active and 5-min passive	2-min at 10W, then 105% WR _{max}	IP vs. VP: $\dot{V}O_{2max}$ difference < 5%
Baumgart, Moes [44]	30-s moving average	NS	SKI	CSI (10 W every 30-s)	5-min passive and 3-min active	110% WR _{max}	NS
Bhammar, Stickford [45]	20-s	NS	CYC	CSI (10 or 15 W·min ⁻¹)	15-min passive	2-min at 20W, then 105% WR _{max}	Difference between measured $\dot{V}O_{2verif}$ and $\dot{V}O_{2max}$ in relation to the difference in predicted $\dot{V}O_{2verif}$ and $\dot{V}O_{2max}$

Table 3. (Continued)

Study	$\dot{V}O_2$ sampling method	Traditional $\dot{V}O_{2max}$ criteria adopted	Ergometer	CPET Protocol	Recovery Phase	Verification Phase (VP) Protocol	Verification Criteria Threshold
Bhammar, Adams-Huet and Babb [46]	20-s	NS	CYC	CSI (10 or 15 $W \cdot \text{min}^{-1}$)	15-min passive	105% WR_{max}	Difference between measured $\dot{V}O_{2verif}$ and $\dot{V}O_{2max}$ in relation to the difference in predicted $\dot{V}O_{2verif}$ and $\dot{V}O_{2max}$
Bhammar and Chien [47]	20-s	$\dot{V}O_2$ plateau (difference between measured $\dot{V}O_2$ between the penultimate and final stage was less than 50% of the “expected” increase).	CYC	CSI (20 $W \cdot \text{min}^{-1}$ for women and 25 $W \cdot \text{min}^{-1}$ for men)	At least 15-min	2-min at 40W (men) or 30W (women), then 105% WR_{max}	Difference between measured $\dot{V}O_{2verif}$ and expected $\dot{V}O_{2verif}$
Blazquez, Guillamo [48]	NS	NS	CYC	CSI (20 $W \cdot \text{min}^{-1}$)	4-min	↑ WR_{max} every 30 sec up to exhaustion	NS
Bowen, Cannon [49]	12-breath rolling average	$RER_{max} \geq 1.10$	CYC	Ramp (4-18 $W \cdot \text{min}^{-1}$)	5-min	95% WR_{max}	$\dot{V}O_{2max}$ was confirmed when $p > 0.05$ in relation to $\dot{V}O_{2verif}$
Causser, Shute [50]	15-s average	predicted $\dot{V}O_{2peak}$, WR_{peak} , HR_{max} , 80% VE_{max} , RER_{max} 1.03 or 1.05, RPE 9 or 17.	CYC	Ramp (10-25 $W \cdot \text{min}^{-1}$)	5-min active and 10-min passive	3-min at 20W, then 110% WR_{max}	IP vs. VP: $\dot{V}O_{2max}$ difference < 9%
Cockcroft, Bond [51]	10-s	NS	CYC	Ramp (details were not reported)	NS	NS	NS
de Groot, Takken [52]	30-s average	$\dot{V}O_2$ plateau of 2.1 $\text{mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ (difference between normalised $\dot{V}O_{2peak}$ and $\dot{V}O_2$ in the last 30 seconds of the minute before the last minute); $RER_{max} \geq 1.00$; > 95% APMHR	TR	CSI (0.25 or 0.5 $\text{km} \cdot \text{min}^{-1}$ at a constant 2% grade)	4-min	110% WR_{max}	NS

Table 3. (Continued)

Study	$\dot{V}O_2$ sampling method	Traditional $\dot{V}O_{2max}$ criteria adopted	Ergometer	CPET Protocol	Recovery Phase	Verification Phase (VP) Protocol	Verification Criteria Threshold
Goosey-Tolfrey, Paulson [53]	NS	NS	TR	CSI (grade elevation of 0.1%/40-s and 0.3%·min ⁻¹ at a constant speed)	5-min active	↑ slope to 0.1 or 0.3% in relation to IP	NS
Lambrick, Bertelsen [54]	10-bins	NS	TR	CSI (0.5 km·h ⁻¹ at a constant 1% grade)	15-min passive	105% WR _{max}	NS
Lambrick, Jakeman [55]	10-s bins	NS	TR	CSI and DiscSI (1 km·h ⁻¹ each min up to 8 km·h ⁻¹ , then 0.5 km·h ⁻¹ each min, at a constant 1% grade)	15-min	105% WR _{max}	NS
Leicht, Tolfrey [56]	30-s rolling average	$\dot{V}O_2$ plateau (negative deviation > 0.1 l/min between observed and predicted peak $\dot{V}O_2$ based on linear regression); RER _{max} > 1.05, 1.10, 1.15, and 1.20; > 85%, 90%, 95%, and 100% APMHR and La _{max} > 4, 5, and 6 mmol·L ⁻¹	TR	CSI (grade elevation of 0.1%/40-s and 0.3%/min at a constant speed)	2-min passive and 5-min active	↑ slope to 0.3 or 0.6% in relation to IP	NS
Mahoney, Baughman [57]	2 x 15-s	NS	CYC	Ramp ($W = kg \times (\dot{V}O_2 - 7)/1.8$)	Different day	5-min at 50W, then 80, 90, 100 and 105% WR _{max}	NS
Manresa-Rocamora, Fuertes-Kenneally [58]	12-breath rolling average	RER _{max} ≥ 1.10	CYC	Ramp (details were not reported)	5-min active	95% WR _{max}	IP vs. VP: $\dot{V}O_{2max}$ difference ≤ 3%
McCreery, Mackintosh [59]	15-s	NS	CYC	Ramp (10 W·min ⁻¹)	NS	110% WR _{max}	IP vs. VP: $\dot{V}O_{2max}$ difference ≤ 9%

Table 3. (Continued)

Study	$\dot{V}O_2$ sampling method	Traditional $\dot{V}O_{2max}$ criteria adopted	Ergometer	CPET Protocol	Recovery Phase	Verification Phase (VP) Protocol	Verification Criteria Threshold
Michalski, Ferreira [60]	30-s	NS	TR	Ramp (predicted initial and final speeds/grades determined during familiarisation)	20-min passive	2-min at 50% WR_{max} , 1 min at 70% WR_{max} , then 1 stage > IP- WR_{max}	NS
Moreno-Cabañas, Ortega [61]	15-s average	$\dot{V}O_2$ plateau (difference between measured $\dot{V}O_2$ between the penultimate and final stage < 50% of the “expected” increase)	CYC	CSI (15 $W \cdot \text{min}^{-1}$ for women and 20 $W \cdot \text{min}^{-1}$ for men)	5-min active and 15-min passive	2-min at 30 or 50W, then 110% WR_{max}	Difference between $\dot{V}O_2$ from CPET vs. $\dot{V}O_{2verif}$ ($\dot{V}O_2$ -work rate relationship) < 50% of the “expected” increase.
Moreno-Cabañas, Ortega [62]	15-s average	$\dot{V}O_2$ plateau (difference between measured $\dot{V}O_2$ between the penultimate and final stage < 50% of the “expected” increase)	CYC	CSI (15 $W \cdot \text{min}^{-1}$ for women and 20 $W \cdot \text{min}^{-1}$ for men)	5-min active and 15-min passive	2-min at 30 or 50W, then 110% WR_{max}	Difference between $\dot{V}O_2$ from CPET vs. $\dot{V}O_{2verif}$ ($\dot{V}O_2$ -work rate relationship) < 50% of the “expected” increase.
Murias, Kowalchuk and Paterson [63]	20-s	NS	CYC	Ramp (12-15 $W \cdot \text{min}^{-1}$)	5-min active	85% WR_{max}	NS
Murias, Kowalchuk and Paterson [64]	20-s	NS	CYC	Ramp (15-20 $W \cdot \text{min}^{-1}$)	5-min active	85% WR_{max}	NS
Murias, Pogliaghi and Paterson [65]	20-s average	NS	CYC	Ramp (15-20 $W \cdot \text{min}^{-1}$)	5-min active	85% and 105% WR_{max}	IP vs. VP: $\dot{V}O_{2max}$ difference $\leq 2.0 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$
Oliveira, Barker [66]	NS	NS	TR	CSI (0.5 km/30-s)	10-min	↑ slope to 5%	NS
Pal, Schneider [67]	20-s average	$RER_{max} \geq 1.10$; $APMHR$; $La_{max} \geq 8 \text{ mmol} \cdot \text{L}^{-1}$; $RPE \geq 18$	CYC	CSI (10 $W \cdot \text{min}^{-1}$)	10-min passive	110% WR_{max}	IP vs. VP: $\dot{V}O_{2max}$ difference $\leq 3\%$

Table 3. (Continued)

Study	$\dot{V}O_2$ sampling method	Traditional $\dot{V}O_{2max}$ criteria adopted	Ergometer	CPET Protocol	Recovery Phase	Verification Phase (VP) Protocol	Verification Criteria Threshold
Paulson, Goosey-Tolfrey [68]	NS	NS	TR	CSI (grade elevation of 0.1%/40-s and 0.3%/min, at a constant speed)	5-min active	↑ slope to 0.1 or 0.3% in relation to IP	NS
Robben, Poole and Harms [69]	20-s average	$\dot{V}O_2$ plateau (difference between modelled and actual > 50% of the regression slope for the linear portion of the $\dot{V}O_2$ -workrate relationship); $RER_{max} \geq 1.00$; $\geq 90\%$ APMHR	CYC	CSI (10 W·min ⁻¹)	15-min passive	100% WR_{max}	Difference between $\dot{V}O_2$ from CPET vs. $\dot{V}O_{2verif}$ ($\dot{V}O_2$ -work rate relationship) < 50% of the "expected" increase
Rowland [70]	15-s	$\dot{V}O_2$ plateau of < 2 mL·kg ⁻¹ ·min ⁻¹	TR	CSI (grade elevation of 2.5% at each 3-min stage, at a constant speed)	Different day	↑ slope to 2.5, 5 and 7% in relation to IP	NS
Sansum, Weston [71]	10 to 15-s average	$\dot{V}O_2$ plateau (linear regression over the "linear" portion of the $\dot{V}O_2$ response); $RER_{max} \geq 1.00$ and 1.1; $HR_{max} > 195$ bpm, ≥ 85 and 95% APMHR	CYC	Ramp (10-30 W·min ⁻¹)	25-min	3-min at 20W, then 105 or 110% WR_{max}	IP vs. VP: $\dot{V}O_{2max}$ difference < 5%
Sawyer, Tucker [72]	2 x 15-s average	NS	CYC	Ramp (15 W·min ⁻¹ for women and 30 W·min ⁻¹ for men)	5-10 min active	100% WR_{max}	≥ 60 -s
Saynor, Barker [31]	15-s average	$\dot{V}O_2$ plateau (linear regression over the "linear" portion of the $\dot{V}O_2$ response); $RER_{max} \geq 1.00$ and 1.10; HR_{max} of 180 bpm and $\geq 95\%$ APMHR	CYC	Ramp (10-25 W·min ⁻¹)	5-min active and 10-min passive	3-min at 20W, then 110% WR_{max}	IP vs. VP: $\dot{V}O_{2max}$ difference < 9%
Saynor, Barker [73]	15-s average	NS	CYC	Ramp (10-25 W·min ⁻¹)	5-min active and 10-min passive	3-min at 20W, then 110% WR_{max}	NS
Schaun, Alberton [32]	2 x 20-s average	$\dot{V}O_2$ plateau (difference between measured $\dot{V}O_2$ between the penultimate and final stage < 50% of the "expected" increase; 150 mL·min ⁻¹); HR_{max} within 10 bpm of APMHR; $RER_{max} \geq 1.10$; RPE ≥ 18	TR	CSI (0.5 km and 1%/min)	10-min passive	2-min at 50% WR_{max} , 1 min at 70% WR_{max} , then 1 stage > IP- WR_{max}	IP vs. VP: $\dot{V}O_{2max}$ difference < 3%

Table 3. (Continued)

Study	$\dot{V}O_2$ sampling method	Traditional $\dot{V}O_{2max}$ criteria adopted	Ergometer	CPET Protocol	Recovery Phase	Verification Phase (VP) Protocol	Verification Criteria Threshold
Schaun, Alberton [74]	2 x 20-s average	$\dot{V}O_2$ plateau of 150 mL·min ⁻¹ ; HR _{max} within 10 bpm of APMHR; RER _{max} ≥ 1.10; RPE ≥ 18	TR	CSI (0.5 km and 1%/min)	10-min passive	2-min at 50% WR _{max} , 1 min at 70% WR _{max} , then 1 stage > IP-WR _{max}	IP vs. VP: $\dot{V}O_{2max}$ difference < 3%
Schneider, Schlüter [75]	20-s average	RER _{max} ≥ 1.1; APMHR; La _{max} ≥ 8 mmol·L ⁻¹ ; RPE ≥ 18	CYC	CSI (10 W·min ⁻¹)	10-min passive	110% WR _{max}	IP vs. VP: $\dot{V}O_{2max}$ difference < 3%
Tomlinson, Barker [76]	10-s	$\dot{V}O_2$ plateau (linear regression over the "linear" portion of the $\dot{V}O_2$ response); RER _{max} ≥ 1.00 and 1.10; HR _{max} of 180 bpm and ≥ 95% APMHR	CYC	Ramp (no details were reported)	5-min active and 10-min passive	3-min at 20W, then 110% WR _{max}	NS
Villanueva, Campbell [77]	3 x 10-s	NS	CYC	Ramp (15 W·min ⁻¹ for women and 20 W·min ⁻¹ for men)	10-min active	85 and 110% WR _{max}	NS
Werkman, Hulzebos [78]	30-s	$\dot{V}O_2$ plateau of ≤ 2.1 mL·kg ⁻¹ ·min ⁻¹ ; RER _{max} ≥ 1.00; HR _{max} ≥ 95% APMHR	CYC	CSI (10, 15 or 20 W·min ⁻¹ based on each participant's height)	10-min passive	1-min unloaded cycling, then an increase in workload every 10-s	NS
Woloschuk, Hodges [79]	2 x 15-s	NS	CYC	CSI (10-15 W·min ⁻¹)	10-min	105% WR _{max}	NS
Wood, Hills [80]	30-s average	$\dot{V}O_2$ plateau < 50% of expected for the change in WR; HR within ± 11 bpm of APMHR; RER _{max} ≥ 1.15; La _{max} ≥ 8 mmol·L ⁻¹ ; RPE ≥ 18	TR	CSI (grade elevation of 2.5%/min at a constant speed)	5-10 min passive	↑ WR _{max} every minute up to exhaustion	NS

Abbreviations: *APMHR* = age-predicted maximal heart rate; *bpm* = beats per minute; *CPET* = cardiopulmonary exercise test; *CSI* = continuous step-incremented; *CYC* = cycling; *DiscSI* = discontinuous step-incremented; *HR* = heart rate; *HR_{max}* = maximal heart rate; *IP* = incremental phase; *kg* = kilogram; *La_{max}* = maximal blood lactate concentration; *NS* = not stated; *O₂* = oxygen; *RER_{max}* = maximal respiratory exchange ratio; *RPE* = rating of perceived exertion; *SD* = standard deviation; *SKI* = ski ergometer; *TR* = treadmill; $\dot{V}O_2$ = oxygen uptake; $\dot{V}O_{2max}$ = maximal oxygen uptake; $\dot{V}O_{2verif}$ = maximal oxygen uptake obtained in the verification phase; *VP* = verification phase; *W* = watts; *WR* = work rate; *WR_{max}* = maximal work rate. *Note:* Authors were contacted to provide unavailable data.

Twenty-five studies (58%) used square-wave verification phase protocols (i.e., the work rate was immediately increased to the target sub or supra peak work rate), whereas 17 (40%) used multistage verification phase protocols characterised by an initial warm-up stage. Only one study (2%) did not describe the verification phase protocol. The peak work rate used in the verification phase protocols ranged from 80% to 110% of the peak work rate attained in the CPET across studies. Most studies applied a supra peak work rate based on the peak work rate achieved in the CPET ($n = 33$; 77%). Three studies (7%) applied both sub and supra peak work rates within the same study, two (5%) used peak work rate, four (9%) applied only a sub peak work rate, and one (2%) did not describe the verification phase work rate. The mean times to exhaustion for the CPET and verification phase were 568 s (SD, 143 s) and 127 s (SD, 57 s), respectively.

Twenty-two studies (51%) employed threshold criteria to analyse differences between the highest $\dot{V}O_2$ attained in the CPET and verification phase and were frequently based on the intra-subject coefficient of variation acquired from the researchers' laboratories or from published literature. Threshold criteria included a difference in $\dot{V}O_2$ ($L \cdot \text{min}^{-1}$) of < 2%, < 3%, < 5% and < 9%, and an absolute difference between measured and predicted $\dot{V}O_2$ from linear extrapolations of $\dot{V}O_2$ / work rate responses during the CPET (such as < 50% of the "expected" increase). Other cut-off points included a typical error of $0.06 L \cdot \text{min}^{-1}$, $2.1 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$, or when the comparison of the highest group mean $\dot{V}O_2$ value obtained in the CPET versus the verification phase resulted in $p > 0.05$.

Methodological quality of the included studies

There was 86% agreement between the two authors in ranking the items initially, and full agreement was reached upon discussion with a third author. Two studies (5%) were rated as having low methodological quality, 37 (86%) as moderate, and four (9%) as good (Fig 2). The quality of the studies included in the main and subgroup comparisons were used to underpin the respective LoE.

Fig 2. Quality scores for the 43 included studies. Grey dots for items scored as 'no', yellow dots are for items scored 'yes' (1), and blue dots for 'yes' (2). Black dots are added where authors retrospectively provided extra information that would have led to a higher quality appraisal score if included in the original publication.

Group-level quantitative data synthesis: Differences between the highest $\dot{V}O_2$ attained in the CPET and verification phase

Table 4 shows comparisons between the highest $\dot{V}O_2$ values elicited in the CPET and verification phase for each study. Fig 3 displays the forest plot of effect sizes and 95% CIs for the highest $\dot{V}O_2$ values (30 studies) based on the random effects meta-analysis results. The highest $\dot{V}O_2$ was not different between CPET and verification phase (mean difference = -0.00 [95% CI = -0.03 to 0.03] L·min⁻¹, $p = 0.87$; LoE: strong). Given the potential for large heterogeneity in mean $\dot{V}O_{2max}$ values across the different populations included in the review, the overall and subgroup meta-analysis findings were robust to a sensitivity check using the alternative statistical approach of using standardised mean differences. This method resulted in a different weighting pattern to the individual study group differences between the CPET and verification phase $\dot{V}O_{2max}$ values (data not presented). However, the overall effect was unchanged (standardised mean difference = -0.01 [95% CI = -0.09 to 0.08] L·min⁻¹, $p = 0.87$, LoE: strong). Pooled data for $\dot{V}O_{2max}$ following the CPET and verification phase showed no significant heterogeneity among all the studies (see Fig 3). Except for one of the included studies judged to be an outlier [44], the meta-analysed studies were judged to have a low-risk of bias as shown by the funnel plot (Fig 4).

Table 4. Overall group-level comparisons for the reviewed studies for the highest $\dot{V}O_2$ values attained in the cardiopulmonary exercise test (CPET) and verification phase (VP) (N = 30).

Study	Specific Experimental Condition	CPET			VP			% Weight	Mean Difference IV, Random, 95% CI [L·min ⁻¹]
		Mean L·min ⁻¹	SD L·min ⁻¹	n	Mean L·min ⁻¹	SD L·min ⁻¹	n		
Armstrong, Welsman and Winsley [41]	↑ slope to 2.5% in relation to the first test (boys)	1.93	0.23	17	1.95	0.24	17	3.5%	-0.02 [-0.18, 0.14]
	↑ slope to 5% in relation to the first test (boys)	1.93	0.23	17	1.98	0.17	17	4.8%	-0.05 [-0.19, 0.09]
	↑ slope to 2.5% in relation to the first test (girls)	1.85	0.28	18	1.90	0.26	18	2.8%	-0.05 [-0.23, 0.13]
	↑ slope to 5% in relation to the first test (girls)	1.85	0.28	18	1.91	0.35	18	2.0%	-0.05 [-0.27, 0.15]
Astorino, Bediamol [42]	N/A	1.30	0.45	10	1.31	0.43	10	0.6%	-0.01 [-0.40, 0.38]
Astorino, De La Rosa [43]	Baseline	1.99	0.37	17	1.98	0.32	17	1.6%	0.01 [-0.22, 0.24]
	Post 3-week	2.00	0.40	17	2.04	0.38	17	1.3%	-0.04 [-0.30, 0.22]
	Post 6-week	2.08	0.34	17	2.08	0.32	17	1.8%	0.00 [-0.22, 0.22]
Barker, Williams [30]	N/A	1.69	0.284	13	1.615	0.307	13	1.7%	0.07 [-0.15, 0.30]
Baumgart, Moes [44]	N/A	2.506	0.463	14	2.058	0.313	14	1.0%	0.45 [0.16, 0.74]
Bhammar, Stickford [45]	Obese children	1.57	0.27	9	1.71	0.31	9	1.2%	-0.14 [-0.41, 0.13]
	Non-obese children	1.84	0.48	9	1.94	0.47	9	0.5%	-0.10 [-0.54, 0.34]
Bhammar, Adams-Huet and Babb [46]	Obese children	1.514	0.373	21	1.656	0.436	21	1.5%	-0.14 [-0.39, 0.10]
	Non-obese children	1.442	0.335	23	1.529	0.341	23	2.3%	-0.09 [-0.28, 0.11]
Bhammar and Chien [47]	Women	1.595	0.14	4	1.663	0.09	4	3.3%	-0.07 [-0.23, 0.10]
	Men	2.596	0.24	7	2.636	0.27	7	1.2%	-0.04 [-0.31, 0.23]
Blazquez, Guillamo [48]	N/A	0.98	0.363	32	0.918	0.258	32	3.7%	0.06 [-0.09, 0.22]
Bowen, Cannon [49]	$\dot{V}O_{2max}$	1.30	0.36	14	1.31	0.36	14	1.2%	-0.01 [-0.28, 0.26]
	$\dot{V}O_{2peak}$	1.23	0.33	9	1.29	0.31	9	1.0%	-0.06 [-0.36, 0.24]
Causer, Shute [50]	Paediatric	1.86	0.70	14	1.84	0.72	14	0.3%	0.02 [-0.51, 0.55]
	Adults	2.08	0.69	28	1.98	0.64	28	0.7%	0.10 [-0.25, 0.45]
Lambrick, Jakeman [55]	CSI	2.489	0.369	21	2.489	0.315	21	2.0%	0.00 [-0.21, 0.21]
	DisCSI	2.448	0.342	21	2.399	0.329	21	2.1%	0.05 [-0.15, 0.25]

Table 4. (Continued)

Study	Specific Experimental Condition	CPET			VP			% Weight	Mean Difference
		Mean L·min ⁻¹	SD L·min ⁻¹	n	Mean L·min ⁻¹	SD L·min ⁻¹	n		IV, Random, 95% CI [L·min ⁻¹]
Leicht, Tolfrey [56]	Tetraplegics (Day 1)	1.574	0.354	8	1.649	0.393	8	0.7%	-0.07 [-0.44, 0.29]
	Tetraplegics (Day 2)	1.605	0.416	8	1.553	0.302	8	0.7%	0.05 [-0.30, 0.41]
	Paraplegics (Day 1)	2.47	0.335	8	2.395	0.335	8	0.8%	0.08 [-0.25, 0.40]
	Paraplegics (Day 2)	2.335	0.242	8	2.298	0.273	8	1.4%	0.04 [-0.22, 0.29]
	Non-SCI (Day 1)	3.059	0.627	8	2.978	0.53	8	0.3%	0.08 [-0.49, 0.65]
	Non-SCI (Day 2)	3.146	0.56	8	3.053	0.474	8	0.3%	0.09 [-0.42, 0.60]
Mahoney, Baughman [57]	80% WR _{max}	3.37	0.39	9	3.53	0.47	9	0.6%	-0.16 [-0.56, 0.24]
	90% WR _{max}	3.37	0.39	9	3.61	0.54	9	0.5%	-0.24 [-0.68, 0.20]
	100% WR _{max}	3.37	0.39	9	3.57	0.51	9	0.5%	-0.20 [-0.62, 0.22]
	105% WR _{max}	3.37	0.39	9	3.41	0.53	9	0.5%	-0.04 [-0.47, 0.39]
McCreery, Mackintosh [59]	Baseline	0.901	0.211	10	0.853	0.31	10	3.4%	0.05 [-0.11, 0.21]
	Post 8-week	0.893	0.179	10	0.788	0.47	10	2.9%	0.10 [-0.07, 0.28]
Michalski, Ferreira [60]	N/A	1.36	0.47	7	1.44	0.51	7	0.3%	-0.08 [0.59, 0.43]
Moreno-Cabañas, Ortega [61]	N/A	2.144	0.538	100	2.207	0.529	100	4.0%	-0.06 [-0.21, 0.08]
Moreno-Cabañas, Ortega [62]	Pre-training	2.04	0.53	44	2.11	0.52	44	1.8%	-0.07 [-0.29, 0.15]
	Post-training	2.26	0.6	44	2.32	0.55	44	1.5%	-0.06 [-0.30, 0.18]
Murias, Pogliaghi and Paterson [65]	85% WR _{max}	2.184	0.554	8	2.184	0.546	8	0.3%	0.00 [-0.54, 0.54]
	105% WR _{max}	2.519	0.538	23	2.574	0.507	23	1.0%	-0.05 [-0.36, 0.25]
Robben, Poole and Harms [69]	Boys	1.26	0.19	12	1.12	0.7	12	0.5%	0.05 [-0.36, 0.46]
Rowland [70]	↑ slope to 2.5% in relation to the first test	1.833	0.146	9	1.87	0.129	9	5.4%	-0.04 [-0.16, 0.09]
	↑ slope to 5% in relation to the first test	1.833	0.146	9	1.836	0.167	9	4.2%	-0.00 [-0.15, 0.14]
	↑ slope to 7.5% in relation to the first test	1.833	0.146	9	1.805	0.109	9	6.2%	0.03 [-0.09, 0.15]
Sansum, Weston [71]	Non-overweight boys	2.39	0.70	65	2.24	0.66	65	1.6%	0.15 [-0.08, 0.38]
	Overweight boys	3.03	0.68	11	3.04	0.70	11	0.3%	-0.01 [-0.59, 0.57]
	Non-overweight girls	1.96	0.30	44	1.90	0.35	44	4.7%	0.06 [-0.08, 0.20]
	Overweight girls	1.94	0.40	8	1.86	0.31	8	0.7%	0.08 [-0.27, 0.43]
Sawyer, Tucker [72]	N/A	2.29	0.71	19	2.34	0.67	19	0.5%	-0.05 [-0.49, 0.39]
Saynor, Barker [31]	Boys	1.995	0.815	10	1.921	0.685	10	0.2%	0.07 [-0.59, 0.73]
	Girls	1.42	0.562	4	1.57	0.661	4	0.1%	-0.15 [-1.00, 0.70]

Table 4. (Continued)

Study	Specific Experimental Condition	CPET			VP			% Weight	Mean Difference IV, Random, 95% CI [L·min ⁻¹]
		Mean L·min ⁻¹	SD L·min ⁻¹	n	Mean L·min ⁻¹	SD L·min ⁻¹	n		
Saynor, Barker [73]	Test 1	1.77	0.57	13	1.68	0.56	13	0.5%	0.09 [-0.34, 0.52]
	Test 2	1.76	0.56	13	1.63	0.52	13	0.5%	0.13 [-0.29, 0.55]
	Test 3	1.68	0.55	13	1.62	0.64	13	0.5%	0.06 [-0.36, 0.48]
Schaun, Alberton [32]	N/A	1.72	0.46	33	1.83	0.50	33	1.6%	-0.11 [-0.34, 0.12]
Schaun, Alberton [74]	Baseline	1.72	0.43	12	1.94	0.45	12	0.7%	-0.22 [-0.57, 0.13]
	Post 12-week	1.87	0.44	12	2.00	0.50	12	0.6%	-0.13 [-0.51, 0.25]
Schneider, Schlüter [75]	N/A	1.65	0.36	75	1.60	0.38	75	6.3%	0.05 [-0.07, 0.17]
Villanueva, Campbell [77]	85% WR _{max}	1.85	0.73	22	1.86	0.72	22	0.5%	-0.01 [-0.44, 0.42]
	110% WR _{max}	1.85	0.57	20	1.79	0.73	20	0.5%	0.06 [-0.35, 0.47]
Werkman, Hulzebos [78]	N/A	1.90	0.60	12	1.90	0.70	12	0.3%	0.00 [-0.52, 0.52]

Abbreviations: *CI* = confidence interval; *CPET* = cardiopulmonary exercise test; *CSI* = continuous step-incremented; *DisCSI* = discontinuous step-incremented; *IV* = inverse variance; *n* = total of participants included; *Non-SCI* = non-spinal cord-injured; *N/A* = not applicable; *SD* = standard deviation; $\dot{V}O_{2peak}$ = peak oxygen uptake; $\dot{V}O_{2max}$ = maximal oxygen uptake; $\dot{V}O_{2peak}$ = peak oxygen uptake; *VP* = verification phase; *WR_{max}* = maximal work rate. Note: whenever possible, authors were contacted to provide unavailable data. %Weight = weight attributed to each study due to its statistical power.

Fig 3. Forest plot for all the studies included in the meta-analysis (n = 30) for the highest $\dot{V}O_2$ (L·min⁻¹) attained in the cardiopulmonary exercise test (CPET) and verification phase using a random-effects model.

Data are reported as mean differences adjusted for control data (95% CI).

Fig 4. Funnel plot assessment of publication bias for the highest $\dot{V}O_2$ (L·min⁻¹) attained in the cardiopulmonary exercise test (CPET) and verification phase. One outlier identified, which may relate to methodological error within the verification phase protocol (see discussion section).

There was no statistically significant difference between CPET and verification-derived $\dot{V}O_{2max}$ for the paediatric group that included seven studies with 18 experimental conditions (mean difference = -0.01 [95% CI = -0.05 to 0.03] L·min⁻¹, $p = 0.68$, LoE: moderate). Additionally, the subgroup analysis of obese and non-obese paediatric participants, composed of two studies with four experimental conditions, revealed no statistically significant difference (mean difference = -0.11 [95% CI = -0.24 to 0.01] L·min⁻¹, $p = 0.08$, LoE: moderate). The wheelchair group consisted of three studies with eight experimental conditions, and no statistically significant difference was observed between the CPET and verification phase (mean difference = 0.11 [95% CI = -0.02 to 0.23] L·min⁻¹, $p = 0.10$, LoE: strong). The chronic respiratory group consisted of 6 studies and 11 experimental conditions that demonstrated no statistically significant significance between the CPET and verification phase (mean difference = 0.07 [95% CI = -0.02 to 0.17] L·min⁻¹, $p = 0.14$, LoE: strong). The 15 girls from the study by Robben et al. [69] were removed from the meta-analysis, since the reported SD was an extreme outlier (e.g., the study would have been weighted at 40.7% in the final meta-analysis). The subgroup of paediatric patients with cystic fibrosis included four studies and seven experimental conditions (mean difference = 0.06 [95% CI = -0.13 to 0.25] L·min⁻¹, $p = 0.55$ LoE: strong). The geriatric group incorporated four studies with seven experimental conditions, and results showed no statistically significant difference between the CPET and verification phase (mean difference of -0.08 [95% CI = -0.21 to 0.05] L·min⁻¹, $p = 0.20$ LoE: moderate). Finally, the metabolic group including individuals with overweight or obesity (with and without metabolic disease), comprised six studies and 13 experimental conditions. These studies used ramp-based cycle ergometry and demonstrated no statistically significant

difference between CPET and the verification phase (mean difference = -0.06 [95% CI = -0.13 to -0.01] L·min⁻¹, $p = 0.09$).

Regarding the CPET and verification phase protocols (Fig 5 combines all subgroup analyses into the same category), there were no statistically significant differences for verification phase work rate (mean difference = 0.00 [95% CI = -0.04 to 0.04] L·min⁻¹, $p = 0.98$, LoE: strong), recovery mode (mean difference = -0.02 [95% CI = -0.07 to 0.02] L·min⁻¹, $p = 0.29$, LoE: strong), adoption of verification phase criteria (mean difference = -0.02 [95% CI = -0.06 to 0.03] L·min⁻¹, $p = 0.43$, LoE: strong), verification phase performed on the same or on a separate day as the CPET (mean difference = 0.00 [95% CI = -0.04 to 0.04] L·min⁻¹, $p = 0.97$, LoE: strong), or verification phase duration (mean difference = 0.01 [95% CI -0.05 to 0.07] L·min⁻¹, $p = 0.82$, LoE: strong). All combined categories were not statistically significantly different when analysed separately ($p = 0.28$ to $p = 0.93$, LoE: strong). The exercise mode used did not influence the results for treadmill (mean difference = -0.02 [95% CI -0.07 to 0.03] L·min⁻¹, $p = 0.41$, LoE: moderate) or cycling (mean difference = 0.00 [95% CI -0.04 to 0.04] L·min⁻¹, $p = 0.93$, LoE: strong) across all included studies. The highest $\dot{V}O_2$ attained in the CPET was similar among protocols and their subsequent verification phase ($p = 0.30$; $p = 0.43$ and $p = 0.56$) for ramp, discontinuous, and continuous step-incremented protocols, respectively; LoE: moderate to strong.

Fig 5. Mean differences (95% confidence intervals [CIs]) between the highest $\dot{V}O_2$ values (L·min⁻¹) in the cardiopulmonary exercise test (CPET) and verification phase according to the verification-phase characteristics for work rate (i.e., sub WR_{peak} vs. supra WR_{peak}), recovery (i.e., active vs. passive), adoption of a criterion threshold (i.e., yes vs. no), timing (performed on the same day vs. a different day to the CPET), and duration (i.e., no longer than 80 s, from 81 to 120-s and longer than 120-s).

Participant-level analysis of the highest $\dot{V}O_2$ values attained in the CPET and verification phase

Only 23 (53.5%) of the 43 reviewed studies reported how many participants achieved a lower, equal, or higher $\dot{V}O_2$ value in the verification phase versus the CPET or supplied participant-level $\dot{V}O_2$ data from which this information could be obtained. Table 5 shows the percentages of participants that achieved a lower, equal, or higher $\dot{V}O_2$ value in the verification phase versus the CPET for each study where this information was available. Fig 6 shows participant-level differences between the highest $\dot{V}O_2$ values obtained in the CPET and verification phase for the seven studies where these data were available [31, 44-47, 56, 60].

Table 5. Frequencies and percentages for whether the verification phase (VP) elicited a lower, similar, or higher $\dot{V}O_2$ than the cardiopulmonary exercise test (CPET) for the 23 studies that mentioned or reported this information. Participant-level data were supplied by the authors for the first seven studies in the table.

Study	Population characteristics and sample size	Highest $\dot{V}O_2$ lower, similar, or higher in VP versus CPET?		
		Lower	Similar	Higher
Baumgart, Moes [44] ^a	Para ice hockey athletes (n = 14)	100%	0%	0%
Bhammar, Stickford [45] ^b	Obese children (n = 9) and non-obese children (n = 9)	NR	NR	66.7%
Bhammar, Adams-Huet and Babb [46] ^b	Obese children (n = 21) and non-obese children (n = 23)	NR	NR	NR
Bhammar and Chien [47] ^c	Adults with prehypertension (n = 11)	NR	NR	27.3%
Leicht, Tolfrey [56]	Wheelchair athletes (n = 24)	33.3%	29.2%	37.5%
Michalski, Ferreira [60]	Stroke patients (n = 7)	28.5%	14.4%	57.1%
Saynor, Barker [31]	Cystic fibrosis (n = 14)	21.4%	57.1%	21.4%
Astorino, De La Rosa [43]	Inactive with obesity (n = 17)	NR	NR	30%, 53% and 41% (baseline, post 3 and 6 week)
Bowen, Cannon [49]	Heart failure patients (n = 24)	20.8%	58.4%	20.8%
Causer, Shute [50]	Cystic fibrosis (paediatric n = 14 and adults n = 28)	14.3% 3.6%	64.3% 75%	21.4% 21.4%
de Groot, Takken [52]	Spina bifida (n = 20)	10%	65%	25%
Lambrick, Jakeman [55]	Healthy children (n = 21)	NR	NR	38% (CSI) and 19% (DSI)
Mahoney, Baughman [57]	Adults with obesity (n = 9)	NR	NR	88.9%
Manresa-Rocamora, Fuertes-Kenneally [58]	Heart failure patients (n = 21)	23.8%	52.4%	23.8%
Moreno-Cabañas, Ortega [61]	Metabolic syndrome adults with obesity (n = 100)	0%	60%	40%
Moreno-Cabañas, Ortega [62]	Metabolic syndrome adults with obesity (n = 44)	0%	59%	41%
Sansum, Weston [71]	Healthy children (n = 128)	0%	88%	12%
Sawyer, Tucker [72]	Adults with obesity (n = 19)	NR	NR	68.4%
Schaun, Alberton [32]	Older adults with hypertension (n = 33)	21.2%	24.2%	54.6%
Schaun, Alberton [74]	Older adults with hypertension (n = 12)	8.3% baseline 16.7% post 12-week	16.7% baseline and 8.3% post 12-week	75% baseline and post 12-week
Schneider, Schlüter [75]	Cancer survivors (n = 75)	0%	68%	32%
Villanueva, Campbell [77]	Older adults (n = 22 sub peak and n = 20 supra peak verification phase)	14% sub peak 35% supra peak	68% sub peak 40% supra peak	18% sub peak 25% supra peak
Wood, Hills [80]	Overweight and obese adults (n = 114 ^d)	NR	45.6%	NR

^a 2% threshold criterion used to decide whether the highest $\dot{V}O_2$ was similar in the CPET and verification phase, which was based on the inherent variability of the metabolic cart; ^b Participant-level data were provided but the missing frequencies in the table could not be calculated as an appropriate verification threshold criterion could not be identified; ^c Could not calculate the missing frequencies in the table as the authors used a verification threshold criterion based on the predicted versus measured $\dot{V}O_{2max}$ in the CPET and verification phase; ^d Total study sample was 135 participants of which 114 performed both a CPET and verification phase. CSI = continuous step-incremented; DSI = discontinuous step-incremented.

Fig 6. Participant-level differences between the highest $\dot{V}O_2$ values obtained in the cardiopulmonary exercise test (CPET) and verification phase for the seven studies where these data were available.

Discussion

Main findings

To the best of our knowledge, this is the first systematic review and meta-analysis to investigate the utility of the verification phase for confirming $\dot{V}O_{2max}$ in special and clinical groups. The major findings were: a) overall, the highest $\dot{V}O_2$ attained in the CPET was not statistically significant different to that obtained in the verification phase across all primary studies included; b) subgroup analysis showed there were no statistically significant differences in the highest $\dot{V}O_2$ attained in the CPET and verification phase for specific groups; c) across all studies, the difference between the highest $\dot{V}O_2$ attained in the CPET and verification phase was not affected by test protocol characteristics; d) participant-level verification phase data might be useful for providing evidence of whether the CPET likely elicited $\dot{V}O_{2max}$ in a given person; e) a $\dot{V}O_2$ plateau in the CPET does not always confirm $\dot{V}O_{2max}$ since the verification phase $\dot{V}O_2$ is sometimes higher in the presence of a $\dot{V}O_2$ plateau; and f) the included studies did not report any adverse events associated with the verification phase.

The mean absolute difference in the highest $\dot{V}O_2$ attained in the verification phase and CPET was less than 1% for the 30 studies included in the meta-analysis. This difference was similar to that found in our recent meta-analysis that investigated the utility of the verification phase in apparently healthy adults [29]. The global analysis (see Fig 3) revealed a statistically significantly higher mean $\dot{V}O_2$ in the CPET for only one study [44]. This study involved 14 elite para ice hockey players and was the only study that used a ski ergometer [44] involving an upper-body poling technique. The CPET involved an initial WR associated with a value of 11 on the Borg RPE scale, followed by WR increments of 10 W every 30 s. The verification phase was performed at 110% of WR_{max} achieved in the CPET. The rapid WR increments during the CPET led to a relatively short mean \pm SD test duration of 365 ± 58 s (unpublished data) and a higher peak WR than would have been the case for a longer protocol. Consequently, the verification phase WR might have been excessively high and resulted in a mean \pm SD test duration of only 90 ± 13 s (unpublished data), likely due to early local fatigue of the upper-body musculature. These findings indicate that a longer CPET duration and/or lower WR for the verification phase are required for achieving a verification phase of sufficient length in athletes with lower-limb impairments performing upper-body exercise modalities.

A plausible explanation for the contrasting findings between this study and the other 29 meta-analysed studies is that the verification phase workload was presumably too high to attain similar $\dot{V}O_{2max}$ values because it induced fatigue of the limited active upper body musculature. These results suggest the adoption of a lower percentage of peak work rate might be a better strategy for this specific population.

Paediatric group

Similar $\dot{V}O_{2max}$ values were observed between the CPET and verification phase in paediatric participants. Accordingly, two original studies found no significant difference in the $\dot{V}O_{2max}$ attained in the CPET and verification phase in children [30, 55]. Conversely, a study involving 128 children and adolescents (76 male and 52 females; aged 9.3–17.4 years with and without overweight [71]), reported a significantly higher $\dot{V}O_{2max}$ in the CPET, although the authors reported a strong correlation ($r = 0.94$) between the tests. When comparing the highest $\dot{V}O_2$ values obtained in the CPET and verification phase on an participant-level basis, 88% ($n = 112$) of

children and adolescents had their highest $\dot{V}O_2$ in the CPET verified as their 'true' $\dot{V}O_{2max}$ (< 5% difference between protocols). For the remaining 12% of participants who did not have their $\dot{V}O_{2max}$ verified, the highest $\dot{V}O_2$ recorded in the verification phase was 6-23% higher than in the CPET. The authors recommended the verification phase be used to confirm $\dot{V}O_{2max}$ attainment in paediatric individuals. In another study [46], eight non-obese children had their $\dot{V}O_{2max}$ confirmed in the verification phase (< 3% difference in $\dot{V}O_{2max}$), while 15 achieved a higher $\dot{V}O_2$ (> 3% difference in $\dot{V}O_{2max}$ between protocols) in the verification phase. Consequently, these data suggest the suitability of the verification phase to confirm $\dot{V}O_{2max}$ has been attained in non-obese, overweight, and obese children.

Wheelchair group

The subgroup of people using a wheelchair attained similar highest $\dot{V}O_2$ values between protocols. Of the three studies included in this sub-analysis, two investigated the verification phase as the focus of the study [42, 56]. The first recruited 10 individuals with spinal cord injury, who performed a verification phase at 105% of the peak work rate (WR_{peak}) attained in the CPET, attained highest $\dot{V}O_2$ values that were not significantly different between the CPET and verification phase. Hence, the authors recommended a verification phase to confirm 'true' $\dot{V}O_{2max}$ in people with spinal cord injury. Similarly, 24 trained wheelchair athletes (tetraplegics, paraplegics and non-spinal cord-injured) had their $\dot{V}O_{2max}$ confirmed through a verification phase performed at the same treadmill speed, but a gradient of 0.3% to 0.6% higher than that of the CPET treadmill gradient.

Chronic respiratory group

The verification phase also confirmed $\dot{V}O_{2max}$ in the chronic respiratory group. Fourteen paediatric and 28 adults with cystic fibrosis performed a supra peak verification phase at 110% WR_{max} . Twelve paediatric (85.7%) and 27 adults (96.4%) had their $\dot{V}O_{2max}$ confirmed through the verification phase by applying a threshold of 9% difference between CPET and verification phase protocols, which represented the within-subject variability in $\dot{V}O_{2max}$ for this population [50]. No exercise-induced hypoxemia or perceived discomfort were reported. The authors

concluded that a verification phase should be incorporated to confirm that a $\dot{V}O_{2max}$ has been attained. Sixteen adolescents with mild-to-moderate cystic fibrosis (8 males and 8 females; age 14.6 ± 1.7 years) also achieved similar highest $\dot{V}O_2$ values in the CPET and verification phase [78]. Notably, the verification phase was reported as being well tolerated. These results reinforce the robustness of the verification phase as a tool to confirm $\dot{V}O_{2max}$ in people with chronic respiratory limitations.

Geriatric group

Regarding the geriatric group, our results demonstrated no significance difference in the highest $\dot{V}O_2$ observed in the CPET and verification phase. The lack of difference in $\dot{V}O_{2max}$ between protocols was reported in eight older adults who performed a verification phase at 85% WR_{max} and 23 older adults who exercised at 105% WR_{max} . Moreover, the highest $\dot{V}O_2$ values observed in the CPET and verification phase were highly correlated ($r = 0.99$). The authors emphasised that only minor within-individual differences were observed between the highest $\dot{V}O_2$ elicited during the CPET and verification phase [65]. Another study investigated the applicability of the verification phase performed at 85% ($n = 22$) and 110% ($n = 20$) WR_{max} in older adults [77]. The mean highest $\dot{V}O_2$ across participants was not significantly different between both sub and supra peak verification phases and the associated CPET. Although not included in any of our subgroup analyses, Bowen, Cannon [49] showed that group highest $\dot{V}O_2$ values were similar between CPET and a verification protocol performed at 95% WR_{peak} (14.5 ± 3.0 vs. 14.7 ± 3.1 $mL \cdot kg^{-1} \cdot min^{-1}$) within 24 older adults with chronic heart failure. However, within-subject comparisons confirmed $\dot{V}O_{2max}$ in only 14 of 24 patients. Importantly, the verification phase protocol was cited to accurately and reliably confirm (or refute) $\dot{V}O_{2max}$ along with a measurement sensitivity (95% confidence interval) that is specific for each individual. Likewise, using a similar protocol, Manresa-Rocamora, Fuertes-Kenneally [58] showed $\dot{V}O_{2max}$ was confirmed in 11 (52.4%) patients and not confirmed in 10 (47.6%). The authors concluded that a submaximal verification phase performed at 95% WR_{peak} is a safe and suitable method to determine $\dot{V}O_{2max}$ in patients with heart failure with reduced ejection fraction. However, this primary study was not included in the meta-analysis since the authors reported non-parametric variables related to $\dot{V}O_2$ and

verification phases, and a non-parametric 95% CI for the differences were used to compare the CPET and verification phase.

Metabolic group

Our final group sub-analysis within the metabolic group revealed a trend for an attenuated $\dot{V}O_2$ in the CPET compared to the verification phase in individuals with overweight or obesity, with or without metabolic disease or prehypertension. Notably, all six studies in the subgroup analysis used cycle ergometry. In a trial with four verification phases performed at 80, 90, 100 and 105% WR_{max} , eight of nine men with obesity attained a higher $\dot{V}O_2$ in the verification phase [57]. Although not statistically significant, the highest $\dot{V}O_2$ during the verification phase at 90% WR_{max} was $0.24 \text{ L}\cdot\text{min}^{-1}$ higher than during the CPET, which is a 7% difference. Other data from studies with large samples of individuals who are sedentary, have obesity, and metabolic disease, indicate that the magnitude (3%-9%) and prevalence (40% of people with overweight/obesity) of the underestimation of $\dot{V}O_{2max}$ during a CPET is high. Regardless, the results from the current meta-analysis and these primary studies indicate that the verification phase appears to be a robust method for confirming $\dot{V}O_{2max}$ in individuals with overweight or obesity, with or without metabolic disease or prehypertension.

Verification phase characteristics

Regarding verification phase characteristics, Fig 5 illustrates the combined effects of work rate, recovery mode, use of a verification phase threshold, day of the test, and protocol duration. No significant results were found for the combined analysis, nor when each category was analysed separately. Consequently, a specific verification phase protocol cannot currently be recommended for any of the groups included in the present systematic review. These results agree with those found in apparently healthy adults, where the verification phase was not affected by test protocol characteristics [29]. Although $\dot{V}O_{2max}$ was similar among different protocols, notably, no included study applied a sub WR_{peak} verification phase on a treadmill. It therefore remains unclear if relative

exercise work rate affects the suitability of the verification phase for confirming $\dot{V}O_{2max}$ on a treadmill in special and clinical groups.

Regarding the mode of recovery (i.e., active or passive) between the CPET and verification phase, our meta-analysis demonstrated that the $20 \text{ mL}\cdot\text{min}^{-1}$ mean difference between the highest $\dot{V}O_2$ values observed in the CPET and verification phase was not statistically significant. Similar to a recent systematic review that investigated the utility of the verification phase in apparently healthy adults [29], we did not find any study that compared active and passive recoveries. In the present systematic review, the time between the CPET and verification phase ranged from 4 to 25 minutes in studies where these tests were performed on the same day. Although no single study compared the verification phase conducted on the same day versus a different day to the CPET, no statistically significant difference was found in the present review after combining the results. We therefore recommend either an active or passive recovery after the CPET, with no need for an additional visit on a separate day.

Considering the importance of a verification phase being individually analysed to identify those who confirmed their $\dot{V}O_{2max}$, we suggest the adoption of a threshold criterion to compare the differences between the CPET and verification phase. The threshold criterion value has commonly been based on the reproducibility of $\dot{V}O_{2max}$ during the CPET and specific to the metabolic cart that was used, or an arbitrary 2-3% difference has been used. Alternatively, some studies have calculated the differences between measured and predicted $\dot{V}O_{2max}$ [45-47, 61, 62]. However, we did not obtain significant differences between studies that did or did not apply a threshold criterion. We systematically recorded whether authors of primary studies commented specifically on participant level data. In fact, 22 studies (51%) in the present systematic review discussed participant level differences between the highest $\dot{V}O_2$ attained in the CPET and verification phase. We strongly suggest future researchers report participant-level data, since the absence of any significant mean differences in the highest $\dot{V}O_2$ attained in the CPET and verification phase, may mask individuals who attain a practically significant higher $\dot{V}O_2$ in the verification phase.

The combined effects of verification phase duration (i.e., short < 80 s, medium 81-120 s, and long > 120 s) resulted in no statistically significant difference between the highest $\dot{V}O_2$ observed in the CPET and verification phase. Similarly, a previous study did not find any significant correlations between different durations at 110% WR_{max} for the highest $\dot{V}O_2$ attained in the CPET and verification phase [77]. Furthermore, 12 paediatric participants with cystic fibrosis confirmed their $\dot{V}O_{2max}$ and two showed 9% higher $\dot{V}O_{2max}$ values during a verification phase performed at 110% WR_{max} (76 ± 22 s). A brief duration may be inadequate to allow sufficient time for oxygen uptake to achieve maximal values, especially in individuals having slow O_2 kinetics such as those with metabolic or respiratory disease [81-83]. Verification phases shorter than 80 s might elicit $\dot{V}O_{2max}$, however, we recommend future studies implement strategies to avoid inappropriately short verification phases and early termination, such as the adoption of multistage verification protocols that incorporate submaximal and supra peak (in relation to the CPET WR_{max}) work rates.

Participant-level data

Whether the verification phase confirmed that $\dot{V}O_{2max}$ was likely elicited during the CPET was highly variable across the 23 studies that either reported this information directly, or where participant-level data were available to obtain this information. For example, the verification phase elicited a higher $\dot{V}O_2$ than the CPET in 0% to 88.9% of participants across studies. This large variability between studies is likely due to differences in the CPET and verification phase protocols that were used for the various special and clinical groups involved.

The verification phase failed to confirm the highest $\dot{V}O_2$ in all participants in one study involving para ice hockey players [44], which included a CPET with rapidly-incrementing work rates and a verification phase performed at 110% of the WR_{max} achieved in the CPET. A longer CPET duration with a slower ramp rate and/or a lower WR in the verification phase might be more effective in eliciting a higher $\dot{V}O_2$ when using upper-body exercise modalities in athletes with lower-limb impairments. In other studies where participants elicited a higher $\dot{V}O_2$ in the CPET, the percentage ranged from 3.6% to 35%. The study that observed 35% involved 22 apparently

healthy older adults that performed verification phases at 85% and 110% WR_{max} [77]. The authors concluded that 85% WR_{max} was preferable for this population as it more likely confirmed $\dot{V}O_{2max}$.

Four out of seven people after stroke achieved a $\dot{V}O_2$ that was > 3% higher in the verification phase versus the CPET (range 4.7% to 24.1%) [60]. Similar to the para ice hockey athletes [44], it is unclear whether the same limiting factors during the CPET also occurred during the verification phase, leading to submaximal $\dot{V}O_2$ values in both protocols. Twelve of 18 (66.7%) children elicited a higher $\dot{V}O_2$ in the verification phase compared to the CPET [45]. A notable finding is five of the seven children that exhibited a $\dot{V}O_2$ plateau during the CPET elicited a higher $\dot{V}O_2$ during the verification phase. One study involving nine men with obesity compared verification phases performed at 80%, 90%, 100% and 105% of WR_{max} achieved in the CPET [57]. Eight of the men attained a higher $\dot{V}O_2$ during a verification phase, with the submaximal verification phase performed at 90% WR_{max} eliciting the highest $\dot{V}O_2$ values. These findings are similar to those observed in older adults in that a sub-peak verification phase during testing appears preferable for adults with obesity, at least during cycle ergometry. In another study including adults with obesity [72], a verification phase performed at 100% WR_{max} elicited a $\dot{V}O_2$ that was 2% to 21% higher than in the CPET in 13 out of 19 participants. A plausible explanation for these findings is that the CPET work rate increment of 30 $W \cdot \text{min}^{-1}$ for men was inappropriately high for this group, resulting in submaximal $\dot{V}O_2$ values during the CPET. Two studies included older adults with hypertension [32, 74]. In the first study [32], 18 out of 33 participants achieved a higher $\dot{V}O_2$ in the verification phase compared to the value obtained in the CPET (ranging from 3 to 22.1%). Notably, 29 out of 33 tests would have been validated as eliciting $\dot{V}O_{2max}$ if the traditional criteria were utilised (i.e. false positives). In the second study [74], 9 out of 12 participants obtained a verification-derived $\dot{V}O_2$ that was 4.9 to 21% higher than the value observed in the CPET at baseline and after a 12 week exercise program. All participants but one would have had their tests validated as eliciting $\dot{V}O_{2max}$ if traditional criteria were applied. Results from the present systematic review therefore question the validity of the $\dot{V}O_2$ plateau as the primary criterion for establishing the occurrence of $\dot{V}O_{2max}$ [13] and highlights the importance of performing a verification phase even in the presence of a $\dot{V}O_2$ plateau in the CPET.

Midgley et al. [12] suggested that it is unlikely that a person would give identical submaximal efforts in the CPET and verification phase, thereby rationalising that similar highest $\dot{V}O_2$ values in the CPET and verification phase likely verifies that $\dot{V}O_{2max}$ has been elicited. Whether this supposition is true is debatable, especially in certain clinical populations where exercise might be terminated due to symptoms related to the clinical condition. Examples are people with pulmonary disease where undue breathlessness typically results in test termination before any cardiovascular limitation [84], severely deconditioned individuals with very poor exercise tolerance, and those conditions associated with chronic pain made worse with exercise. In such people, the verification phase might simply verify the reproducibility of the $\dot{V}O_2$ associated with symptom tolerance (i.e., $\dot{V}O_{2peak}$). This is likely also true for CPET involving upper body ergometry, where the CPET is often limited by peripheral fatigue [85].

Limitations and recommendations for future research

A limitation of the present systematic review is that 13 (30%) of the 43 studies included in the systematic review were not included in the meta-analysis due to unsuccessful attempts to acquire the required unpublished information from the corresponding authors for these studies. Some authors responded to our correspondence declaring they no longer have access to the requested data. In regards group-level data, the verification phase appears useful for helping verify that, on average, a specific CPET protocol elicited $\dot{V}O_{2max}$ in special and clinical groups. A verification phase therefore seems unnecessary for this purpose if a new study uses the same CPET protocol with the same population as a previous study that confirmed the CPET was appropriate for eliciting $\dot{V}O_{2max}$ using a verification phase.

Another limitation of the present systematic review is that participant-level data were available for only seven of the 43 reviewed studies, limiting inferences that could be made. We recommend that participant-level data are made available for future studies. Further research is needed to optimise verification phase protocols so that evidence-based guidelines can be published in the future. This includes identifying robust verification threshold criteria to establish when the highest $\dot{V}O_2$ values elicited in the CPET and verification phases are sufficiently

similar for $\dot{V}O_{2\max}$ or $\dot{V}O_{2\text{peak}}$ to be 'verified'. Few studies have directly compared test protocols and this is a particularly important area for future research.

Further critical discussion is also needed among researchers regarding how the verification phase should be interpreted when applied to participant-level data. For example, In the first edition of the Canadian Association of Sports Sciences guidelines for the physiological testing of high-performance athletes, Thoden et al. [86] suggested that the highest $\dot{V}O_2$ value elicited in either the CPET or verification phase should be regarded as $\dot{V}O_{2\max}$. In the second edition of the guidelines, however, Thoden et al. [87] recommended that an increase in the highest $\dot{V}O_2$ elicited in the verification phase that is not more than 2% higher than that elicited in the CPET, verifies $\dot{V}O_{2\max}$ was elicited in the CPET. At the participant level, the verification phase can therefore be used to verify whether $\dot{V}O_{2\max}$ was likely elicited in the CPET or used simply as another opportunity for a person to elicit $\dot{V}O_{2\max}$.

Conclusion

The main finding was that the mean difference in $\dot{V}O_2$ between the CPET and associated verification phase was similar. In other words, the verification phase confirmed that $\dot{V}O_{2\max}$ had been attained in the CPET. Moreover, a 10-to-15-minute recovery phase and short verification phase following a CPET appears to be safe, well-tolerated, and time-efficient for a diverse range of special and clinical groups. Unlike traditional $\dot{V}O_{2\max}$ criteria, the application of the verification phase was not affected by differences in test protocol or procedures, or participant characteristics, except perhaps for those with overweight or obesity. In those with overweight or obesity, the $\dot{V}O_2$ attained in the CPET was significantly lower than that obtained in the verification phase. For individuals with obesity and apparently healthy older adults, the selection of sub peak work rates above critical power is desirable. For the remaining groups, it remains somewhat unclear which work rate is most suitable. For paediatrics, a verification phase of 100-105% WR_{\max} conducted 15-min after the CPET on a cycle or treadmill has been useful. For the wheelchair group, we can only advise to use a verification phase < 110% WR_{\max} . Regarding

individuals with chronic respiratory problems, a cycle ergometer is applicable for a verification phase of 3-min at 20 W, then 110% WR_{max} , performed after 5-min of active rest and 10-min of passive rest after the CPET. Older adults, including those with chronic heart failure, may perform a verification phase 5-10 min after the CPET, either on a cycle ergometer or treadmill. If cycling is chosen, a work rate equal to 85-95% WR_{peak} would optimise $\dot{V}O_{2max}$ attainment. On a treadmill, 2-min at 50%, 1-min at 70%, and then one stage higher than WR_{peak} is recommended. Lastly, adults with obesity or metabolic conditions may perform a verification phase above critical power on a cycle ergometer up to 110% WR_{peak} on the CPET.

Some researchers might decide not to conduct a verification phase where results from the present meta-analysis indicate that particular CPET test protocols and procedures for certain special and clinical groups appear to elicit 'true' $\dot{V}O_{2max}$ values. However, there remains the issue of identifying whether a participant has likely elicited $\dot{V}O_{2max}$ during a CPET or elicited a submaximal $\dot{V}O_2$ value due to early test termination. Considering the limitations associated with traditional $\dot{V}O_{2max}$ criteria, the verification phase remains a reasonable time-efficient alternative in special and clinical populations; however, further research and critical debate are required before robust evidence-based guidelines can be provided.

Acknowledgments

We sincerely thank the following authors for kindly providing additional data: Dr. Tony Babb, Dr. Dharini Bhammar, Prof. James Faulkner, Dr. Christof Leicht, Prof. Melitta McNarry, Dr. Friederike Rosenberger, Dr. Brandon Sawyer, Dr. Zoe Saynor, Justine Schneider, Dr. Owen Tomlinson, and Prof. Craig Williams.

References

1. Fletcher GF, Ades PA, Kligfield P, Arena R, Balady GJ, Bittner VA, et al. Exercise standards for testing and training: a scientific statement from the American Heart Association. *Circulation*. 2013;128(8):873-934. doi: 10.1161/CIR.0b013e31829b5b44.

2. Hill AV, Lupton H. Muscular exercise, lactic acid, and the supply and utilization of oxygen. *QJM: An International Journal of Medicine*. 1923;os-16(62):135-171. doi: <https://doi.org/10.1093/qjmed/os-16.62.135>.
3. Hill AV, Long CNH, Lupton H. Muscular exercise, lactic acid and the supply and utilisation of oxygen.—Parts VII–VIII. *Proc R Soc Lond B Biol Sci*. 1924;97(682):155-176.
4. ATS/ACCP. American Thoracic Society/American College of Chest Physicians statement on cardiopulmonary exercise testing. *Am J Respir Crit Care Med*. 2003;167(2):211-277. doi: 10.1164/rccm.167.2.211.
5. McMurray RG, Ainsworth BE, Harrell JS, Griggs TR, Williams OD. Is physical activity or aerobic power more influential on reducing cardiovascular disease risk factors? *Med Sci Sports Exerc*. 1998;30(10):1521-1529. doi: 10.1097/00005768-199810000-00009.
6. Weber KT, Janicki JS. Cardiopulmonary exercise testing for evaluation of chronic cardiac failure. *Am J Cardiol*. 1985;55(2):A22-A31. doi: 10.1016/0002-9149(85)90792-1.
7. Riebe D, Ehrman JK, Liguori G, Magal M. ACSM's guidelines for exercise testing and prescription. 10th ed. Philadelphia: Wolters Kluwer; 2018.
8. da Cunha FA, Farinatti PTV, Midgley AW. Methodological and practical application issues in exercise prescription using the heart rate reserve and oxygen uptake reserve methods. *J Sci Med Sport*. 2011;14(1):46-57. doi: 10.1016/j.jsams.2010.07.008.
9. Weatherwax R, Harris N, Kilding AE, Dalleck L. Time course changes in confirmed 'true' $\text{VO}_{2\text{max}}$ after individualized and standardized training. *Sports Med Int Open*. 2019;3(02):E32-E39. doi: 10.1055/a-0867-9415.
10. Cooper CB, Storer TW. Exercise testing and interpretation: a practical approach. Cambridge, UK: Cambridge University Press; 2001.
11. Ross R, Blair SN, Arena R, Church TS, Després JP, Franklin BA, et al. Importance of assessing cardiorespiratory fitness in clinical practice: a case for fitness as a clinical vital sign: a scientific statement from the American Heart Association. *Circulation*. 2016;134(24):e653-e699. doi: 10.1161/CIR.0000000000000461.
12. Midgley AW, McNaughton LR, Polman R, Marchant D. Criteria for determination of maximal oxygen uptake: a brief critique and recommendations for future research. *Sports Med*. 2007;37(12):1019-1028. doi: 10.2165/00007256-200737120-00002.
13. Taylor HL, Buskirk E, Henschel A. Maximal oxygen intake as an objective measure of cardio-respiratory performance. *J Appl Physiol*. 1955;8(1):73-80. doi: 10.1152/jappl.1955.8.1.73.

14. Davis JA, Whipp BJ, Lamarra N, Huntsman DJ, Frank MH, Wasserman K. Effect of ramp slope on determination of aerobic parameters from the ramp exercise test. *Med Sci Sports Exerc.* 1982;14(5):339-343.
15. Buchfuhrer MJ, Hansen JE, Robinson TE, Sue DY, Wasserman K, Whipp BJ. Optimizing the exercise protocol for cardiopulmonary assessment. *J Appl Physiol Respir Environ Exerc Physiol.* 1983;55(5):1558-1564. doi: 10.1152/jappl.1983.55.5.1558.
16. Myers J, Buchanan N, Walsh D, Kraemer M, McAuley P, Hamilton-Wessler M, Froelicher VF. Comparison of the ramp versus standard exercise protocols. *J Am Coll Cardiol.* 1991;17(6):1334-1342. doi: 10.1016/s0735-1097(10)80144-5.
17. Keir DA, Paterson DH, Kowalchuk JM, Murias JM. Using ramp-incremental VO_2 responses for constant-intensity exercise selection. *Appl Physiol Nutr Metab.* 2018;43(9):882-892. doi: 10.1139/apnm-2017-0826.
18. Day JR, Rossiter HB, Coats EM, Skasick A, Whipp BJ. The maximally attainable VO_2 during exercise in humans: the peak vs. maximum issue. *J Appl Physiol.* 2003;95(5):1901-1907. doi: 10.1152/jappphysiol.00024.2003.
19. Myers J, Walsh D, Sullivan M, Froelicher V. Effect of sampling on variability and plateau in oxygen uptake. *J Appl Physiol* 1990;68(1):404-410. doi: 10.1152/jappl.1990.68.1.404.
20. Poole DC, Wilkerson DP, Jones AM. Validity of criteria for establishing maximal O_2 uptake during ramp exercise tests. *Eur J Appl Physiol.* 2008;102(4):403-410. Epub 2007/10/31. doi: 10.1007/s00421-007-0596-3. PubMed PMID: 17968581.
21. Rossiter HB, Kowalchuk JM, Whipp BJ. A test to establish maximum O_2 uptake despite no plateau in the O_2 uptake response to ramp incremental exercise. *J Appl Physiol* 2006;100(3):764-770. doi: 10.1152/jappphysiol.00932.2005.
22. Midgley AW, Carroll S, Marchant D, McNaughton LR, Siegler J. Evaluation of true maximal oxygen uptake based on a novel set of standardized criteria. *Appl Physiol Nutr Metab.* 2009;34(2):115-123. doi: 10.1139/H08-146.
23. Astorino TA. Alterations in $\text{VO}_{2\text{max}}$ and the VO_2 plateau with manipulation of sampling interval. *Clin Physiol Funct Imaging.* 2009;29(1):60-67. doi: 10.1111/j.1475-097X.2008.00835.x.
24. Gordon D, Mehter M, Gernigon M, Caddy O, Keiller D, Barnes R. The effects of exercise modality on the incidence of plateau at $\text{VO}_{2\text{max}}$. *Clin Physiol Funct Imaging.* 2012;32(5):394-399. doi: 10.1111/j.1475-097X.2012.01142.x.
25. Edvardsen E, Hem E, Anderssen SA. End criteria for reaching maximal oxygen uptake must be strict and adjusted to sex and age: a cross-sectional study. *PLoS One.* 2014;9(1):e85276. doi: 10.1371/journal.pone.0085276.
26. Howley ET, Bassett DR, Welch HG. Criteria for maximal oxygen uptake: review and commentary. *Med Sci Sports Exerc.* 1995;27:1292-1301.

27. Poole DC, Jones AM. Measurement of the maximum oxygen uptake VO_{2max} : VO_{2peak} is no longer acceptable. *J Appl Physiol* 2017;122(4):997-1002. doi: 10.1152/jappphysiol.01063.2016.
28. Midgley AW, Carroll S. Emergence of the verification phase procedure for confirming 'true' VO_{2max} . *Scand J Med Sci Sports*. 2009;19(3):313-322. doi: 10.1111/j.1600-0838.2009.00898.x.
29. Costa VAB, Midgley AW, Carroll S, Astorino TA, de Paula T, Farinatti P, Cunha FA. Is a verification phase useful for confirming maximal oxygen uptake in apparently healthy adults? A systematic review and meta-analysis. *PLoS One*. 2021;16(2):e0247057. doi: 10.1371/journal.pone.0247057.
30. Barker AR, Williams CA, Jones AM, Armstrong N. Establishing maximal oxygen uptake in young people during a ramp cycle test to exhaustion. *Br J Sports Med*. 2011;45(6):498-503. doi: 10.1136/bjism.2009.063180.
31. Saynor ZL, Barker AR, Oades PJ, Williams CA. A protocol to determine valid VO_{2max} in young cystic fibrosis patients. *J Sci Med Sport*. 2013;16(6):539-544. doi: 10.1016/j.jsams.2013.01.010.
32. Schaun GZ, Alberton CL, Gomes MLB, Santos LP, Bamman MM, Mendes GF, et al. Maximal oxygen uptake is underestimated during Incremental testing in hypertensive older adults: Findings from the HAEL study. *Med Sci Sports Exerc*. 2021;53(7):1452-1459. doi: 10.1249/MSS.0000000000002598.
33. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4(1):1-9. doi: 10.1186/2046-4053-4-1.
34. Nelson ME, Rejeski WJ, Blair SN, Duncan PW, Judge JO, King AC, et al. Physical activity and public health in older adults: recommendation from the American College of Sports Medicine and the American Heart Association. *Circulation*. 2007;116(9):1094. doi: 10.1161/CIRCULATIONAHA.107.185650.
35. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health*. 1998;52(6):377-384. doi: 10.1136/jech.52.6.377.
36. Hébert-Losier K, Supej M, Holmberg HC. Biomechanical factors influencing the performance of elite alpine ski racers. *Sports Med*. 2014;44(4):519-533. doi: 10.1007/s40279-013-0132-z.
37. Baumgart JK, Brurok B, Sandbakk Ø. Comparison of peak oxygen uptake between upper-body exercise modes: A systematic literature review and meta-analysis. *Front Physiol*. 2020;11:412. doi: 10.3389/fphys.2020.00412.

38. Van Tulder M, Furlan A, Bombardier C, Bouter L, Group EBotCCBR. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. *Spine*. 2003;28(12):1290-1299. doi: 10.1097/01.BRS.0000065484.95996.AF.
39. Higgins JP, Li T, Deeks JJ. Choosing effect measures and computing estimates of effect. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. *Cochrane handbook for systematic reviews of interventions*. Oxford, UK: John Wiley & Sons; 2019. pp. 143-176.
40. Hardy RJ, Thompson SG. Detecting and describing heterogeneity in meta-analysis. *Stat Med*. 1998;17(8):841-856. doi: 10.1002/(sici)1097-0258(19980430)17:8<841::aid-sim781>3.0.co;2-d.
41. Armstrong N, Welsman J, Winsley R. Is peak VO₂ a maximal index of children's aerobic fitness? *Int J Sports Med*. 1996;17(05):356-359. doi: 10.1055/s-2007-972860.
42. Astorino TA, Bediamol N, Cotoia S, Ines K, Koeu N, Menard N, et al. Verification testing to confirm VO_{2max} attainment in persons with spinal cord injury. *J Spinal Cord Med*. 2019;42(4):494-501. doi: 10.1080/10790268.2017.1422890.
43. Astorino TA, De La Rosa AB, Clark A, De Revere JL. Verification testing to confirm VO_{2max} attainment in inactive women with obesity. *Int J Exerc Sci*. 2020;13(4):1448.
44. Baumgart JK, Moes M, Skovereng K, Ettema G, Sandbakk Ø. Examination of gas exchange and blood lactate thresholds in Paralympic athletes during upper-body poling. *PLoS One*. 2018;13(10):e0205588. doi: 10.1371/journal.pone.0205588.
45. Bhammar DM, Stickford JL, Bernhardt V, Babb TG. Verification of maximal oxygen uptake in obese and nonobese children. *Med Sci Sports Exerc*. 2017;49(4):702-710. doi: 10.1249/MSS.0000000000001170.
46. Bhammar DM, Adams-Huet B, Babb TG. Quantification of cardiorespiratory fitness in children with obesity. *Med Sci Sports Exerc*. 2019;51(11):2243-2250. doi: 10.1249/MSS.0000000000002061.
47. Bhammar DM, Chien LC. Quantification and verification of cardiorespiratory fitness in adults with prehypertension. *Sports*. 2021;9(1):9. doi: 10.3390/sports9010009.
48. Blazquez A, Guillamo E, Alegre J, Ventura JL, Barbany JR, Javierre C. Cardiorespiratory responses to exercise in female chronic fatigue syndrome patients affect the psychological health of their partners. *JEPonline*. 2011;14(2):1-16.

49. Bowen TS, Cannon DT, Begg G, Baliga V, Witte KK, Rossiter HB. A novel cardiopulmonary exercise test protocol and criterion to determine maximal oxygen uptake in chronic heart failure. *J Appl Physiol* 2012;113(3):451-458. doi: 10.1152/jappphysiol.01416.2011.
50. Causer AJ, Shute JK, Cummings MH, Shepherd AI, Bright V, Connett G, et al. Cardiopulmonary exercise testing with supramaximal verification produces a safe and valid assessment of VO_{2max} in people with cystic fibrosis: a retrospective analysis. *J Appl Physiol* 2018;125(4):1277-1283. doi: 10.1152/jappphysiol.00454.2018.
51. Cockcroft EJ, Bond B, Williams CA, Harris S, Jackman SR, Armstrong N, Barker AR. The effects of two weeks high-intensity interval training on fasting glucose, glucose tolerance and insulin resistance in adolescent boys: a pilot study. *BMC Sports Sci Med Rehabil.* 2019;11(1):1-9. doi: 10.1186/s13102-019-0141-9.
52. de Groot JF, Takken T, de Graaff S, Gooskens RHJM, Helders PJM, Vanhees L. Treadmill testing of children who have spina bifida and are ambulatory: does peak oxygen uptake reflect maximum oxygen uptake? *Phys Ther.* 2009;89(7):679-687. doi: 10.2522/ptj.20080328.
53. Goosey-Tolfrey VL, Paulson TAW, Tolfrey K, Eston RG. Prediction of peak oxygen uptake from differentiated ratings of perceived exertion during wheelchair propulsion in trained wheelchair sportspersons. *Eur J Appl Physiol.* 2014;114(6):1251-1258. doi: 10.1007/s00421-014-2850-9.
54. Lambrick D, Bertelsen H, Eston R, Stoner L, Faulkner J. Prediction of peak oxygen uptake in children using submaximal ratings of perceived exertion during treadmill exercise. *Eur J Appl Physiol.* 2016;116(6):1189-1195. doi: 10.1007/s00421-016-3377-z.
55. Lambrick D, Jakeman J, Grigg R, Kaufmann S, Faulkner J. The efficacy of a discontinuous graded exercise test in measuring peak oxygen uptake in children aged 8 to 10 years. *Biol Sport.* 2017;34(1):57. doi: 10.5114/biolSport.2017.63734.
56. Leicht CA, Tolfrey K, Lenton JP, Bishop NC, Goosey-Tolfrey VL. The verification phase and reliability of physiological parameters in peak testing of elite wheelchair athletes. *Eur J Appl Physiol.* 2013;113(2):337-345. doi: 10.1007/s00421-012-2441-6.
57. Mahoney JM, Baughman BR, Sheard AC, Sawyer BJ. Determining the optimal workrate for cycle ergometer verification phase testing in males with obesity. *Sports.* 2021;9(2):30. doi: 10.3390/sports9020030.
58. Manresa-Rocamora A, Fuertes-Kenneally L, Blasco-Peris C, Sempere-Ruiz N, Sarabia JM, Climent-Paya V. Is the verification phase a suitable criterion for the determination of maximum oxygen uptake in patients with heart failure

- and reduced ejection fraction? A validation study. *Int J Environ Res Public Health*. 2023;20(4):2764. doi: 10.3390/ijerph20042764.
59. McCreery JL, Mackintosh KA, Mills-Bennett R, McNarry MA. The effect of a high-intensity PrO₂Fit inspiratory muscle training intervention on physiological and psychological health in adults with bronchiectasis: A mixed-methods study. *Int J Environ Res Public Health*. 2021;18(6):3051. doi: 10.3390/ijerph18063051.
 60. Michalski AC, Ferreira AS, Midgley AW, Costa VAB, Fonseca GF, da Silva NSL, et al. Mixed circuit training acutely reduces arterial stiffness in patients with chronic stroke: a crossover randomized controlled trial. *Eur J Appl Physiol*. 2023;123(1):121-134. doi: 10.1007/s00421-022-05061-8.
 61. Moreno-Cabañas A, Ortega JF, Morales-Palomo F, Ramirez-Jimenez M, Mora-Rodriguez R. Importance of a verification test to accurately assess VO_{2max} in unfit individuals with obesity. *Scand J Med Sci Sports*. 2020;30(3):583-590. doi: 10.1111/sms.13602.
 62. Moreno-Cabañas A, Ortega JF, Morales-Palomo F, Ramirez-Jimenez M, Alvarez-Jimenez L, Pallares JG, Mora-Rodriguez R. The use of a graded exercise test may be insufficient to quantify true changes in $\dot{V}O_{2max}$ following exercise training in unfit individuals with metabolic syndrome. *J Appl Physiol* 2020;129(4):760-767. doi: 10.1152/jappphysiol.00455.2020.
 63. Murias JM, Kowalchuk JM, Paterson DH. Mechanisms for increases in VO_{2max} with endurance training in older and young women. *Med Sci Sports Exerc*. 2010;42(10):1891-1898. doi: 10.1249/MSS.0b013e3181dd0bba.
 64. Murias JM, Kowalchuk JM, Paterson DH. Time course and mechanisms of adaptations in cardiorespiratory fitness with endurance training in older and young men. *J Appl Physiol* 2010;108(3):621-627. doi: 10.1152/jappphysiol.01152.2009.
 65. Murias JM, Pogliaghi S, Paterson DH. Measurement of a true VO_{2max} during a ramp incremental test is not confirmed by a verification phase. *Front Physiol*. 2018;9:143. doi: 10.3389/fphys.2018.00143.
 66. Oliveira RS, Barker AR, Debras F, Kranen SH, Williams CA. Effects of exercise intensity on vascular and autonomic components of the baroreflex following glucose ingestion in adolescents. *Eur J Appl Physiol*. 2019;119(4):867-878. doi: 10.1007/s00421-019-04076-y.
 67. Pal A, Schneider J, Schlüter K, Steindorf K, Wiskemann J, Rosenberger F, Zimmer P. Different endurance exercises modulate NK cell cytotoxic and inhibiting receptors. *Eur J Appl Physiol*. 2021;121(12):3379-3387. doi: 10.1007/s00421-021-04735-z.

68. Paulson TAW, Goosey-Tolfrey VL, Lenton JP, Leicht CA, Bishop NC. Spinal cord injury level and the circulating cytokine response to strenuous exercise. *Med Sci Sports Exerc.* 2013;45(9):1649-1655. doi: 10.1249/MSS.0b013e31828f9bbb.
69. Robben KE, Poole DC, Harms CA. Maximal oxygen uptake validation in children with expiratory flow limitation. *Pediatr Exerc Sci.* 2013;25(1):84-100. doi: 10.1123/pes.25.1.84.
70. Rowland TW. Does peak VO_2 reflect VO_{2max} in children?: Evidence from supramaximal testing. *Med Sci Sports Exerc.* 1993;25(6):689-693.
71. Sansum KM, Weston ME, Bond B, Cockcroft EJ, O'Connor A, Tomlinson OW, et al. Validity of the supramaximal test to verify maximal oxygen uptake in children and adolescents. *Pediatr Exerc Sci.* 2019;31(2):213-222. doi: 10.1123/pes.2018-0129.
72. Sawyer BJ, Tucker WJ, Bhammar DM, Gaesser GA. Using a verification test for determination of VO_{2max} in sedentary adults with obesity. *J Strength Cond Res.* 2015;29(12):3432-3438. doi: 10.1519/JSC.0000000000001199.
73. Saynor ZL, Barker AR, Oades PJ, Williams CA. Reproducibility of maximal cardiopulmonary exercise testing for young cystic fibrosis patients. *J Cyst Fibros.* 2013;12(6):644-650. doi: 10.1016/j.jcf.2013.04.012.
74. Schaun GZ, Alberton CL, Brizio Gomes ML, Mendes GF, Häfele MS, Andrade LS, et al. Exercise intervention does not reduce the likelihood of VO_{2max} underestimation in older adults with hypertension. *J Sports Sci.* 2022;40(12):1399-1405. doi: 10.1080/02640414.2022.2081403.
75. Schneider J, Schlüter K, Wiskemann J, Rosenberger F. Do we underestimate maximal oxygen uptake in cancer survivors? Findings from a supramaximal verification test. *Appl Physiol Nutr Metab.* 2020;45(5):486-492. doi: 10.1139/apnm-2019-0560.
76. Tomlinson OW, Barker AR, Chubbock LV, Stevens D, Saynor ZL, Oades PJ, Williams CA. Analysis of oxygen uptake efficiency parameters in young people with cystic fibrosis. *Eur J Appl Physiol.* 2018;118(10):2055-2063. doi: 10.1007/s00421-018-3926-8.
77. Villanueva IR, Campbell JC, Medina SM, Jorgensen TM, Wilson SL, Angadi SS, et al. Comparison of constant load exercise intensity for verification of maximal oxygen uptake following a graded exercise test in older adults. *Physiol Rep.* 2021;9(18):e15037. doi: 10.14814/phy2.15037.
78. Werkman MS, Hulzebos HJ, van de Weert-van PB, Arets HGM, Helders PJM, Takken T. Supramaximal verification of peak oxygen uptake in adolescents with cystic fibrosis. *Pediatr Phys Ther.* 2011;23(1):15-21. doi: 10.1097/PEP.0b013e318208ca9e.

79. Woloschuk A, Hodges GJ, Massarotto RJ, Klentrou P, Falk B. The skin blood flow response to exercise in boys and men and the role of nitric oxide. *Eur J Appl Physiol.* 2020;120(4):753-762. doi: 10.1007/s00421-019-04286-4.
80. Wood RE, Hills AP, Hunter GR, King NA, Byrne NM. VO_{2max} in overweight and obese adults: do they meet the threshold criteria? *Med Sci Sports Exerc.* 2010;42(3):470-477. doi: 10.1249/MSS.0b013e3181b666ad.
81. Combret Y, Medrinal C, Prieur G, Robledo Quesada A, Gillot T, Gravier FE, et al. Oxygen uptake kinetics during treadmill walking in adolescents with clinically stable cystic fibrosis. *Physiother Theory Pract.* 2022;38(10):1389-1397. Epub 2021/01/05. doi: 10.1080/09593985.2020.1868029. PubMed PMID: 33390080.
82. O'Connor E, Kiely C, O'Shea D, Green S, Egaña M. Similar level of impairment in exercise performance and oxygen uptake kinetics in middle-aged men and women with type 2 diabetes. *Am J Physiol Regul Integr Comp Physiol.* 2012;303(1):R70-76. Epub 2012/04/28. doi: 10.1152/ajpregu.00012.2012. PubMed PMID: 22538515.
83. Wolpat A, Lima FV, Silva FM, Tochetto M, de Freitas A, Grandi T, et al. Association between inspiratory muscle weakness and slowed oxygen uptake kinetics in patients with chronic obstructive pulmonary disease. *Appl Physiol Nutr Metab.* 2017;42(12):1239-1246. Epub 2017/07/28. doi: 10.1139/apnm-2016-0568. PubMed PMID: 28750180.
84. Boutou AK, Zafeiridis A, Pitsiou G, Dipla K, Kioumis I, Stanopoulos I. Cardiopulmonary exercise testing in chronic obstructive pulmonary disease: An update on its clinical value and applications. *Clin Physiol Funct Imaging.* 2020;40(4):197-206. Epub 20200402. doi: 10.1111/cpf.12627. PubMed PMID: 32176429.
85. Loughney L, West M, Pintus S, Lythgoe D, Clark E, Jack S, Torella F. Comparison of oxygen uptake during arm or leg cardiopulmonary exercise testing in vascular surgery patients and control subjects. *B J Anaesth.* 2014;112(1):57-65.
86. Thoden JS, Wilson BA, MacDougall JD. Testing aerobic power. In: MacDougall JD, Wenger HA, Green HJ, editors. *Physiological testing of the elite athlete.* Ithaca, NY: Movement Publications, Inc; 1982. pp. 39-60.
87. Thoden JS, Wislon BA, MacDougall JD. Testing aerobic power. In: MacDougall JD, Wenger HA, Green HJ, editors. *Physiological testing of the high-performance athlete.* 2nd ed. Champaign, IL: Human Kinetics; 1991. pp. 107-173.

Supporting information

S1 Fig. PRISMA checklist.

S2 Fig. Search strategies for all databases.

S3 Fig. Modified Downs and Black checklist.

Figure 1

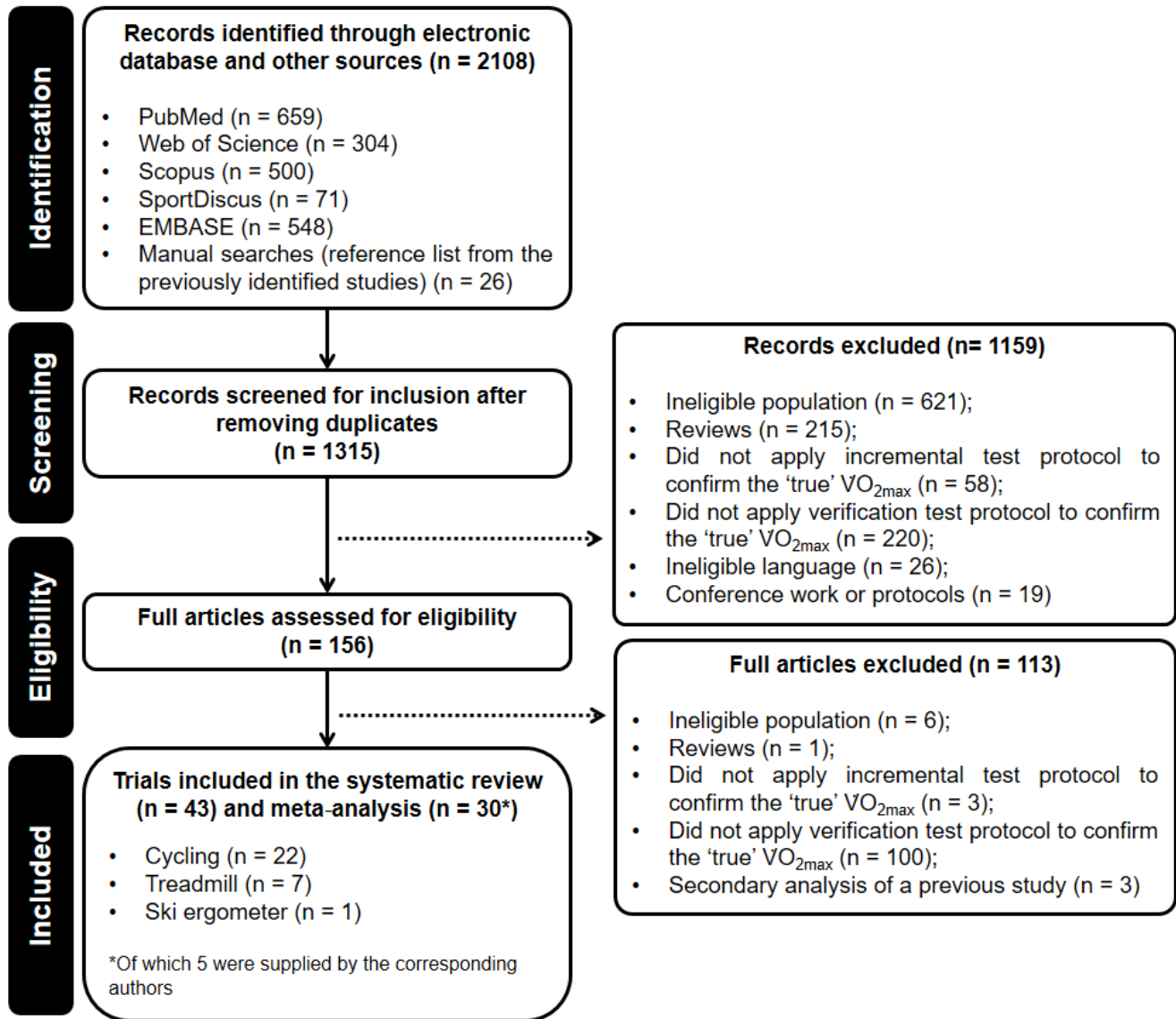


Figure 2

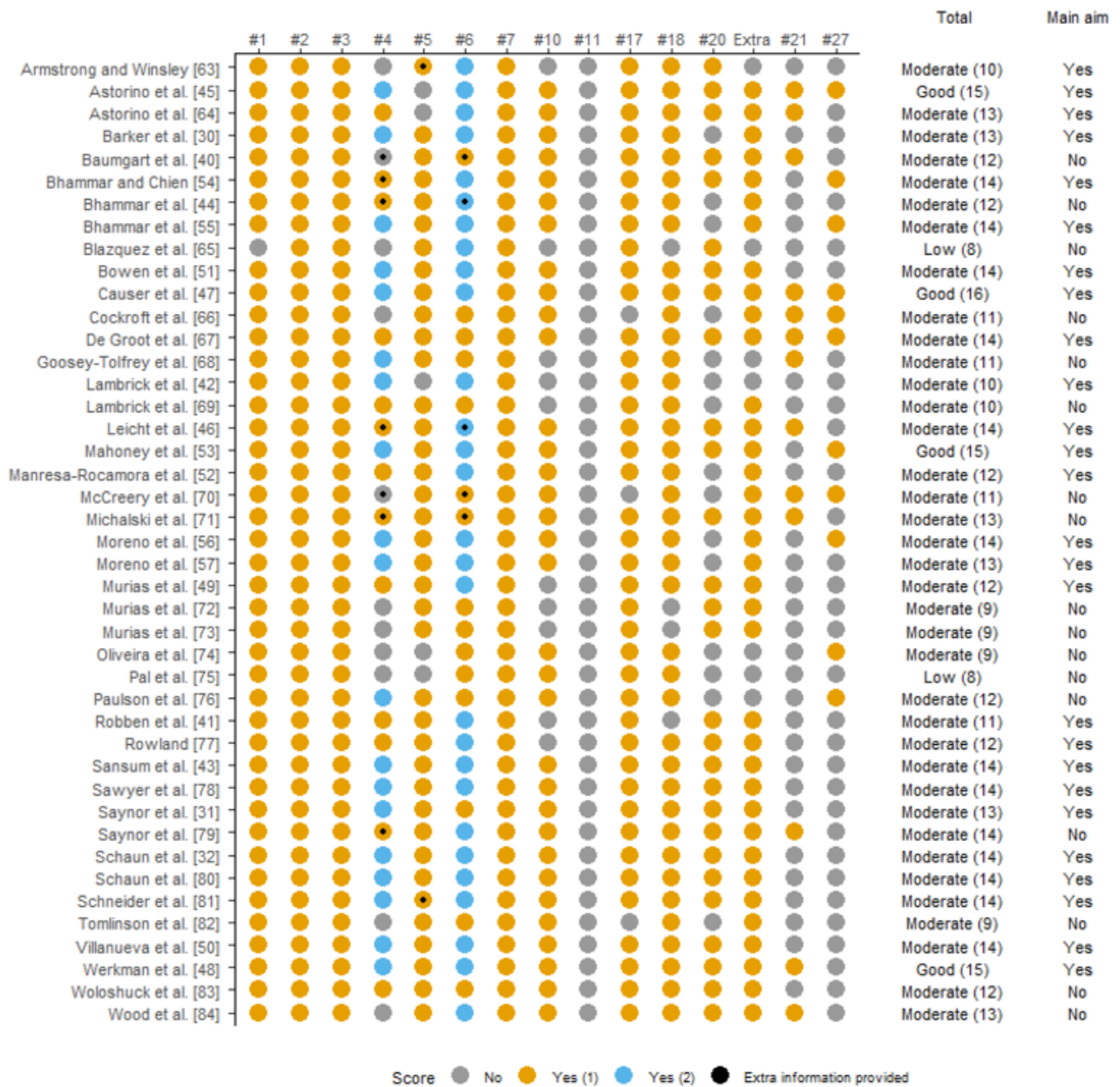


Figure 3

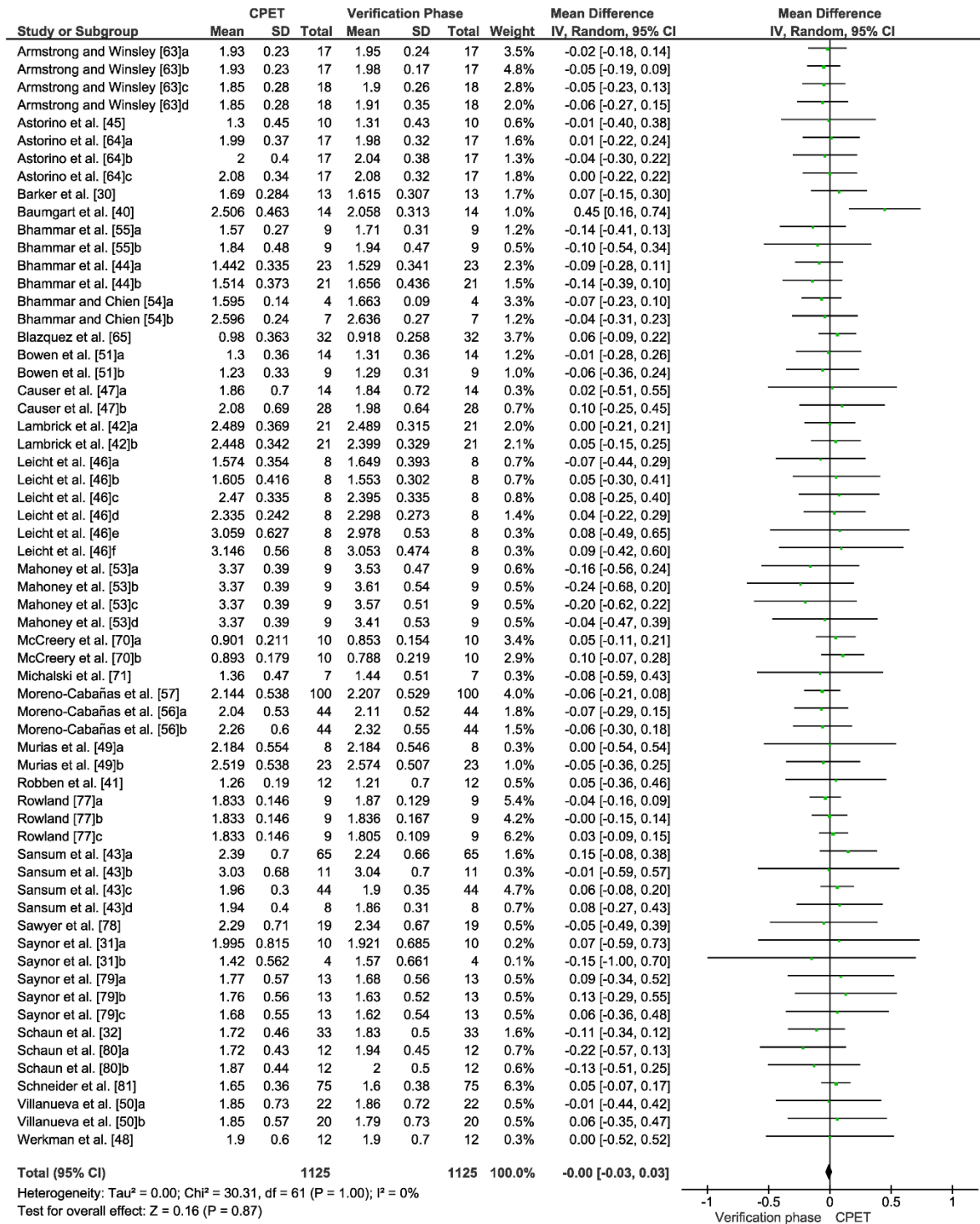


Figure 5

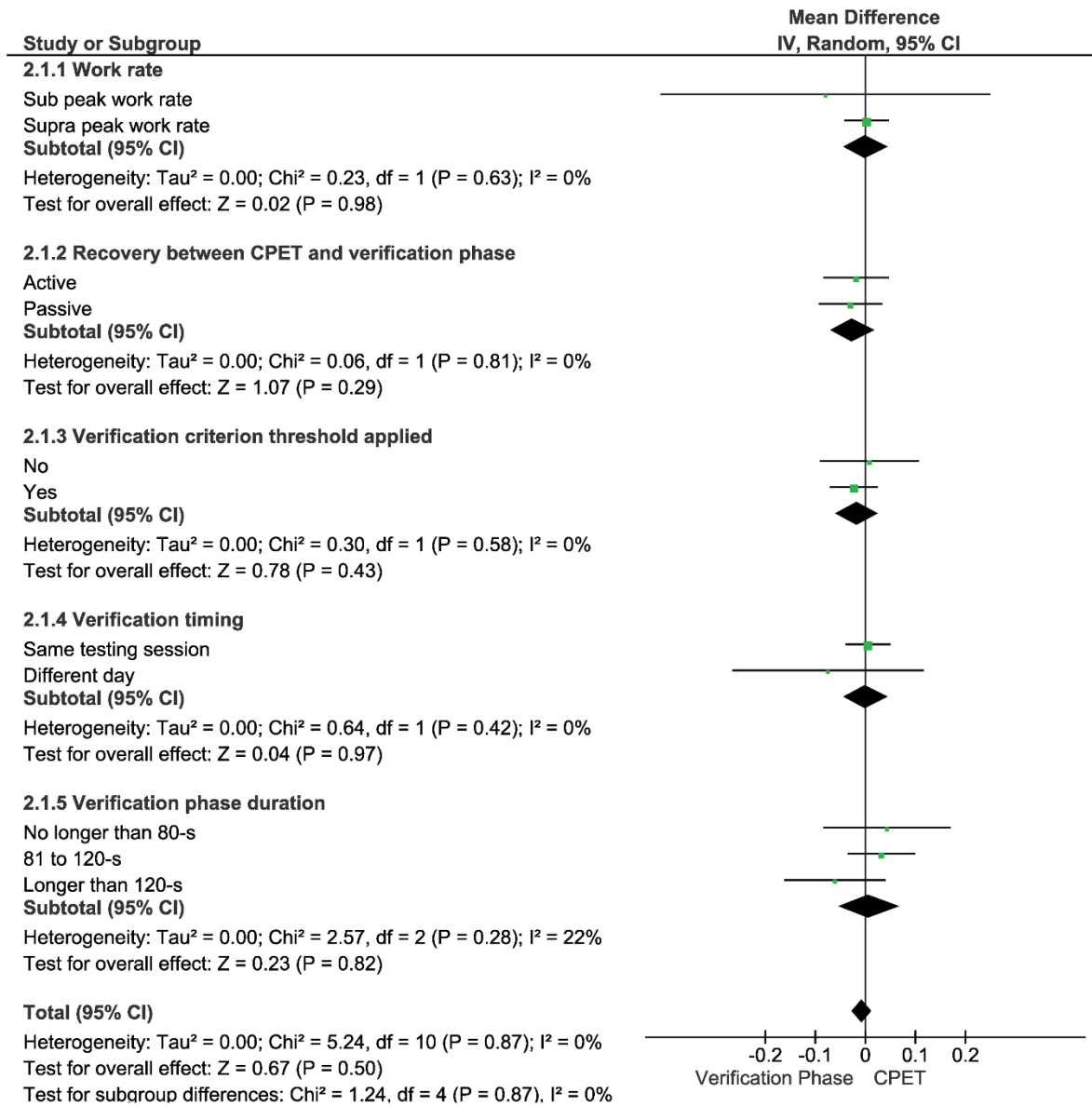
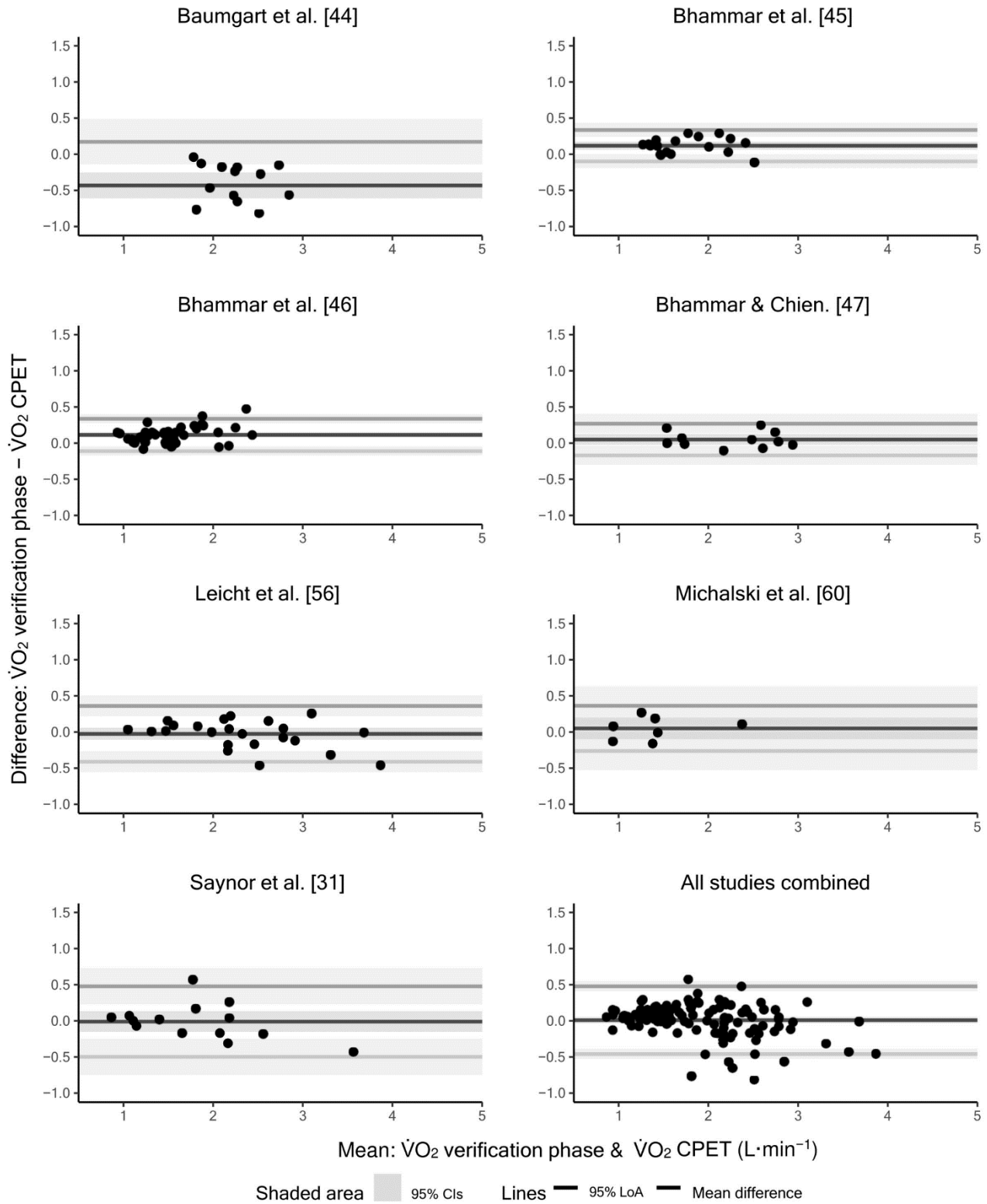


Figure 6



Supplemental file 1 – PRISMA flow diagram

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Confirming the attainment of maximal oxygen uptake in special and clinical populations: a systematic review and meta-analysis of verification phase protocols	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	4-6
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	6-7
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	6-7
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	7
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	7
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	7
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	7
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	7-8
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	7-8
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data	8-9

		synthesis.	
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	8-9
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I ²) for each meta-analysis.	8-9

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	8-9
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	8-9
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	9-10
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	10-35
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	N/A
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	26-35
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	27-35
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	34
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	35-39
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	40

Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	40-41
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	1

Supplemental file 2 – Search strategies for all databases

5. Search Strategy for MEDLINE (Pubmed):

((“Exercise Test”[mh] OR “Exercise Test” OR “Exercise Testing” OR “Treadmill Test” OR “Treadmill Testing” OR “Bicycle Test” OR “Bicycle Testing” OR “Cardiopulmonary Exercise Test” OR “Cardiopulmonary Exercise Testing” OR “CPET” OR “Incremental Test” OR “Incremental Testing” OR “Incremental Exercise” OR “Graded-Exercise Test” OR “Graded-Exercise Testing” OR “GXT” OR “Step-incremented” OR “Step-exercise” OR “Ramp Test” OR “Ramp Testing” OR “Ramp-incremental”) AND (“Verification Phase” OR “Verification Test” OR “Verification Testing” OR “Verification Bout” OR “Verification Stage” OR “Verification Criteria” OR “Supramaximal Test” OR “Square-wave” OR “Exhaustive Test” OR “Exhaustive Testing” OR “CWR” OR “RISE”)) AND (“Oxygen Consumption”[mh] OR “Oxygen Consumption” OR “Oxygen Uptake” OR “VO2” OR “VO2 max” OR “VO2max” OR “VO2 peak” OR “VO2peak”) Sort by: Most Recent

2. Search Strategy for Web of Science:

TÓPICO: (“Exercise Test” OR “Exercise Testing” OR “Treadmill Test” OR “Treadmill Testing” OR “Bicycle Test” OR “Bicycle Testing” OR “Cardiopulmonary Exercise Test” OR “Cardiopulmonary Exercise Testing” OR “CPET” OR “Incremental Test” OR “Incremental Testing” OR “Incremental Exercise” OR “Graded-Exercise Test” OR “Graded-Exercise Testing” OR “GXT” OR “Step-incremented” OR “Step-exercise” OR “Ramp Test” OR “Ramp Testing” OR “Ramp-incremental”) AND TÓPICO: (“Verification Phase” OR “Verification Test” OR “Verification Testing” OR “Verification Bout” OR “Verification Stage” OR “Verification Criteria” OR “Supramaximal Test” OR “Square-wave” OR “Exhaustive Test” OR “Exhaustive Testing” OR “CWR” OR “RISE”) AND TÓPICO: (“Oxygen Consumption” OR “Oxygen Uptake” OR “VO2” OR “VO2 max” OR “VO2max” OR “VO2 peak” OR “VO2peak”)
Tempo estipulado: Todos os anos. Índices: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.

3. Search Strategy for SCOPUS:

“Exercise Test” OR “Exercise Testing” OR “Treadmill Test” OR “Treadmill Testing” OR “Bicycle Test” OR “Bicycle Testing” OR “Cardiopulmonary Exercise Test” OR “Cardiopulmonary Exercise Testing” OR “CPET” OR “Incremental Test” OR “Incremental Testing” OR “Incremental Exercise” OR “Graded-Exercise Test” OR “Graded-Exercise Testing” OR “GXT” OR “Step-incremented” OR “Step-exercise” OR “Ramp Test” OR “Ramp Testing” OR “Ramp-incremental” AND “Verification Phase” OR “Verification Test” OR “Verification Testing” OR “Verification Bout” OR “Verification Stage” OR “Verification Criteria” OR “Supramaximal Test” OR “Square-wave” OR “Exhaustive Test” OR “Exhaustive Testing” OR “CWR” OR “RISE” AND “Oxygen Consumption” OR “Oxygen Uptake” OR “VO2” OR “VO2 max” OR “VO2max” OR “VO2 peak” OR “VO2peak” AND NOT INDEX (medline)

4. Search Strategy for SPORTDiscus (Ebsco):

(“Exercise Test” OR “Exercise Testing” OR “Treadmill Test” OR “Treadmill Testing” OR “Bicycle Test” OR “Bicycle Testing” OR “Cardiopulmonary Exercise Test” OR “Cardiopulmonary Exercise Testing” OR “CPET” OR “Incremental Test” OR “Incremental Testing” OR “Incremental Exercise” OR “Graded-Exercise Test” OR “Graded-Exercise Testing” OR “GXT” OR “Step-incremented” OR “Step-exercise” OR “Ramp Test” OR “Ramp Testing” OR “Ramp-incremental”) AND (“Verification Phase” OR “Verification Test” OR “Verification Testing” OR “Verification Bout” OR “Verification Stage” OR “Verification Criteria” OR “Supramaximal Test” OR “Square-wave” OR “Exhaustive Test” OR “Exhaustive Testing” OR “CWR” OR “RISE”) AND (“Oxygen Consumption” OR “Oxygen Uptake” OR “VO2” OR “VO2 max” OR “VO2max” OR “VO2 peak” OR “VO2peak”)

5. Search Strategy for EMBASE

(‘exercise test’ OR ‘exercise testing’ OR ‘treadmill test’ OR ‘treadmill testing’ OR ‘bicycle test’ OR ‘bicycle testing’ OR ‘cardiopulmonary exercise test’ OR ‘cardiopulmonary exercise testing’ OR ‘cpet’ OR ‘incremental test’ OR ‘incremental testing’ OR ‘incremental exercise’ OR ‘graded-exercise test’ OR ‘graded-exercise testing’ OR ‘gxt’ OR ‘step-incremented’ OR ‘step-exercise’ OR ‘ramp test’ OR ‘ramp testing’ OR ‘ramp-incremental’) AND (‘verification phase’ OR ‘verification test’ OR ‘verification testing’ OR ‘verification bout’ OR ‘verification stage’ OR ‘verification criteria’ OR ‘supramaximal test’ OR ‘square-wave’ OR ‘exhaustive test’ OR ‘exhaustive

testing' OR 'cwr' OR 'rise') AND ('oxygen consumption' OR 'oxygen uptake' OR 'vo2' OR 'vo2 max' OR 'vo2max' OR 'vo2 peak' OR 'vo2peak')

Supplemental file 3 – Modified Downs and Black checklist

Color coding

Black = original questions

Blue = original comments

~~Blue~~ = we deleted parts of the original comments

Green = specifications/adjustments made for this review

Note that the following words have been changed from the original format:

- “Patient” was replaced with “participant”
- “Principal confounders” was replaced with “participant characteristics”

	Yes	No	U	N/A	Comments
Reporting					
1. <i>Is the hypothesis/aim/objective of the study clearly described? This item is only rated as a Yes, if the aim <u>OR</u> purpose <u>OR</u> hypothesis are clearly described. The aim <u>OR</u> purpose <u>OR</u> hypothesis do not necessarily need to include the cardiopulmonary exercise test (CPET) to exhaustion and/or verification phase, as we acknowledge that this may not be the main aim <u>OR</u> purpose of the included studies.</i>					
2. <i>Are the main outcomes to be measured clearly described in the introduction or methods section? If the main outcomes are first mentioned in the results section, the question should be answered with a no.</i>					
3. <i>Are the characteristics of the <u>participants</u> included in the study clearly described? In cohort studies and trials, inclusion and/or exclusion criteria should be given. In case-control studies, a case-definition and the source for controls should be given. It needs to be clear that all study participants belong to one of the special populations targeted in this review (i.e., children, older adults, people with a disability, people with obesity). If it this is clear and it is clear from the inclusion criteria which participants are excluded, rate this item as a Yes. If this is not clear and the study does not mention any exclusion criteria, rate this item as a No.</i>					
4. <i>Are the interventions of interest clearly described? Treatments and placebo (where relevant) that are to be compared should be clearly described. For this review the CPET to exhaustion can be considered the control/ placebo and the verification phase the treatment.</i> For the CPET, the following information needs to be provided: - Duration of stages and magnitude of increments - Duration of the test For the verification phase, the following information needs to be provided: - If the test was sub- or supra-peak or at the peak power output/speed obtained during the maximal test to exhaustion including the work rate/ speed at which it was performed					

<p>- Duration of the test</p> <p>Furthermore, the following information needs to be included:</p> <ul style="list-style-type: none"> - Exercise modality used for both tests - Time and type of recovery between tests (if performed on separate tests, it needs to be specified that the testing was performed at the approximately same time of day) <p>Both tests need to be described in sufficient detail for this item to be ranked as a yes (2 points) or yes (1 point). Two points are given if all six items are described. One point is given if four-five of the six items are described.</p>					
<p>5. <i>Are the participant characteristics distribution of principal confounders in each group of subjects to be compared clearly described?</i> A list of participant characteristics is provided. This contains sex, age, body height and body mass of the participants. For studies on wheelchair users, body height does not need to be provided. In addition, for the clinical populations, sufficient detail needs to be provided in the type of disability/disease.</p>					
<p>6. <i>Are the main findings of the study clearly described?</i> Simple outcome data (including denominators and numerators) should be reported for all major findings so that the reader can check the major analyses and conclusions. (This question does not cover statistical tests, which are considered below). In addition, oxygen uptake (VO₂) and its unit of measurement need to be provided for both the maximal test to exhaustion and the verification phase.</p> <p>Both (i.e., simple outcome data for the major findings + VO₂ and its unit of measurement for both tests) need to be sufficiently described for this item to a yes with 2 points. Only providing the Δ% or absolute difference in VO₂ for both tests is not sufficient. If only one (i.e., simple outcome data for the major findings + VO₂ and its unit of measurement for both tests) is sufficiently described, it is rated as a yes with 1 point.</p>					
<p>7. <i>Does the study provide estimates of the random variability in the data for the main outcomes?</i> In non normally distributed data the inter-quartile range of results should be reported. In normally distributed data the standard error, standard deviation or confidence intervals should be reported. If the distribution of data is not described, it must be assumed that the estimates used were appropriate and the question should be answered yes. If no mean and SD are calculated and only individual data is provided, rate this item as a No. If it is a case-study this item should be rated as N/A. In addition, for VO₂ a measure of central tendency (for example mean or median) and dispersion (for example standard deviation or interquartile range needs to be provided).</p>					

<p>10. Have actual probability values been reported (e.g., 0.035 rather than <0.05) for the main outcomes except where probability is less than 0.001? This refers to the main outcomes of the study and not necessarily to the comparison of CPET to exhaustion and verification phase.</p>					
<p>External validity All the following criteria attempt to address the representativeness of the findings of the study and whether they may be generalised to the population from which the subjects were derived.</p>					
<p>11. Were the subjects asked to participate in the study representative of the entire population from which they were recruited?</p> <p>The study must identify the source population for participants and describe how the participants were selected recruited. It must be clear what the source population is, and if study participants can be considered representative for this population (e.g., through random sampling). If it is likely that the study participants were selected by convenience sampling, rate this item as a no. Participants would be representative if they comprised the entire source population, an unselected sample of consecutive participants, or a random sample. Random sampling is only feasible where a list of all members of the relevant population exists. Where a study does not report the proportion of the source population from which the patients are derived, the question should be answered unable to determine.</p>					
<p>Internal validity – bias</p>					
<p>17. In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients, or in case-control studies, is the time period between the intervention and outcome the same for cases and controls?</p> <p>Where follow-up was the same for all study patients the answer should be yes. If different lengths of follow-up were adjusted for by, for example, survival analysis the answer should be yes. Studies where differences in follow-up are ignored should be answered no. For this review the maximal test to exhaustion can be considered the control/ placebo and the verification phase the treatment. The time period between the maximal test to exhaustion and the verification phase(s) needs to be specified and standardized within each included study for this item to be rated as a yes. If it is specified that the tests were performed on separate days rate this item as yes.</p>					
<p>18. Were the statistical tests used to assess the main outcomes appropriate?</p> <p>The statistical techniques must be appropriate for the data. For example non-parametric methods should be use for small sample sizes. Where little statistical analysis has been undertaken but where there is no evidence of bias, the question should be answered yes. If the distribution of the data (normal or not) is not described it must be assumed that the estimates used were appropriate and the question should be answered yes. At a minimum, the following needs to be described for this</p>					

<p>item to be ranked as a yes: 1) the specific type of test (only stating that an ANOVA/ T-test or correlations was performed without further specification is insufficient) and 2) what specific outcome measures the test(s) target(s).</p>					
<p>20. <i>Were the main outcome measures used accurate (valid and reliable)?</i> For studies where the outcome measures are clearly described, the question should be answered yes. For studies which refer to other work or that demonstrates the outcome measures are accurate, the question should be answered yes. Information on the calibration procedures of the metabolic cart/ ergospirometer needs to be provided. The statement “was calibrated according to the manufacturer’s specifications/instructions” is sufficient. Referring to another study or the manufacturers instructions for the calibration procedures is also sufficient for this item to be ranked as a yes.</p>					
<p><i>Extra item. Was the sampling method of the ergometer (i.e., mixing chamber or breath-by-breath) and the sampling interval specified?</i></p>					
Internal validity – confounding (selection bias)					
<p>21. <i>Were the participants in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited from the same population?</i> For example, participants for all comparison groups should be selected from the same hospital. The question should be unable to determine for cohort and case-control studies where there is no information concerning the source of patients in the study. The source population in this review is defined as individuals with the same type of special populations targeted in this review (i.e., children, older adults, people with a disability) at a distinct institution/in a distinct area (region, county, country).</p>					
Power					
<p>27. <i>Did the study have sufficient power to detect a clinically important effect where the probability value for a difference due to chance is less than 5%?</i> Sample sizes have been calculated to detect a difference from x% and y%. Prospective sample size calculations need to be included in the study for this item to be ranked as a yes. This item may also be ranked as a yes, if a reference is made to another study where sample size calculations were appropriately performed.</p>					
Extra item (not included in the quality appraisal score)					
<p><i>Extra item. Was the comparison between the incremental test and verification stage a main outcome of the study?</i></p>					