Sports Med (2016) 46:737–750
DOI 10.1007/s40279-015-0445-1

SYSTEMATIC REVIEW

Validity of Submaximal Step Tests to Estimate Maximal Oxygen Uptake in Healthy Adults

Hunter Bennett[1] · **Gaynor Parfitt**[1] · **Kade Davison**[1] · **Roger Eston**[1]

Abstract

Background Aerobic capacity ($V\text{O}_2$max) is a strong predictor of health and fitness and is considered a key physiological measure in the healthy adult population. Submaximal step tests provide a safe, simple and ecologically valid means of assessing $V\text{O}_2$max in both the general population and a rehabilitation setting. However, no studies have attempted to synthesize the existing knowledge regarding the validity of the multiple step-test protocols available to estimate $V\text{O}_2$max in the healthy adult population.

Objectives The objective of this study was to systematically review literature on the validity and reliability of submaximal step-test protocols to estimate $V\text{O}_2$max in healthy adults (age 18–65 years).

Data Sources and Study Selection A systematic literature search of the MEDLINE, EMBASE, Scopus, Web of Science, and Cochrane Library databases was performed. The search returned 690 studies that underwent the initial screening process. To be included, the study had to (1) have participants deemed to be healthy and aged between 18 and 65 years; (2) assess $V\text{O}_2$max by means of a submaximal step test against a graded exercise test (GXT) to volitional exhaustion; and (3) be available in English. Reference lists from included articles were screened for additional articles.

Data Analysis and Study Appraisal Methods The primary outcome measures used were the validity statistics between the actual measured $V\text{O}_2$max and predicted $V\text{O}_2$max values, and the reported direction of the statistically significant difference between the measured $V\text{O}_2$max and the predicted $V\text{O}_2$max. The Quality Assessment Tool for Quantitative Studies was used to assess the risk of bias in each included study, and was adapted to the type of quantitative study design used.

Results The combined database search produced 690 studies, from which 644 were excluded during the screening process. Following full-text assessment, a further 39 studies were excluded based on the eligibility criteria detailed previously. Four additional studies were located via the reference lists of the included studies, leaving 11 studies that fulfilled the inclusion criteria and which compared eight different step-test protocols against a direct measure of $V\text{O}_2$max incurred during a maximal GXT. Validity measures varied, with a broad range of correlation coefficients reported across the 11 studies ($r = 0.469$–0.95). Of the 11 studies, two reported reliability measures, demonstrating good test–retest reliability [mean -0.8 ± 3.7 mL kg^{-1} min^{-1} (±7.7 % of the mean measured $V\text{O}_2$max)].

Conclusions Considering the relationship between $V\text{O}_2$max and various markers of health, the use of step tests as a measure of health in both the general adult population and rehabilitation settings is advocated. Step tests provide a simple, effective and ecologically valid method of submaximally assessing $V\text{O}_2$max that can be implemented in a variety of situations within the general adult population. Future research is needed to assess the reliability of the majority of the step-test procedures reviewed. Based on the validity measures, submaximal step-test protocols are an acceptable means of estimating VO2max in the generally

---

✉ Roger Eston
roger.eston@unisa.edu.au

[1] Alliance for Research in Exercise, Nutrition and Activity, Sansom Institute for Health Research, University of South Australia, Adelaide, SA, Australia

healthy adult population. For tracking changes in cardiorespiratory fitness, the Chester Step test appears to be an appropriate tool due to its high test–retest reliability.

Key Points

Validity of step-test protocols vary significantly but they appear to provide high reliability on a test–retest basis.

Step tests provide a practical, submaximal method of assessing the cardiorespiratory fitness of healthy adults.

1 Introduction

Cardiorespiratory fitness is a health-related component of physical fitness, requiring the integration of the circulatory, respiratory, and muscular systems to supply oxygen to the working tissue during physical activity [1]. There is a considerable body of evidence suggesting that poor cardiorespiratory fitness is associated with increased risk of morbidity and mortality in both men and women through various cardiovascular and metabolic risk factors [2]. The evaluation and maintenance of cardiorespiratory fitness could therefore be considered an integral component to stem declines in aerobic capacity and reduce associated risk to health.

Maximal oxygen uptake ($V\mathrm{O}_2$max) is the highest oxygen uptake that can be achieved despite increases in intensity of exercise [3]. It is the highest rate at which oxygen can be taken in and utilised by the body, and provides measures of cardiorespiratory fitness and cardiovascular health and function [4, 5]. It is the most commonly applied measure to indicate a change in aerobic capacity as a result of training [6].

Direct measurement of $V\mathrm{O}_2$max is often obtained from a graded exercise test (GXT) requiring exercise to volitional exhaustion, with the expired air of the individual undergoing analysis [7]. This protocol is time-consuming, expensive, ecologically invalid in real-world settings, and induces high physical stress [7–9].

Submaximal exercise testing is therefore commonly used to predict $V\mathrm{O}_2$max when time is limited, laboratory equipment is unavailable, or it may be considered unsafe to exercise at high intensities [5, 10, 11]. In this regard, step tests are inexpensive, simple, portable and an ecologically valid means of estimating $V\mathrm{O}_2$max submaximally [12]. They provide a safe and practical method of assessing cardiorespiratory fitness under submaximal conditions and therefore offer high potential to be used to assess health in the general adult population, and in a rehabilitation setting. Their capacity has been demonstrated successfully as a tool to assess cardiorespiratory fitness in fire brigades in Britain, USA, Europe and Asia [13], and in primary care and home settings, in adults of varying fitness levels across broad age ranges in Canada [14–17].

There are a wide variety of step-test protocols which differ in stepping frequency, test duration and number of test stages [18]. Despite their simplicity and ease of use, it has been suggested that protocols utilising a fixed step height or fixed step rate may produce a less accurate estimation of cardiorespiratory fitness. Use of a fixed step height is common in step tests, but as leg length differs significantly between different people, the energy required to perform each step varies [19, 20], and a step that is too high for a particular individual may infer a mechanical disadvantage, and may therefore be more likely to be dependent on muscular endurance than cardiorespiratory fitness. Conversely, a step that is too small may not provide adequate resistance to stimulate the required cardiorespiratory response. Step tests that use a fixed cadence have also produced a higher exercise intensity during the test in individuals with higher body mass index, lower body height and lower exercise capacity [18]. In some scenarios, fixed-rate step tests may result in vigorous exercise intensities, eliminating their benefit as a submaximal exercise test as medical supervision might be required, and may impact the accuracy and validity of the test.

Additionally, the estimation of $V\mathrm{O}_2$max from step tests is typically determined through the combination of establishing an absolute intensity estimate of $V\mathrm{O}_2$ for any given step rate at a particular step height, and then extrapolating that $V\mathrm{O}_2$ to a maximum by way of corresponding this value to a relative exercise intensity using a percentage of maximal heart rate (HR), which is typically based on an age-predicted maximal HR. To date, with the exception of one study assessing the validity and reliability of the Chester step test (CST) [13] and recommendations for estimating $V\mathrm{O}_2$max from the assumed relationships between HR and rating of perceived exertion (RPE) during stepping [21], no current step-test protocols appear to have used RPE as a means of representing this relative exercise intensity component, as observed when using cycle or arm ergometry and treadmill exercise [22–28].

The purpose of this systematic review was to assess the validity and reliability of submaximal step-test protocols as methods to estimate $V\mathrm{O}_2$max in healthy adults (18–65 years) against a validated measure of $V\mathrm{O}_2$max.

2 Methods

2.1 Search Strategy

A systematic analysis of the scientific literature was undertaken to find as many studies as possible that reported the validity of submaximal step tests to estimate maximal oxygen consumption. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [29] for systematic reviews was used. Candidate studies published between 1950 and 2015 were searched for between 2 March and 13 April 2015 by first searching a number of relevant online databases (Table 1).

The following keyword string was used to search the MEDLINE, EMBASE, Scopus, Web of Science, and Cochrane Library databases: (step test) AND ((Maximal Oxygen Uptake) OR peak oxygen uptake) OR VO_2max) OR aerobic capacity) OR aerobic fitness) OR aerobic power) OR cardiorespiratory fitness.

All titles and abstracts returned from the search were assessed to identify suitable articles. When the title and abstract provided insufficient data to ensure an article's eligibility, the full-text paper was retrieved and analysed.

2.2 Inclusion/Exclusion Criteria

Studies were included if they explicitly reported on the validity of a submaximal step-test protocol to estimate VO_2max in comparison with measured VO_2max (mL kg^{-1} min^{-1}); the actual VO_2max (or VO_2peak) was directly measured using a gas analysis system; participants were apparently healthy (asymptomatic of disease and free from acute or chronic injury); and the age of the participants were reported to be between 18 and 65 years.

Studies were excluded if they did not compare estimated VO_2max from a step-test protocol to a directly measured VO_2max; the population included children and elderly; the population included adults deemed unhealthy, either through a diagnosed clinical condition or considered obese; the population included adult athletes; intervention-based studies; or they did not report one validity statistic and either a predicted and measured VO_2max value or a directional difference between the measured and predicted VO_2max value.

No limits were placed on the year of publication but only published, full-text manuscripts written in the English language were included.

The reference lists of all retained studies were then examined in an attempt to locate further studies.

2.3 Quality Assessment

The Quality Assessment Tool for Quantitative Studies [30] was used to assess risk of bias in each included study. This is a standardised critical appraisal tool, which consists of six sections: selection, study design, confounding factors, blinding, data collection method (reliability and validity), and withdrawal. The tool can be adapted dependent on the type of quantitative study design used [31]. As the included studies were cross-sectional in nature, the following sections were excluded: study design, confounding factors, and blinding, as they principally relate to interventional study designs [32]. For each individual study, each of the components were rated as 'strong', 'moderate', or 'weak', based on standardised criteria. These ratings were then combined to attain an overall rating for each study ('strong', no weak ratings; 'moderate', one weak rating; and 'weak', two or more weak ratings). No studies were excluded on the basis of risk of bias assessment.

2.4 Data Extraction

The authors independently extracted the data (the quality criteria, participant details, validity measures, reliability measures and main conclusion). Discrepancies were resolved by referring back to the initial paper during an in-depth discussion.

2.5 Data Analysis

The primary outcome measures used were the validity statistics between the actual measured VO_2max and predicted VO_2max values, and the reported direction of the statistically significant difference between the measured VO_2max and the predicted VO_2max.

3 Results

Figure 1 details the study selection process. The combined database search produced 690 results, from which 644 were excluded during the screening process. Following full-text assessment, a further 39 studies were excluded based on the eligibility criteria detailed previously. Four additional studies were located via the reference lists of the included studies, leaving a total of 11 studies for analysis.

Table 1 provides a summary of the 11 included studies reporting submaximal step test to predict VO_2max, and detailing study sample information, the step-test protocol utilised, the predictive equations used, the variables included in those equations, the step height used, validity, the direction of difference between the measured and predicted VO_2max values, and the reported reliability.

Sample sizes ranged from 13 to 80 participants. Three study samples comprised only female participants [33–35], three studies comprised only male participants [36–38], whilst the remainder comprised both male and female

Table 1 Studies included in this review

Study	No. of participants (sex)	Age, years (mean ± SD, or range)	Protocol	Equation	Variables	Step height (cm)	Mean measured VO_2max	Mean predicted VO_2max	Validity	Direction of prediction difference	Reliability
Buckley et al. [13]	13 (7M, 6F)	22.4 ± 4.6	Chester step test	Chester step test graphical data sheet—line of best fit	HR, RPE	30	48.2 ± 7.7 mL kg^{-1} min^{-1}	T1: 45.4 ± 8.1 mL kg^{-1} min^{-1} T2: 46.4 ± 9.2 mL kg^{-1} min^{-1}	Chester step test 1 $p = 0.006$ Chester step test 2 (nonsignificant)	−	−0.8 ± 3.7 mL kg^{-1} min^{-1}
Sykes and Roberts [12]	68, sex not specified	30.6 ± 9.7	Chester step test	Chester step test graphical data sheet—line of best fit	HR, RPE	30	54.5 ± 8.7 mL kg^{-1} min^{-1}	T1: 53.2 ± 7.7 mL kg^{-1} min^{-1} T2: 53.9 ± 7.6 mL kg^{-1} min^{-1}	$r = 0.92$	=	−0.7 mL kg^{-1} min^{-1}
Webb et al. [20]	80 (38M, 42F)	18–29	Individualised protocol	$VO_2max = 45.938 + 9.253(G) - 0.140$ (body mass) + 0.670 (PFA score) + 0.429 (final step rate) − 0.149 (rested HR 45 s post-test)	Body mass, HR, step rate, PFA score	Individualised	47.6 ± 7.7 mL kg^{-1} min^{-1}	NA	$r = 0.81$	=	NR
Knight et al. [39]	40, sex not specified	43 ± 14	STEP tool protocol	$VO_2max = 3.9 + (1511/time) \times [(body\ mass/HR) \times 0.124] - (age \times 0.032) - (sex \times 0.633)$	Time, body mass, HR, age, sex	20	43.3 ± 9.7 mL kg^{-1} min^{-1}	43.9 ± 10.8 mL kg^{-1} min^{-1}	$r = 0.78$	+	NR
Chatterjee et al. [36]	30M	22.6 ± 0.2	Queen's College step test	$VO_2max = 111.332 - (0.426 \times HR$ in beats/min)	HR	41.3	39.8 mL kg^{-1} min^{-1}	39.3 mL kg^{-1} min^{-1}	−0.07 mL kg^{-1} min^{-1} $r = 0.95$	−	NR
Chatterjee et al. [33]	40F	21.9 ± 3.2	Queen's College step test	$VO_2max = 65.81 - (0.1847 \times HR$ in beats/min)	HR	41.3	32.8 ± 3.8 mL kg^{-1} min^{-1}	35.5 ± 4.4 mL kg^{-1} min^{-1}	+2.7 mL kg^{-1} min^{-1} ($p < 0.001$)	+	NR
McArdle et al. [34]	41F	19–23	Queen's College step test	$VO_2max = 65.81 - (0.1847 \times HR$ in beats/min)	HR	41.3	38.1 ± 3.8 mL kg^{-1} min^{-1}	NA	$r = 0.75$	±	NR
			Skubic and Hodgkins step test	$VO_2max = 55.92 - (0.1517 \times HR$ in beats/min)		45.7			$r = 0.64$	±	NR
Perroni et al. [37]	15M	31 ± 6	Queen's College step test	$VO_2max = 111.332 - (0.426 \times HR$ in beats/min)	HR	40	52.8 ± 5.7 mL kg^{-1} min^{-1}	49.2 ± 7.6 mL kg^{-1} min^{-1}	$r = 0.47$	+	NR
Francis and Culpepper [35]	17F	19–33	Height-adjusted, rate-specific, single-stage step test	$VO_2max = 71.97 - (0.776 \times HR$ in beats/min)	HR	Individualised	43.4 ± 4.6 mL kg^{-1} min^{-1}	NA	$r = 0.80$	±	NR
Astrand and Ryhming [38]	18M	18–19	Astrand–Ryhming step test	Astrand–Ryhming nomogram—line of best fit	HR, body mass	40	4.03 L min^{-1}	4.03 L min^{-1}	$SEE = 0.28$ L min^{-1}	±	NR

Table 1 continued

Study	No. of participants (sex)	Age, years (mean ± SD, or range)	Protocol	Equation	Variables	Step height (cm)	Mean measured VO_2max	Mean predicted VO_2max	Validity	Direction of prediction difference	Reliability
Santo and Golding [19]	60 (33M, 27F)	18–55	Modified YMCA 3-min step test	VO_2max = (−0.9675 × post-test HR at 15s) + 77.643 OR VO_2max = (−0.2805 × post-test HR at 60s) + 76.710	HR	Individualised	48.4 ± 8.9 mL kg^{-1} min^{-1}	NA	Post-test HR at 15 s: $r = 0.73$ Post-test HR at 60 s: $r = 0.75$	=	NR

M male, *F* female, *SD* standard deviation, *VO₂max* maximal oxygen uptake, *G* gender factor, *PFA* perceived functional ability, *HR* heart rate, *RPE* rating of perceived exertion, *SEE* standard error of the estimate, *NR* not reported, *NA* not applicable

subjects [12, 13, 19, 20, 39]. The publication year ranged from 1954 to 2014.

The aim of this section is to summarise the studies that have examined the validity of submaximal step-tests to estimate VO_2max. In Sect. 3.1, each study is reviewed and then the overall validity of submaximal step-test protocols is summarised in Sect. 3.2. In Sect. 3.3, the results of the risk of bias assessment are summarised.

3.1 Step-Test Protocols

The 11 included studies examined the validity of eight different submaximal step tests to estimate VO_2max in healthy adults: the CST [12, 13], a personalised step test [20], the STEP tool step test [39], the Queen's College step test [33, 34, 36, 37], the Skubic and Hodgkins step test [34], a height-adjusted, rate-specific, single-state step test [35], the Astrand–Ryhming step test [38], and a modified YMCA 3-min step test [19].

3.1.1 Chester Step Test

Sykes and Roberts [12] investigated the reliability and validity of the CST in 68 individuals (mean age 30.6 ± 9.7 years) of both sexes and varying fitness levels, for the prediction of VO_2max, when compared with measured VO_2max from a maximal treadmill test.

The CST required participants to step up and down a single step, 30 cm in height, to a metronome beat (on a prerecorded audio tape). Stepping commenced at 15 steps/min for 2 min, after which both HR and RPE were recorded (level 1) during exercise. Step rate then increased by 5 steps/min to 20 steps/min for a further 2 min, where HR and RPE were again recorded (level 2). The test followed this incremental pattern until participants either reached an HR of 80 % of the predicted maximum (220 − age), or completed the test. The maximum test duration was 10 min (level 5). VO_2max was then predicted by plotting the exercise HRs on a graphical datasheet, where a visual line of best-fit is drawn between data points, projecting the line up to maximum HR and then estimating the matching oxygen uptake value. The step test was run twice on 2 individual days (CST1 and CST2).

There was a high overall correlation ($r = 0.92$) between directly measured VO_2max from a GXT and predicted VO_2max from the CST, with slightly higher values seen for females ($r = 0.95$) than males ($r = 0.87$). The standard error of the predicted VO_2max for CST1 was ±3.9 mL kg^{-1} min^{-1}. The test–retest reliability of the CST was reported to be good. The mean difference between predicted measures between CST1 and CST2 was −0.7 mL kg^{-1} min^{-1}, leading to the conclusion that the CST is

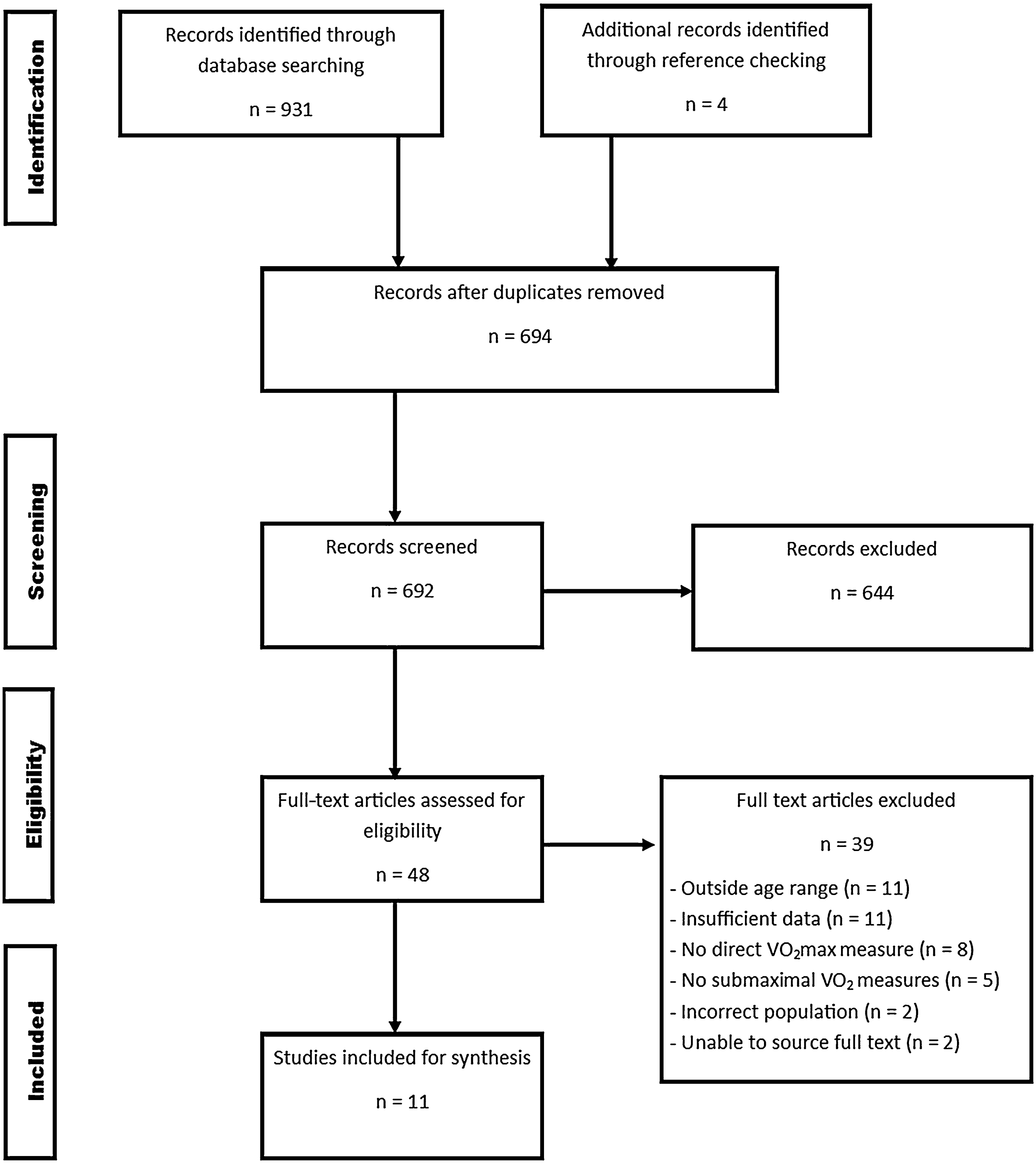

Fig. 1 Study identification process. *VO_2max* maximal oxygen uptake, *VO_2* oxygen uptake

a valid predictor of aerobic capacity in males and females of varying ages and fitness levels.

Buckley et al. [13] further investigated the reliability and validity of measures taken during the CST to predict VO_2max in 13 participants (7 males, 6 females; mean age 22.4 ± 4.6 years) against measured VO_2max taken during a GXT performed on a treadmill. In their study, participants performed the CST twice on two separate occasions (CST1, CST2), 5–7 days apart, using the protocol outlined by Sykes and Roberts [12], with the endpoint of the test raised to 90 % of the age-predicted HR maximum and/or RPE 17. This was done as a means to determine oxygen uptake measures from as many stages of the CST as possible. HR and RPE were recorded during the last 15 s of each stage. According to the procedures described for the CST, the VO_2max was predicted from the CST manual data sheet using the points at or below the 80 % of the age-estimated HR maximum, as previously stated [12].

The CST1 underestimated actual VO_2max by 2.8 mL kg^{-1} min^{-1} ($p = 0.006$) and CST2 by 1.6 mL kg^{-1} min^{-1} (nonsignificant). The agreement between predicted VO_2max was relatively narrow, with a bias ±95 % limits of agreement of -0.8 ± 3.7 mL kg^{-1} min^{-1}. Estimated versus measured oxygen uptake at each stage of the CST during both trials produced errors ranging between 11 and 17 %.

The authors concluded that VO_2max prediction using the CST has questionable validity but is reliable on a test–retest basis, and as such may be an appropriate field-based test to detect changes in aerobic capacity but not to provide a valid measure of VO_2max. The VO_2max prediction error is suggested to be due to estimation error of oxygen uptake at each stage of the CST. A practice trial of the CST is recommended to increase accuracy of measurement, associated with a potential learning effect of the test.

It is important to note that in both studies the use of a visual line of best-fit and the prediction of maximal HR using an age-predicted maximal HR equation may have introduced potential error, impacting on the accuracy of the results.

3.1.2 *Personalised Step Test*

Webb et al. [20] developed and validated an individualised step-test protocol to predict VO_2max in 80 college students (38 males, 42 females; aged 18–29 years) against measured VO_2max via a maximal GXT performed on a treadmill.

The step test used an individualised step height determined by the Culpepper and Francis [40] equation (0.19 × height in cm). To further personalise the step test, the participant began the step test at a stepping rate of 10, 15, 20, or 25 steps/min (stage 1, 2, 3 or 4, respectively) depending on a predetermined perceived functional ability questionnaire (PFA) score. The PFA score was determined by the following method:

Each participant completed two PFA questions: "Suppose you were going to exercise continuously on an indoor track for 1 mile. Which exercise pace is just right for you "not too easy and not too hard?", and "how fast could you cover a distance of 3-miles and not become breathless or overly fatigued? Be realistic." For each of the two questions, participants were instructed to select one numbered response that best described their current level of functional ability to walk, jog, or run 1 or 3 miles. The participant could choose one of 13 possible responses for each PFA question. The participant's PFA score was calculated as the sum of the responses to the two PFA questions. Thus, the range of possible PFA scores was 2–26 [20].

For each participant, 75 % of their age-predicted maximal HR (207 − (0.7 × age)) was calculated as the end-point of the step test. Each individual was familiarised with the step sequence required, prior to commencing the step test. A metronome was used to help participants maintain the required step rate during the familiarisation period, and during the step test. Following familiarisation, subjects began the step test at the predetermined step rate. Each stage was 2 min in duration, with HR and RPE recorded during the final 30 s of each stage. At the completion of each stage, step rate was increased by 5 steps/min. This continued until the measured HR was ≥75 % of the age-predicted maximal HR. When HR reached 75 % of the age-predicted maximum, the individual finished the current stage and the test was terminated. Immediately following test completion, HR was recorded immediately post-exercise and every 15 s thereafter for 1 min in a seated position. Multiple linear regression analysis yielded the following equation to predict VO_2max.

$$\begin{aligned} VO_2\text{max} = {} & 45.94 + 9.25 \times (\text{sex}) - 0.14 \times (\text{body mass}) \\ & + 0.67 \times (\text{PFA score}) + 0.43 \\ & \times (\text{final step rate}) - 0.15 \\ & \times (\text{resting HR 45s post-test}). \end{aligned}$$

Results showed a strong relationship with measured and predicted VO_2max ($r = 0.90$), with a standard error of measurement of 3.4 mL kg^{-1} min^{-1}. This study led the authors to conclude that the individualised step-test protocol provides a model to predict VO_2max from non-exercise data and data collected during an individualised multistage step test that is accurate, time-efficient, and easy to administer.

3.1.3 *STEP Tool Protocol*

Knight et al. [39] demonstrated the validity of the STEP tool, a two-step, step-test protocol, to estimate VO_2max in 40 healthy adults (mean age 43 ± 14 years) against measured VO_2max obtained by a maximal GXT performed on a treadmill.

The STEP tool required participants to step up and down two steps, 20 cm in height, at a self-selected pace, for a total of 20 step cycles. Participants completed one to two practice step cycles before the test to get accustomed to a comfortable pace. Time to complete test, post-test HR (bpm) obtained from the 6 s immediately following the completion of the test, body mass (kg), age (years) and sex (females = 1; males = 2) were recorded and entered into the following prediction equation to estimate VO_2max;

$$VO_2\text{max} = 3.9 + (1511/\text{time}) \times ((\text{body mass}/\text{HR}) \times 0.124) - (\text{age} \times 0.032) - (\text{sex} \times 0.633)$$

There was a strong relationship between measured and estimated VO_2max ($r = 0.78$), which remained strong irrespective of sex (female, $r = 0.79$; male, $r = 0.78$). Some systematic bias was observed between tests, with the STEP tool protocol overestimating VO_2max (+6.4 mL kg^{-1} min^{-1}). The authors concluded that the STEP tool is a valid measure for the estimation of VO_2max, although recommended further research to explore age corrections for younger populations. As this test has shown to overestimate VO_2max, there may be some refinement of the prediction equation required to provide a more accurate estimation in healthy adults. In its current state, it may be an acceptable field measure to provide an estimate of aerobic capacity, but not an exact measure.

It should be noted that moderate physical exertion is required to complete the STEP tool protocol, and as such it may not be suitable in certain populations where exercise above a low intensity is not recommended.

3.1.4 *Queen's College Step Test*

McArdle et al. [34] first demonstrated the validity of the Queen's College step test to predict VO_2max in 41 female college students (aged 19–23 years) against a direct measure of VO_2max obtained during a GXT on a treadmill.

The step test required participants to step up and down a step, 41.3 cm high, at a rate of 22 step cycles (up, up, down, down)/min, for a total duration of 3 min. Step rate was set by a metronome. After completion of the step test, the subject remained standing and recovery HR was measured for 15 s, from 5–20 s post-test. This post-exercise HR was converted to beats/min and used to predict VO_2max using the following equation: VO_2max (mL kg^{-1} min^{-1}) = 65.81 − (0.1847 × HR in beats/min).

A significant correlation was observed between predicted and measured VO_2max ($r = 0.75$), with a standard error of prediction ±2.9 mL kg^{-1} min^{-1}, suggesting an appropriate way to predict VO_2max if an exact measure is not required.

Chatterjee et al. [36] evaluated the validity of the Queen's College step test for the estimation of VO_2max in 30 sedentary male college students (mean age 22.6 ± 0.2 years) against a direct measurement of VO_2max taken during a maximal GXT on a cycle ergometer. The protocol used was the same as reported in the study by McArdle et al. [34]. Post-exercise HR was converted to beats/min and used to predict VO_2max using the following equation: VO_2max (mL kg^{-1} min^{-1}) = 111.3 − (0.426 × pulse rate in beats/min).

The difference between the mean measured and predicted VO_2max values (39.8 ± 1.03 and 39.3 ± 1.1 mL kg^{-1} min^{-1}, respectively) was nonsignificant ($p > 0.10$), with a high correlation ($r = 0.95$; $p < 0.01$). The authors concluded the Queen's College step test can be used in the studied population to provide a good prediction of VO_2max.

Chatterjee et al. [33] further evaluated the validity of the Queen's College step test for the estimation of VO_2max in 40 female students (mean age 21.9 ± 3.2 years) against a direct measurement of VO_2max taken during a maximal GXT on a cycle ergometer. The same step-test protocol was used as above. VO_2max was predicted using the following equation: VO_2max (mL kg^{-1} min^{-1}) = 65.8 − (0.1847 × HR in beats/min).

The mean value of predicted VO_2max (35.5 ± 4.4 mL kg^{-1} min^{-1}) was significantly higher than measured VO_2max (32.8 ± 3.8 mL kg^{-1} min^{-1}), overestimating by an average of 2.7 mL kg^{-1} min^{-1}. The authors concluded that the Queen's College step test is an unacceptable method of predicting VO_2max in this population.

Perroni et al. [37] further evaluated the validity of the Queen's College step test to estimate VO_2max in 15 male firefighters (aged 31 ± 6 years) against a direct measure of VO_2max taken during a maximal GXT on a treadmill. The step-test protocol used was the same used in McArdle et al. [34], although step height was reported at 40 cm rather than 41.3 cm. The VO_2max was predicted using the following equation: VO_2max (mL kg^{-1} min^{-1}) = 111.3 − (0.426 × pulse rate in beats/min).

Significant differences were observed between mean predicted and measured VO_2max (52.8 ± 4.7 and 49.2 ± 7.6 mL kg^{-1} min^{-1}, respectively), with an average overestimation of 3.6 mL kg^{-1} min^{-1}. Only a moderate correlation between estimated and measured VO_2max was observed ($r = 0.469$), indicating that the Queen's College step test is not a suitable test to predict VO_2max in this population.

It is worthy of note that the authors did not indicate whether the low correlation between measured and predicted VO_2max was a result of significant random error associated with the procedure or the fixed figure associated with the regression equation, leading to the possibility that using the same procedure with an alternate fixed variable more suitable for this specific population may provide a more accurate estimate of VO_2max.

3.1.5 *Height-Adjusted, Rate-Specific, Single-Stage Step Test*

Francis and Culpepper [35] demonstrated the validity of a rate-specific, single-stage step test, in which step height was individualised, to predict VO_2max in 17 female college students (19–33 years) against a direct measure of VO_2max obtained during a maximal GXT performed on a treadmill.

The step test protocol was the same as reported in the study by McArdle et al. [34] in which participants stepped up and down a step at a rate of 22 step cycles (up, up, down, down)/min, for a total duration of 3 min. Step rate was set by a metronome. Step height was individualised to each participant, based on the height of the foot when the hip was flexed at an angle of 73.3 degrees. Each participant also completed the step test at a step rate of 26 and 30 step cycles/min as a means to investigate the impact of step rate on prediction validity. Each individual test was completed 24–48 h apart.

After completion of the step test, the subject remained standing and recovery heart beats were counted for 15 s, from 5–20 s post-test. This post-exercise heart beat count was converted to beats/min and was used to predict VO_2max.

Statistically significant correlations were reported between predicted and measured VO_2max ($r = 0.74$, 0.80 and 0.77 for 22, 26 and 30 step cycles/min, respectively), with the prediction of VO_2max from 26 step cycles/min demonstrating the strongest relationship. Standard error of measurement was reported at ±3.09, 2.87 and 2.59 mL kg^{-1} min^{-1} for prediction of VO_2max from 22, 26 and 30 step cycles/min, respectively.

The authors concluded that their height-adjusted, rate-specific, single-stage step test described provides an effective method of predicting VO_2max in young, healthy adult females when more complex methods of laboratory testing are unavailable or unfeasible.

3.1.6 Astrand–Ryhming Step Test

Astrand and Ryhming [38] evaluated the validity of a 5-min step test to predict VO_2max in 18 well-trained male adults (18–19 years), in comparison to measured VO_2max obtained during a maximal GXT performed on a treadmill.

The Astrand–Ryhming step test requires subjects to step up and down a bench for 5 min, at a rate of 22.5 individual steps/min. Step height is set at 40 cm for males and 30 cm for females. Exercise HR is measured for the final 15 s each minute of exercise. A steady HR value is used to predict VO_2max, in conjunction with the participant's body mass and the Astrand–Ryhming nomogram [38, 41]. If a steady HR value is not achieved, the last value recorded is used to estimate VO_2max from the nomogram.

The Astrand–Ryhming step test reported a standard error of measurement of 0.28 L min^{-1} (6.8 %), leading the authors to conclude that this step test is an appropriate method of predicting VO_2max.

3.1.7 Skubic and Hodgkins Step Test

McArdle et al. [34] investigated the validity of the Skubic and Hodgkins step test to predict VO_2max in 41 female college students (aged 19–23 years) against a direct measure of VO_2max obtained during a GXT on a treadmill.

The Skubic and Hogkins step test consisted of stepping up and down a bench, 45.7 cm in height, at a rate of 24 steps/min, for 3 continuous min. At 1 min post-test, 30-s pulse count was taken (60–90 s post-test) and converted to beats/min. The VO_2max was predicted using the following equation: $VO_2\text{max} = 55.9 - (0.1517 \times \text{pulse rate in beats/min})$.

A low but statistically significant correlation ($r = 0.64$) was seen between predicted and measured VO_2max using this method, with a standard error of measurement of ±3.5 mL kg^{-1} min^{-1}, leading to the conclusion that this method does not provide an accurate prediction of VO_2max

3.1.8 Modified YMCA 3-Min Step Test

Santo and Golding [19] demonstrated the validity of a modified 3-min step test to estimate VO_2max in 60 (27 females, 30 males) healthy participants (aged between 18 and 55 years), against a maximal treadmill test.

The YMCA 3-min step test was altered by adjusting the step height to the individual participant's stature, based on the following equations:

$$\begin{aligned}&\text{Step height for women (cm)}\\&= (0.189) \times (\text{participant height in cm});\\&\quad\text{step height for men (cm)} = (0.192)\\&\quad\times (\text{participant height in cm}).\end{aligned}$$

All other parameters used were the same as the original YMCA step test [20], in which participants step up and down a single step at a rate of 24 steps/min for 3 min. Following the 3-min test, recovery heart beat count is measured for 15 s, commencing 5 s after the completion of the test (5–20 s), and 1 min post-test (60–75 s), and converted to beats/min.

The correlation coefficient was calculated between measured VO_2max and both the 15 s and 1-min post-test HRs. Linear regression was used to develop the following prediction equations for VO_2max from these data:

$VO_2\text{max} = (-0.9675 \times \text{post-test HR 15 s}) + 77.643$; $VO_2\text{max} = (-0.2805 \times \text{post-test HR at 60 s}) + 76.710$. The correlation coefficient between measured and predicted VO_2max from recovery HR at 15 s and 1 min from the modified YMCA was 0.73 and 0.75, respectively, with no definitive difference in measured direction. The authors therefore concluded that the modified YMCA step-test protocol may be a valid means of predicting VO_2max.

It is worth noting the limitations in the study. Their procedure assumes a similar fitness between participants by providing a fixed step rate, leading to the assumption that workload is standardised to a percentage of HR maximum, whereas, in reality, individuals with a lower fitness level will show a larger HR response at any given workload. Validity and reliability of the modified step test may have been affected due to the small participant group and the lack of a cross-validation group. Standard error of the estimate (SEE) is often used to determine the degree of error related to prediction equations established from regression statistics. These SEE values were higher than desired (15–20 % of the measured mean), potentially indicating a limitation to using recovery HR to estimate VO_2max.

3.2 Overall Validity and Reliability of Submaximal Step Tests

The results suggest a relatively strong ability for submaximal step tests to predict VO_2max, although poor validity was shown in some populations. Three studies [13, 36, 37] showed poor validity in the prediction of VO_2max, which is suggested to be as a result of the protocol and equation used for prediction. The remaining studies [13, 19, 20, 34–36, 39] demonstrated a significant relationship between predicted VO_2max from submaximal step test and measured VO_2max ($r = 0.64$–0.95), or a small standard error of measurement [38] (0.28 L min^{-1}, or 6.8 % of mean measured VO_2max), suggesting a valid tool for the submaximal estimation of VO_2max.

Only two studies [12, 13] undertook multiple step tests to provide a reliability measure, both of which were for the CST protocol. Both showed high reliability on a test–retest basis (-0.7 and -0.8 mL kg^{-1} min^{-1}, 1.3 and 1.7 % of mean measured VO_2max, respectively), and suggested a potential learning effect between the first and second test, with prediction of VO_2max from the second test demonstrating a stronger relationship with measured VO_2max.

3.3 Risk of Bias

Figure 2 depicts relatively poor performance for a majority of the studies in the participant selection (selection bias) section. Whilst they adequately report sample selection procedures, the population does not often represent an accurate depiction of a healthy adult population. A relatively poor performance from the majority of studies in data collection (data collection method), as studies often failed to report reliability and validity data for their data collection tools, resulted in a weak rating for the data collection method (Fig. 2; data collection method). Overall, five of the nine studies were rated globally as weak, two were rated moderate, and two were rated strong. Table 2 provides a summary of how each individual paper was rated in each area.

4 Discussion

4.1 Validity and Reliability

This review provides moderate evidence for the efficacy of submaximal step tests to estimate VO_2max in healthy adults, with seven of the included studies demonstrating a strong relationship between predicted and measured VO_2max. Results suggest that the individualised protocol used by Webb et al. [20] provides the most accurate prediction of VO_2max in healthy adults, irrespective of sex. Sykes and Roberts [12] reported a positive relationship ($r = 0.92$) between predicted VO_2max using the CST and measured VO_2max. Observing no significant difference between

Fig. 2 Risk of bias assessment

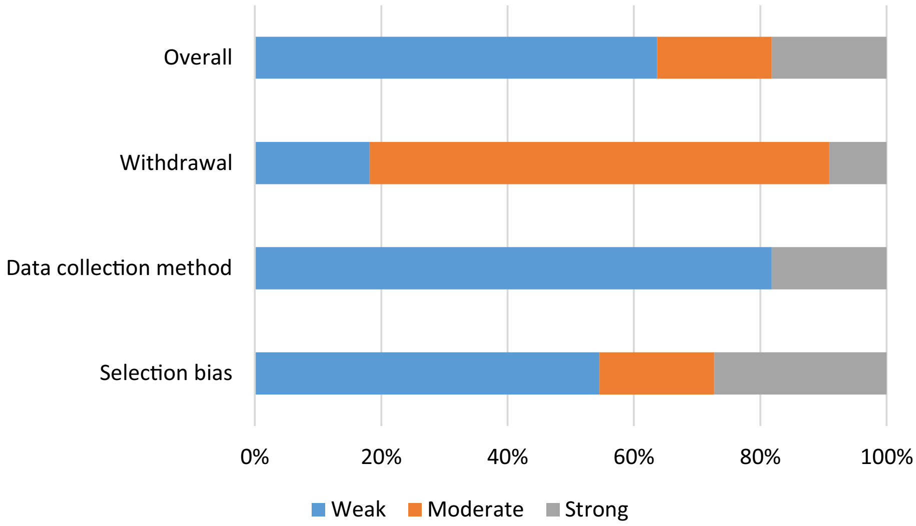

Table 2 Summary of risk of bias assessment by individual paper

Study	Selection bias	Study design	Confounders	Blinding	Data collection method	Withdrawals and dropouts	Overall
Buckley et al. [13]	Moderate	NA	NA	NA	Strong	Strong	Strong
Sykes and Roberts [12]	Moderate	NA	NA	NA	Strong	Moderate	Strong
Webb et al. [20]	Strong	NA	NA	NA	Weak	Moderate	Moderate
Knight et al. [39]	Strong	NA	NA	NA	Weak	Moderate	Moderate
Chatterjee et al. [36]	Weak	NA	NA	NA	Weak	Moderate	Weak
Chatterjee et al. [33]	Weak	NA	NA	NA	Weak	Moderate	Weak
McArdle et al. [34]	Weak	NA	NA	NA	Weak	Moderate	Weak
Perroni et al. [37]	Weak	NA	NA	NA	Weak	Moderate	Weak
Francis and Culpepper [35]	Weak	NA	NA	NA	Weak	Moderate	Weak
Astrand and Ryhming [38]	Weak	NA	NA	NA	Weak	Weak	Weak
Santo and Golding [19]	Strong	NA	NA	NA	Weak	Weak	Weak

NA not applicable

measured and predicted VO_2max, the authors concluded that the test was a valid and accurate method for estimating VO_2max. Buckley et al. [13] showed conflicting results, reporting a poor relationship using the same step-test protocol, leading the authors to conclude that the CST is not a valid predictor of VO_2max.

Both the modified YMCA 3-min step test [19] and STEP tool protocol [39] showed strong relationships between predicted and measured VO_2max ($r = 0.75$, 0.73 and $r = 0.78$, respectively) irrespective of sex, supporting their use is appropriate within the healthy adult population when an estimation of VO_2max is needed, rather than an exact measure.

Predicted VO_2max from the height-adjusted, rate-specific, single-stage step test [35] demonstrated a strong relationship with measured VO_2max in young, healthy female adults when using a fixed step rate of 22 ($r = 0.74$), 26 ($r = 0.80$), and 30 ($r = 0.77$) steps/min. The Astrand–Ryhming step test [38] also demonstrated good predictive validity of VO_2max in young, healthy, male adults (18–19 years) in comparison to a direct measure of VO_2max, with a standard error of estimate of 0.28 L min^{-1} (6.8 %) reported.

The Queen's College step test provided conflicting results [33, 36], showing poor validity when estimating VO_2max in sedentary, but healthy, female adults, overestimating by an average of 2.7 mL kg^{-1} min^{-1} (8.2 % of mean measured VO_2max), but a valid predictor of VO_2max (r – 0.95) in young, sedentary adult males. McArdle et al. [34] demonstrated its validity ($r = 0.75$) to predict VO_2max in healthy, female adults. Perroni et al. [37] demonstrated a poor correlation between predicted and measured VO_2max ($r = 0.469$) in healthy male firefighters, concluding it an inappropriate tool to predict VO_2max in this population. The large variations in validity measures observed between Chatterjee et al. [33] and Perroni et al. [37] could be explained by a number of factors. First, there were differences in the average age, height and fitness of the participants, which could have impacted the results of the study. Second, Chatterjee et al. [33] used cycle ergometry to assess VO_2max, whereas Perroni et al. [37] used a treadmill-based GXT. Cycling could be considered to have greater similarity to step exercise than treadmill running due to its lower-limb dominance, and as such VO_2max prediction may be more accurate when compared with measured VO_2max obtained via cycling exercise.

Also, the Skubic and Hogkins step test [34] demonstrated a low, but statistically significant, correlation ($r = 0.64$) between predicted and measured VO_2max in young, healthy female adults.

As only two of the included studies [12, 13] assessed reliability, a definite conclusion is difficult. Both studies demonstrated good reliability on a test–retest basis, with a slight decrease in estimated VO_2max between step-test 1 and step-test 2, providing a more accurate prediction of VO_2max on step-test 2. The similar results seen were likely due to both studies validating the CST. The slight increase in accuracy between step tests indicates a possible learning effect, leading to the suggestion that one to two practice tests may be used to ensure an accurate estimation of VO_2max on the final step test used. Due to the limited number of studies assessing reliability, it is apparent that more research is warranted in this area to demonstrate reliability measures of different step-test protocols.

4.2 Limitations of Included Studies

Appropriate individualisation of protocols appears to provide valid estimates of VO_2max [19, 20, 35]. Altering step height to individual parameters and limiting step rate to a predetermined PFA score appears to ensure the work performed is of an appropriate intensity for the individual, limiting muscular fatigue while ensuring adequate work is completed by the cardiorespiratory system. Validity appeared to increase with self-pacing [39], with a self-selected step rate ensuring a moderate intensity is maintained throughout the test duration, eliminating the potential issues associated with a fixed step-rate.

Given the success seen with the inclusion of RPE as a means of regulating submaximal exercise tests, as evidenced by the perceptually-regulated exercise test (PRET) procedures to accurately and reliably estimate VO_2max in a variety of populations and via a variety of exercise modalities, irrespective of sex and fitness levels [22–28], the use of a perceptually-regulated protocol with a step test may have the capacity to improve accuracy and reliability of prediction.

Nine of the included studies did not account for body mass and variability in body composition [12, 13, 19, 20, 33–37]. Different body compositions may impact on the results attained during submaximal exercise testing and may therefore limit VO_2max prediction. At any given submaximal workload, an individual with a greater body mass will work at a greater percentage of VO_2max [19]. This suggests that if two individuals have the same VO_2max, but different body mass, and perform the same step test, the individual with a greater body mass will have a higher exercising HR, and thus a lower VO_2max estimation.

Both studies attempting to validate the CST [12, 13] indicated the use of a visual line of best-fit on a graphical data sheet as a potential source of error, and suggested that the validity of the result could be altered significantly as a result of this. The CST also makes the assumption that the HR and oxygen uptake responses will be linear, in relation to the successive increases in work rate with each stage of the CST. Buckley et al. [13] demonstrated that the responses were nonlinear. The significantly greater measured oxygen uptake compared with estimated oxygen uptake at stage 1 of the CST in both trials is suggested to be the cause of this curvilinear relationship.

Similarly, reliance on the Astrand–Ryhming nomogram in the Astrand–Ryhming step test [38] could be considered a potential source of error. The nomogram is based on the supposition of a linear relationship between HR and oxygen intake during exercise, in which there are many exceptions to this scenario [13].

Age-predicted maximal HR equations were used to determine the endpoint of both the CST [12, 13] and the personalised step test [20]. Such equations have a considerable prediction error (average standard deviation of ±10 beats/min) [42], and significantly underestimate maximal HR in older adults [43]. The use of these equations is identified as a potential source of error because if the age-predicted maximal HR is incorrect, the corresponding VO_2max will also be incorrect.

As with all step tests, the subject's ability to maintain the correct or steady stepping tempo and technique is paramount to getting an accurate measure. Alterations in step technique can affect mechanical efficiency and therefore physiological responses in HR and oxygen consumption. The individual differences in step technique, and the inability to maintain a constant rate of stepping, could introduce potential error in each of the studies included.

4.3 Limitations of Review Methods

This review includes only papers published in the English language, potentially missing sources of information relevant to the topic that were published in a different language.

Many studies reviewed were excluded based on insufficient data reporting and population characteristics outside of that defined as healthy adults aged between 18 and 65 years. As such, it is likely that there are more step tests that could be applied to healthy adults, but they did not meet the inclusion criteria for this review. Of these, the Canadian Home Fitness Test in particular is worthy of note [14, 17]. This is a two-step exercise test in which the individual is required to step up and down two steps, 20.3 cm in height, at an age- and sex-specific step-rate, for 3 min. The rate is controlled by prerecorded music. Ten-second post-exercise heart beat is recorded between 5 and 15 s post-test, and is then used to predict VO_2max. Using this test, Jetté et al. [17] demonstrated a strong correlation ($r = 0.74$) between predicted and measured VO_2max, with a standard error of measurement of ± 4.08 mL kg^{-1} min^{-1}, in 59 healthy individuals (15–74 years).

Four of the studies included in this systematic review [13, 35, 37, 38] had relatively small sample sizes of 13, 15, 17 and 18 participants, respectively, potentially limiting the statistical power of results. Six of the included studies [33–38] had very specific populations, which may not provide an accurate representation of how the step-test protocol used may estimate VO_2max in a more generalised adult population.

When globally rated, 7 of the 11 included studies were rated as weak, and another two were rated as moderate, suggesting potential bias. This finding suggests either poor methodology and/or poor reporting in certain studies. Of particular note, only two studies reported both validity and reliability data in regard to their testing procedures [12, 13]. If a submaximal protocol has poor test–retest reliability, then the equation which it is based on will also have poor test–retest reliability. Failure to report reliability measures will reduce confidence in the final predictive equation.

5 Conclusions

Considering the relationship between VO_2max and various markers of health, its use as an assessment of health in both the general adult population and rehabilitation settings is recommended in order to monitor and evaluate health safely and effectively. Testing and maintaining cardiorespiratory fitness could help prevent declines in health within the general population. The step test provides a simple, effective and ecologically valid method of submaximally assessing cardiorespiratory fitness that can be implemented in a variety of settings.

Little information exists regarding the validity of submaximal step tests to predict VO_2max in healthy adults. This systematic review provides clinicians, practitioners and researchers with information regarding the accuracy and other factors that will help aid in the selection of which submaximal step test to use.

Future research is needed to assess the reliability of the majority of the step-test procedures reviewed. Based on the validity measures, submaximal step-test protocols are an acceptable method of estimating VO_2max in the general healthy adult population. If tracking changes in cardiorespiratory fitness, the CST appears to be an appropriate tool due to its high test–retest reliability.

Compliance with Ethical Standards

Funding No sources of funding were used to assist the preparation of this article.

Conflicts of interest Hunter Bennett, Roger Eston, Gaynor Parfitt and Kade Davison declare they have no conflicts of interest relevant to content of this review.

References

1. Lee DC, Artero EG, Sui X, et al. Review: mortality trends in the general population: the importance of cardiorespiratory fitness. J Psychopharmacol. 2010;24(4):27–35.
2. Pedersen BK. Body mass index-independent effect of fitness and physical activity for all-cause mortality. Scand J Med Sci Spor. 2007;17(3):196–204.
3. Katch V, Katch F, Mcardle W. Essentials of exercise physiology. Philadelphia: Lippincott Williams & Wilkins; 2011.
4. American College of Sports Medicine (ACSM). ACSM's guidelines for exercise testing and prescription. Philadelphia: Lippincott Williams & Wilkins; 2008.
5. Sartor F, Vernillo G, Morree H, et al. Estimation of maximal oxygen uptake via submaximal exercise testing in sports, clinical and home settings. Sports Med. 2013;43(3):865–73.
6. Bassett DR, Howley ET. Limiting factors for maximal oxygen uptake and determinants of endurance performance. Med Sci Sport Exer. 2000;32(1):70–84.
7. Hartung GH, Blancq RJ, Lally DA, et al. Estimation of aerobic capacity from submaximal cycle ergometry in women. Med Sci Sport Exer. 1995;27(3):452–7.
8. Grant S, Corbett K, Amjad AM, et al. A comparison of methods of predicting maximum oxygen uptake. Br J Sports Med. 1995;29(3):147–52.
9. Noonan V, Dean E. Submaximal exercise testing: clinical application and interpretation. Phys Ther. 2000;80(8):782–807.
10. Astrand PO. Quantification of exercise capability and evaluation of physical capacity in man. Prog Cardiovasc Dis. 1976;19(1):51–67.
11. Balderrama C, Ibarra G, De La Riva J, et al. Evaluation of three methodologies to estimate the VO_2max in people of different ages. Appl Ergon. 2010;42(1):162–8.
12. Sykes K, Roberts A. The Chester step test—a simple yet effective tool for aerobic capacity. Physiotherapy. 2004;90(1):183–8.
13. Buckley J, Sim J, Eston R, et al. Reliability and validity of measures taken during the Chester step test to predict aerobic power and to prescribe aerobic exercise. Br J Sports Med. 2004;38(2):197–205.
14. Shephard R, Bailey D, Mirwald R. Development of the Canadian home fitness test. Can Med Assoc J. 1976;114(8):675.
15. Shephard R. The current status of the Canadian home fitness test. Br J Sports Med. 1980;14(2):114–25.
16. Shephard RJ, Thomas S, Weiler I. The Canadian home fitness test. Sports Med. 1991;11(6):358–66.
17. Jetté M, Campbell J, Mongeon J, et al. The Canadian home fitness test as a predictor of aerobic capacity. Can Med Assoc J. 1976;114(8):680–2.
18. Hansen D, Jacobs N, Bex S, et al. Are fixed-rate step tests medically safe for assessing physical fitness? Eur J Appl Physiol. 2011;111(2):2593–9.
19. Santo A, Golding LA. Predicting maximum oxygen uptake from a modified 3-minute step test. Res Q Exer Sport. 2003;74(1):110–5.
20. Webb C, Vehrs PR, George JD, et al. Estimating VO_2max using a personalised step test. Meas Phys Educ Exer Sci. 2014;18(1):184–97.
21. Jones J, Bell J, Buckley J, et al. Physical activity and exercise in the management of cardiovascular disease. Estimation of MET max from submaximal exercise. London: The British Cardiovascular Society; 2014. p. 211.
22. Eston R, Lamb K, Parfitt G, et al. The validity of predicting maximal oxygen uptake from a perceptually-regulated graded exercise test. Eur J Appl Physiol. 2005;94(3):221–7.
23. Eston R, Faulkner J, Mason E, et al. The validity of predicting maximal oxygen uptake from perceptually-regulated graded exercise tests of different durations. Eur J Appl Physiol. 2006;97(5):535–54.
24. Eston R, Lambrick D, Sheppard K, et al. Prediction of maximal oxygen uptake in sedentary males from a perceptually regulated, sub-maximal graded exercise test. J Sports Sci. 2008;26(2):131–9.
25. Eston R, Evans H, Faulkner J, et al. A perceptually regulated, graded exercise test predicts peak oxygen uptake during treadmill exercise in active and sedentary participants. Eur J Appl Physiol. 2012;112(10):3459–68.
26. Faulkner J, Parfitt G, Eston R. Prediction of maximal oxygen uptake from the ratings of perceived exertion and heart rate during a perceptually-regulated sub-maximal exercise test in active and sedentary participants. Eur J Appl Physiol. 2007;101(3):397–407.
27. Morris M, Lamb KL, Hayton J, et al. The validity and reliability of predicting maximal oxygen uptake from a treadmill-based submaximal perceptually regulated exercise test. Eur J Appl Physiol. 2010;109(5):983–8.
28. Al-Rahamneh H, Eston R. The validity of predicting peak oxygen uptake from a perceptually guided graded exercise test during

arm exercise in paraplegic individuals. Spinal Cord. 2010;49(3):430–4.

29. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 2009;151(4):264–9.
30. Effective Public Health Practice Project. Quality Assessment Tool for Quantitative Studies. 2008. http://www.ephpp.ca/PDF/Quality%20Assessment%20Tool_2010_2.pdf. Accessed 15 Mar 2015.
31. Oja P, Titze S, Bauman A, et al. Health benefits of cycling: a systematic review. Scand J Med Sci Spor. 2011;21(4):496–509.
32. Ferrar K, Evans H, Smith A, et al. A systematic review and meta-analysis of submaximal exercise-based equations to predict maximal oxygen uptake in young people. Pediatr Exer Sci. 2014;26(3):342–57.
33. Chatterjee S, Chatterjee P, Bandyopadhyay A. Validity of Queen's College step test for estimation of maximum oxygen uptake in female students. Indian J Med Res. 2005;121(1):32–5.
34. McArdle WD, Katch FI, Pechar GS, et al. Reliability and inter-relationships between maximal oxygen intake, physical work capacity and step-test scores in college women. Med Sci Sports. 1971;4(4):182–6.
35. Francis K, Culpepper M. Height-adjusted, rate-specific, single-stage step test for predicting maximal oxygen consumption. South Med J. 1989;82(5):602–6.
36. Chatterjee S, Chatterjee P, Mukherjee PS, et al. Validity of Queen's College step test for use with young Indian men. Br J Sports Med. 2004;38(3):289–91.
37. Perroni F, Cortis C, Minganti C, et al. Maximal oxygen uptake of Italian firefighters: laboratory vs. field evaluations. Sport Sci. Health. 2013;9(2):31–5.
38. Astrand P, Ryhming I. A nomogram for calculation of aerobic capacity (physical fitness) from pulse rate during submaximal work. J Appl Physiol. 1954;7(2):218–21.
39. Knight E, Stuckey M, Petrella R. Validation of the step test and exercise prescription tool for adults. Can J Diabetes. 2014;38(1):164–71.
40. Culpepper M, Francis K. An anatomical model to determine step height in step testing for estimating aerobic capacity. J Theor Biol. 1987;129(1):1–8.
41. Ryhming I. A modified Harvard step test for the evaluation of physical fitness. Eur J Appl Physiol Occup Phys. 1953;15(3):235–50.
42. Robergs R, Landwehr R. The surprising history of the '$HR_{max} = 220$-age" equation. J Exerc Physiol. 2002;5(2):1–10.
43. Tanaka H, Monahan KD, Seals DR. Age-predicted maximal heart rate revisited. J Am Coll Cardiol. 2001;37(1):153–6.