

RESEARCH ARTICLE

# How Accurate Is the Prediction of Maximal Oxygen Uptake with Treadmill Testing?

John R. Wicks<sup>1\*</sup>, Neil B. Oldridge<sup>2</sup>

**1** Department of Rehabilitation, Robina Hospital, Robina, Queensland, Australia, **2** College of Health Sciences, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin, United States of America

\* [john\\_wicks@health.qld.gov.au](mailto:john_wicks@health.qld.gov.au)

## Abstract

### Background

Cardiorespiratory fitness measured by treadmill testing has prognostic significance in determining mortality with cardiovascular and other chronic disease states. The accuracy of a recently developed method for estimating maximal oxygen uptake ( $VO_{2peak}$ ), the heart rate index (HRI), is dependent only on heart rate (HR) and was tested against oxygen uptake ( $VO_2$ ), either measured or predicted from conventional treadmill parameters (speed, incline, protocol time).

### Methods

The HRI equation,  $METs = 6 \times HRI - 5$ , where  $HRI = \text{maximal HR} / \text{resting HR}$ , provides a surrogate measure of  $VO_{2peak}$ . Forty large scale treadmill studies were identified through a systematic search using MEDLINE, Google Scholar and Web of Science in which  $VO_{2peak}$  was either measured (TM- $VO_{2meas}$ ;  $n = 20$ ) or predicted (TM- $VO_{2pred}$ ;  $n = 20$ ) based on treadmill parameters. All studies were required to have reported group mean data of both resting and maximal HRs for determination of HR index-derived oxygen uptake (HRI- $VO_2$ ).

### Results

The 20 studies with measured  $VO_2$  (TM- $VO_{2meas}$ ), involved 11,477 participants (median 337) with a total of 105,044 participants (median 3,736) in the 20 studies with predicted  $VO_2$  (TM- $VO_{2pred}$ ). A difference of only 0.4% was seen between mean ( $\pm$ SD)  $VO_{2peak}$  for TM- $VO_{2meas}$  and HRI- $VO_2$  ( $6.51 \pm 2.25$  METs and  $6.54 \pm 2.28$ , respectively;  $p = 0.84$ ). In contrast, there was a highly significant 21.1% difference between mean ( $\pm$ SD) TM- $VO_{2pred}$  and HRI- $VO_2$  ( $8.12 \pm 1.85$  METs and  $6.71 \pm 1.92$ , respectively;  $p < 0.001$ ).

### Conclusion

Although mean TM- $VO_{2meas}$  and HRI- $VO_2$  were almost identical, mean TM- $VO_{2pred}$  was more than 20% greater than mean HRI- $VO_2$ .



## OPEN ACCESS

**Citation:** Wicks JR, Oldridge NB (2016) How Accurate Is the Prediction of Maximal Oxygen Uptake with Treadmill Testing? PLoS ONE 11(11): e0166608. doi:10.1371/journal.pone.0166608

**Editor:** Gordon Fisher, University of Alabama at Birmingham, UNITED STATES

**Received:** March 2, 2016

**Accepted:** November 1, 2016

**Published:** November 22, 2016

**Copyright:** © 2016 Wicks, Oldridge. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** The authors received no specific funding for this work.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

When assessed as oxygen consumption ( $\text{VO}_2$ ), cardiorespiratory fitness (CRF) may be measured either using a treadmill with conventional gas analysis equipment (TM- $\text{VO}_{2\text{meas}}$ ) or predicted from equations based on treadmill speed, incline or treadmill time (TM- $\text{VO}_{2\text{pred}}$ ) [1]. The prognostic importance of CRF has been extensively investigated in recent meta-analyses confirming the strong inverse relationships between CRF and all-cause mortality in healthy individuals [2] and in patients with either coronary artery disease (CAD) or congestive heart failure (CHF) [3–6]. The prospective studies included in these reviews involve large numbers of subjects and have shown that a 1 MET (equal to  $3.5 \text{ mL O}_2 \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ ) increment increase in CRF is associated with an approximate 10–20% reduction in all cause and cardiovascular mortality [2,7] with a similar effect being observed with CHF [6,8].

Logistics of large studies necessitate prediction of peak  $\text{VO}_2$  ( $\text{VO}_{2\text{peak}}$ ) as measurement of  $\text{VO}_2$  is costly and time consuming. Equations have been determined for the various treadmill protocols based on the variables of treadmill speed, incline or the test time for a particular protocol, a common reference being ACSM publications [1]. However, many factors may contribute to the error of TM- $\text{VO}_{2\text{pred}}$ . They include 1) treadmill handrail support [9–13], 2) failure to use population specific equations [14–18], 3) inappropriate testing protocol [19–21], 4) delayed oxygen kinetics [22–24], 5) reproducibility of cardiopulmonary parameters [25,26], 6) altered mechanical efficiency with treadmill walking [27] and 7) lack of treadmill calibration [28].

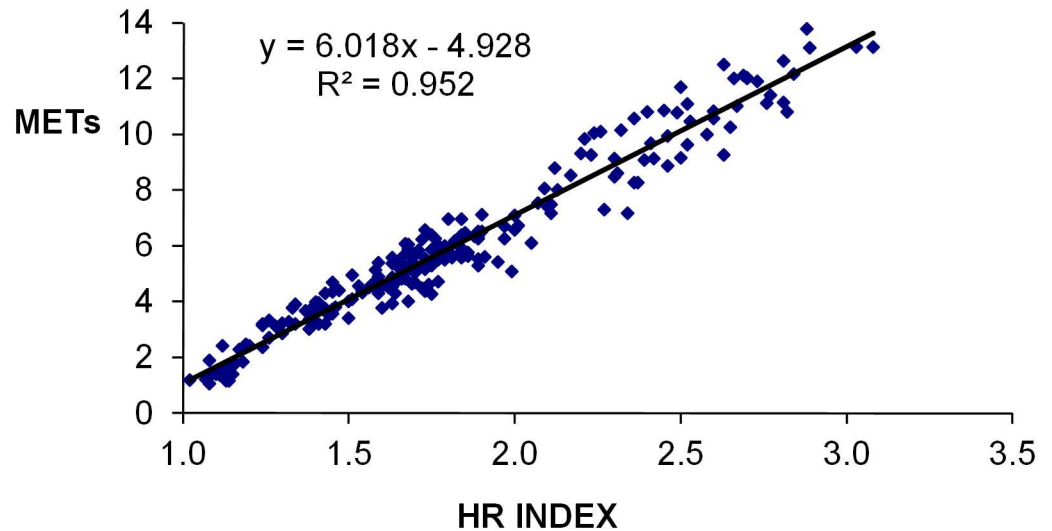
Cardiovascular pathology frequently screened for with treadmill testing includes both CAD and CHF. In using CRF as an outcome measure from a treadmill test,  $\text{VO}_{2\text{peak}}$  is commonly expressed as METs with 1 MET being the  $\text{VO}_2$  at rest with current convention stating that it is equal to  $3.5 \text{ mL O}_2 \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$  [29]. Kaplan-Meier curves have been used extensively to document the link between CRF and long-term morbidity/mortality [30,31]. Although  $\text{VO}_2$  can be predicted from treadmill speed, incline or the test time for a particular protocol, currently the only way to ensure an accurate measurement of  $\text{VO}_2$  is direct measurement with gas analysis. Using only two simple measurements, rest HR and an activity HR (either sub-maximal or maximal), the recently published HR index (HRI = activity HR/rest HR), equation for predicting  $\text{VO}_2$  expressed as METs is associated with a high correlation between HRI and  $\text{VO}_2$ , the equation being METs =  $6 \times \text{HRI} - 5$  [32]. The HRI equation was derived from group mean data from 60 studies in which an exercise test contained a resting HR ( $\text{HR}_{\text{rest}}$ ), and a  $\text{VO}_2$  measured at the activity HR (either submaximal or peak) and expressed in the form of  $\text{mLO}_2 \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$  or METs. The original data are shown as a regression plot in Fig 1. The utility of this equation is that it provides a simple independent surrogate method of estimating  $\text{VO}_2$  using only the rest and either the sub-maximal or maximal activity HR measurements. Though the HRI equation was developed from aggregate data, there has been no analysis to date that has established its predictive accuracy for assessment of  $\text{VO}_2$ .

The objective of this study was to compare aggregate HRI-derived  $\text{VO}_2$  (HRI- $\text{VO}_2$ ) data against  $\text{VO}_{2\text{peak}}$  from two different treadmill tests, either: 1)  $\text{VO}_2$  measured with conventional gas analysis equipment (TM- $\text{VO}_{2\text{meas}}$ ) or 2)  $\text{VO}_2$  predicted from equations based on treadmill speed, incline or treadmill time (TM- $\text{VO}_{2\text{pred}}$ ).

## Methods

### Study selection

Treadmill studies involving assessment of  $\text{VO}_{2\text{peak}}$ , reporting either TM- $\text{VO}_{2\text{meas}}$  or TM- $\text{VO}_{2\text{pred}}$ , were identified through a systematic search conducted on at least a monthly basis



**Fig 1. Linear regression plot of HR index equation.** An analysis from data ( $n = 220$ ) derived from 60 studies with the HR index equation simplified to METs =  $6 \times \text{HR index} - 5$ .

doi:10.1371/journal.pone.0166608.g001

from October 2011 till March 2013 using MEDLINE, Google Scholar and Web of Science. Search terms included (in various combinations) exercise testing, oxygen uptake,  $\text{VO}_2$ , CRF, cardiovascular disease (CVD), CAD, CHF and physical activity. With publications having the prerequisite HR data extensive cross-referencing was undertaken to source other publications with eligible criteria [33].

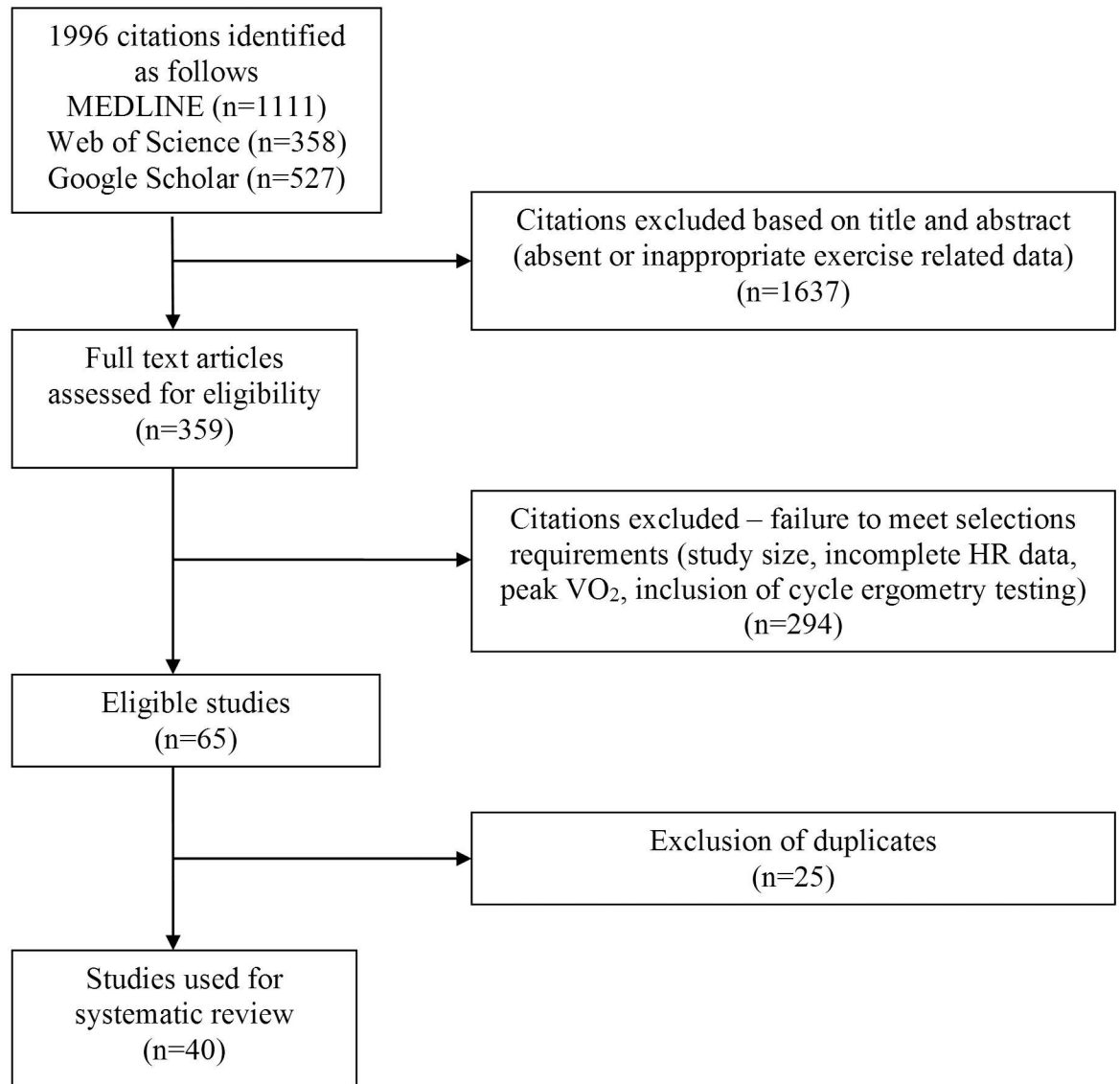
Eligibility criteria for study inclusion are 1)  $>100$  patients enrolled, 2) documented  $\text{VO}_{2\text{peak}}$  (either measured or predicted) expressed as either  $\text{mL O}_2 \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$  or as METs, 3) measured maximal HR ( $\text{HR}_{\text{max}}$ ) associated with  $\text{VO}_{2\text{peak}}$ , and 4) measured  $\text{HR}_{\text{rest}}$ . Where large scale studies included cycle ergometry in conjunction with treadmill testing, the study was excluded. In publications likely to have used a similar subject cohort based on 1. participating authors, 2. study location, 3. time period when the study was performed and 4. characteristics of the study population e.g. healthy, suspected or known CAD the most recent publication was chosen. From the HR data, a predicted MET value ( $\text{VO}_{2\text{peak}}$ ) was derived using the HRI equation ( $\text{METs} = 6 \times \text{HR index} - 5$ , where HR index is  $\text{HR}_{\text{max}}/\text{HR}_{\text{rest}}$ ).

At the time of closure of data acquisition in March 2013 a total of 40 studies (TM- $\text{VO}_{2\text{meas}}$ ;  $n = 20$  studies, TM- $\text{VO}_{2\text{pred}}$ ;  $n = 20$  studies) had been identified with all but one being published since 1991. MEDLINE searching identified 19 of the 40 studies (TM- $\text{VO}_{2\text{meas}}$ ;  $n = 11$  studies, TM- $\text{VO}_{2\text{pred}}$ ;  $n = 8$  studies) used in this analysis with the remaining 21 studies being sourced through Web of Science, Google Scholar and cross referencing. The TM- $\text{VO}_{2\text{meas}}$  studies had a bias towards clinical outcomes related to CHF whereas the TM- $\text{VO}_{2\text{pred}}$  studies were frequently associated with long-term outcome (survival) in screening for CVD. Though multiple search strategies were used to obtain studies meeting selection criteria it is acknowledged that even with rigorous attention to search detail, suitable studies may have been missed.

Fig 2 details the study selection process at the completion of data acquisition in March 2013.

## Statistical analysis

Categorical variables were expressed as numbers and percentages with continuous variables expressed as mean  $\pm$  standard deviation. Student's paired t-test was used to compare HRI- $\text{VO}_2$



**Fig 2. Study selection process used for data acquisition.**

doi:10.1371/journal.pone.0166608.g002

against both TM-VO<sub>2meas</sub> and TM-VO<sub>2pred</sub>. Results are expressed in two formats, namely 1) pooled data for each of TM-VO<sub>2meas</sub> and TM-VO<sub>2pred</sub> against HRI-VO<sub>2</sub> expressed as group means and shown in the form of line of identity and Bland Altman plots [34] and 2) CRF data shown in tertiles for both TM-VO<sub>2meas</sub> and TM-VO<sub>2pred</sub> groups against HRI-VO<sub>2</sub>.

## Results

### Studies used in the analyses

There were 11,477 subjects in the 20 TM-VO<sub>2meas</sub> studies (range 110 to 4631, median 337) and, with each study mean VO<sub>2meas</sub> value representing a data point, there was a total of 45 data points. There was a considerably larger number of subjects at 105,044 (range 772 to 22,275, median 3,736) in the 20 TM-VO<sub>2pred</sub> studies and, with each study mean VO<sub>2pred</sub> value

representing a data point, there were 57 data points. Age and gender distribution was similar for the TM-VO<sub>2meas</sub> (51.0 years and 64.9% males) and TM-VO<sub>2pred</sub> groups (52.9 years and 71.0% males).

The principal details of the 40 treadmill studies used in the analysis are outlined in [Table 1](#). These include the test protocol, use of handrail support and the health status of participants. Of the 20 TM-VO<sub>2meas</sub> studies, 14 (70%) involved subjects with CHF and all 14 used protocols other than the standard Bruce protocol [35]. The design of these alternate protocols reduced the stage increment of VO<sub>2</sub> usually to 2 METs or less with certain ramp protocols having increments of less than 1 MET per minute. In only two of the TM-VO<sub>2meas</sub> studies was hand rail support mentioned, being 'not permitted' in one study (Dressendorfer [36]) and 'discouraged' in the other (Oliveira [37]).

Typically, subjects with known or suspected CVD or with significant cardiovascular risk factors were involved in the TM-VO<sub>2pred</sub> studies ([Table 1](#)). A Bruce protocol, either as the standard or a modified protocol, was used in 13 (65%) of the 20 TM-VO<sub>2pred</sub> studies. With TM-VO<sub>2pred</sub> studies, the use of handrail support was defined in seven studies (35%) and not stated in the remaining 13 studies. Descriptors of handrail support used for these seven studies were 'discouraged' in 3 studies, 'not permitted' in 3 studies and 'light hand rail support' in 1 study. Predictive treadmill equations in TM-VO<sub>2pred</sub> studies were either given or referenced in only 12 (60%) of the 20 studies.

## Characterization of study groups

**A. Group means: oxygen consumption and heart rate.** The mean TM-VO<sub>2pred</sub> reported in the 20 studies was 8.12 METS; the mean TM-VO<sub>2meas</sub> reported in the 20 studies was 6.51 METS, a difference of 1.61 Mets or 24.7% ([Table 2](#)). The mean HR<sub>rest</sub> with TM-VO<sub>2pred</sub> was 75.6 beats·min<sup>-1</sup> and with TM-VO<sub>2meas</sub> was 77.6 beats·min<sup>-1</sup>; the mean HR<sub>max</sub> for TM-VO<sub>2pred</sub> 146.3 beats·min<sup>-1</sup> and TM-VO<sub>2meas</sub> 147.1 beats·min<sup>-1</sup> ([Table 2](#)). However, the absolute differences in group means for HR<sub>rest</sub> and HR<sub>max</sub> between TM-VO<sub>2pred</sub> and TM-VO<sub>2meas</sub> were small at 2.0 beats·min<sup>-1</sup> for HR<sub>rest</sub> and only 0.8 beat·min<sup>-1</sup> for HR<sub>max</sub> ([Table 2](#)).

Alternatively if VO<sub>2peak</sub> is determined by HRI-VO<sub>2</sub> the difference between TM-VO<sub>2pred</sub> and TM-VO<sub>2meas</sub> is reduced to only 0.17 MET or 2.6% (TM-VO<sub>2pred</sub> 6.71 METs, TM-VO<sub>2meas</sub> 6.54 METs), a not unexpected result in view of the small differences in HR<sub>rest</sub> and HR<sub>max</sub> between these two groups ([Table 2](#)).

**B. Comparison of measured VO<sub>2</sub> and predicted VO<sub>2</sub> versus VO<sub>2</sub> predicted by HRI.** When using the HRI to calculate VO<sub>2peak</sub>, there was no significant difference (0.4%,  $p = 0.84$ ) in the pooled VO<sub>2</sub> data with mean ( $\pm$  SD) MET values of 6.51 ( $\pm 2.25$ ) for TM-VO<sub>2meas</sub> and 6.54 ( $\pm 2.28$ ) for HRI-VO<sub>2</sub> ([Fig 3A](#)). However, a highly significant difference (21.1%,  $p < 0.001$ ) was seen between TM-VO<sub>2pred</sub> and HRI-VO<sub>2</sub> with respective values of 8.12 ( $\pm 1.85$ ) METs and 6.71 ( $\pm 1.92$ ) METs ([Fig 3A](#)).

Even when expressed in tertiles based on HRI-VO<sub>2</sub>, there were no significant differences between TM-VO<sub>2meas</sub> and HRI-VO<sub>2</sub> by VO<sub>2</sub> tertile; tertile 1, 2.4% ( $p = 0.42$ ), tertile 2, -4.1% ( $p = 0.18$ ) and tertile 3, 0.7% ( $p = 0.83$ ) ([Fig 3B](#)). By comparison, each tertile for the TM-VO<sub>2pred</sub> groups showed a significant difference from HRI-VO<sub>2</sub>; tertile 1, 31.2% ( $p < 0.001$ ), tertile 2, 29.6% ( $p < 0.001$ ) and tertile 3, 9.1% ( $p = 0.03$ ) ([Fig 3C](#)).

The plot of TM-VO<sub>2meas</sub> against HRI-VO<sub>2</sub> shows a uniform distribution around the line of identity with the Bland Altman plot suggesting that there is no bias between these two separate methods of determining VO<sub>2peak</sub> ([Fig 4A and 4B](#)). However, a similar line of identity plot for TM-VO<sub>2pred</sub> against HRI-VO<sub>2</sub> indicates a strong bias with the Bland Altman plot indicating a systematic error in support of over-prediction of TM-VO<sub>2pred</sub> ([Fig 5A and 5B](#)).

**Table 1. Description of studies, patient diagnosis, and test protocol in which oxygen uptake was either measured or predicted using a prediction equation (Pred EQ).**

First Author	Year	n	Age (years)	Male%	Category	Test	Rail support	Pred EQ
<b>Measured VO<sub>2</sub></b>								
Bard	2006	355	51	72	CHF	ramp	ns	
Diller	2006	727	33	52	ACHD	MB	ns	
Dressendorfer	1993	182	57	100	CAD	MB	NP	
Elmariah	2006	594	52	72	CHF	ramp	ns	
Harrington	1997	131	59	100	CHF, H	MB	ns	
Ingle	2007	394	65	74	CHF	MB	ns	
Jorde	2008	278	52	77	CHF	Na	ns	
Kohrt	1991	110	64	50	H	B,O	ns	
Kubrychtova	2009	712	56	72	CHF	O	ns	
Lanier	2012	320	52	75	CHF	Na	ns	
McDonough	1970	144	51	100	H	B	ns	
Nes	2012	4631	48	49	H	ramp	ns	
Oliveira	2009	948	57	100	CPD, H	ramp	DIS	
Osada	1998	154	52	75	CHF	MB, MNa	ns	
Peterson	2003	369	51	72	CHF	O	ns	
Robbins	1999	487	52	71	CHF, H	Na	ns	
Schalcher	2003	146	52	88	CHF	ramp	ns	
Stolker	2006	221	49	68	CHF	O	ns	
Williams	2001	219	56	76	CHF	B, MB	ns	
Witte	2006	355	66	68	CHF, H	MB	ns	
<b>Predicted VO<sub>2</sub></b>								
Adabag	2008	12555	46	100	CAD <sup>°</sup>	B	ns	EQ-S
Aijaz	2008	10897	54	75	CVD, CVD <sup>°</sup>	B	ns	ns
Arruda-Olson	2002	5798	62	57	CAD, CAD <sup>°</sup>	B, MB, Na	ns	ns
Carnethon	2003	4487	25	45	H	MBa	ns	ns
Cheng	2003	2333	49	100	DM	MBa	ns	EQ-S
Elhendy	2001	1618	55	35	CAD <sup>°</sup>	B, MB, Na	ns	ns
Gulati	2010	5437	52	0	CAD <sup>°</sup>	B	LS	EQ-R
Kim	2007	22275	51	59	CVD <sup>°</sup>	B, MB, O	NP	EQ-R
Kokkinos	2009	4631	61	100	HT	B, ramp	DIS	EQ-R
Lai	2004	5625	59	100	CVD <sup>°</sup>	ramp, O	ns	EQ-R
Lauer	1999	2953	58	64	CVD <sup>°</sup> , CVD <sup>°</sup>	B, MB	NP	EQ-R
Lipinski	2005	1914	52	100	CAD, CHF, CAD <sup>°</sup>	ramp, O	ns	ns
Mahenthiran	2005	1268	60	52	CAD, CAD <sup>°</sup>	B	ns	ns
McAuley	2007	6876	58	97	CAD, CAD <sup>°</sup>	ramp	DIS	EQ-R
Mora	2003	2985	47	0	CAD <sup>°</sup>	B	ns	EQ-R
Morrow	1993	2546	59	100	CAD <sup>°</sup> , CAD, CHF	ramp, O	ns	EQ-R
Myers	2002	6213	59	100	CAD, CAD <sup>°</sup>	ramp	DIS	EQ-R
Negishi	2013	914	56	56	DM	B, MB	NP	EQ-R
Peteiro	2010	2947	62	61	CAD, CAD <sup>°</sup>	B, MB, Na	ns	ns
Shaw	2011	772	63	0	CAD <sup>°</sup>	B, MB	ns	ns

References are available in the supplementary digital content. **Category:** ACHD, adult congenital heart disease; CAD, coronary artery disease (CAD<sup>°</sup>, absent; CAD<sup>°</sup>, suspected); CHF, congestive heart failure; CPD, cardiopulmonary disease; CVD, cardiovascular disease; DM, diabetes mellitus; H, healthy; HT, hypertension. **Treadmill test:** B, Bruce protocol; Ba, Balke protocol; Na, Naughton protocol; ramp, ramp protocol; M, modified protocol; O, other protocol; **Rail support:** ns, not stated; NP, not permitted; DIS, discouraged; LS, light support; **Equation:** EQ-S, stated equation; EQ-R, referenced equation; ns, not stated.

doi:10.1371/journal.pone.0166608.t001



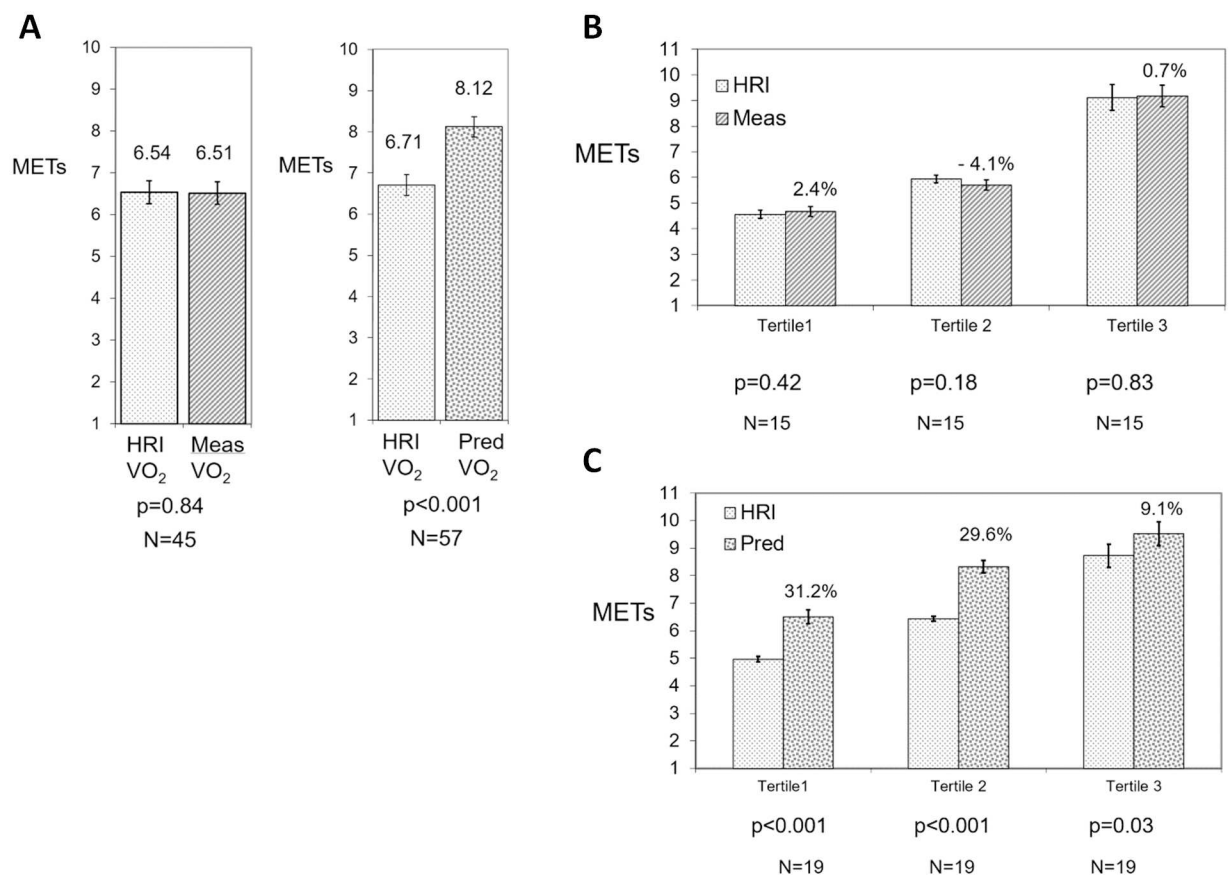
**Table 2. Heart rate and oxygen consumption data for TM-VO<sub>2meas</sub> and TM-VO<sub>2pred</sub>.** Group mean (± 1SD) heart rate (HR) and oxygen consumption (VO<sub>2</sub>) data. HR<sub>rest</sub>, HR<sub>peak</sub>, HRI-VO<sub>2</sub> and VO<sub>2peak</sub> for TM-VO<sub>2meas</sub> and TM-VO<sub>2pred</sub>.

	Studies	Data points	HR <sub>rest</sub> beats·min <sup>-1</sup>	HR <sub>peak</sub> beats·min <sup>-1</sup>	VO <sub>2peak</sub> METs	HRI-VO <sub>2</sub> METs
TM-VO <sub>2pred</sub>	20	57	75.6 ± 5.3	146.3 ± 16.6	8.12 ± 1.85	6.71 ± 1.92
TM-VO <sub>2meas</sub>	20	45	77.6 ± 7.7	147.1 ± 18.8	6.51 ± 2.25	6.54 ± 2.28

doi:10.1371/journal.pone.0166608.t002

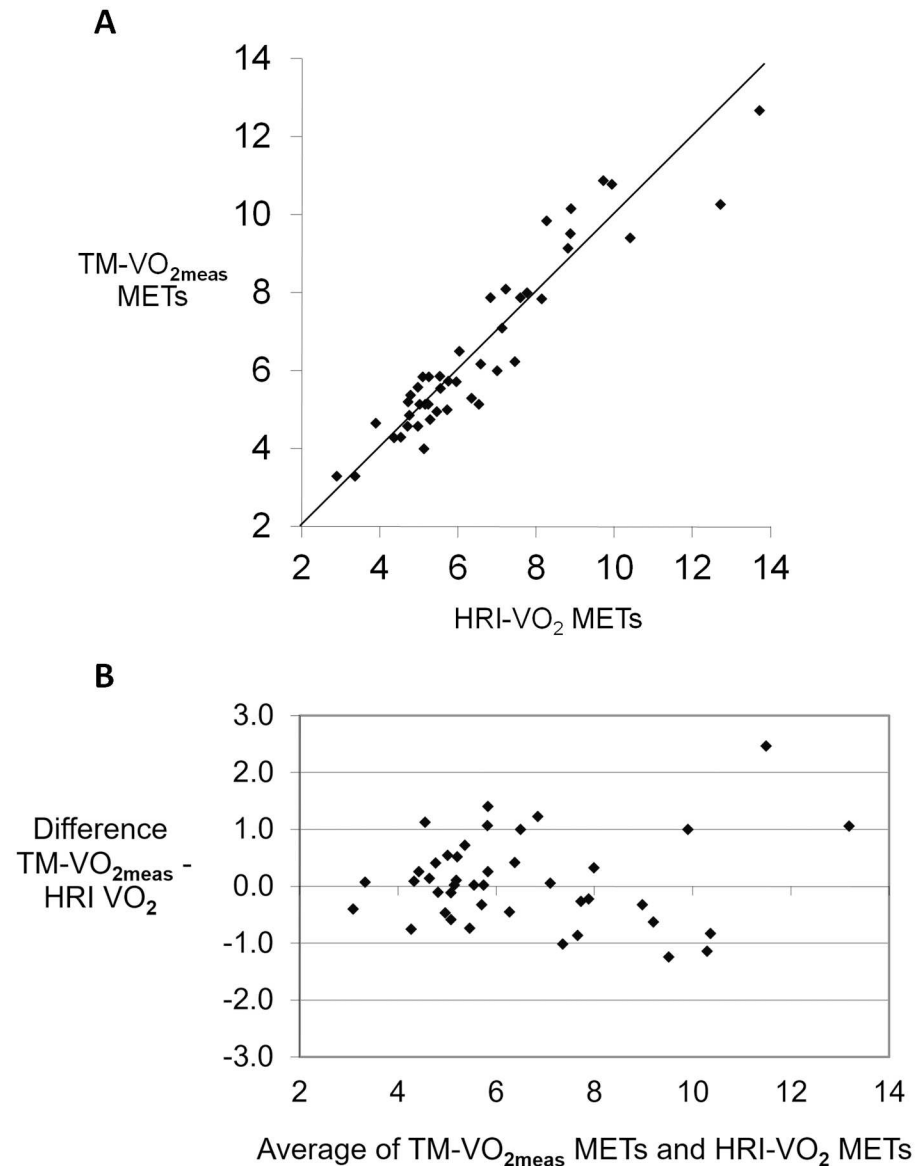
### Discussion

It is crucial to have high quality CRF data for use in epidemiological studies as management strategies involving both pharmacological and lifestyle intervention rely on this accuracy. The utility of the HRI equation [32] as a surrogate measure of VO<sub>2</sub> expressed in METs is confirmed in this study when assessed against VO<sub>2peak</sub> for both TM-VO<sub>2meas</sub> measured with conventional gas analysis equipment and for TM-VO<sub>2pred</sub> predicted from equations based on treadmill speed, incline or treadmill time. A close agreement between HRI-VO<sub>2</sub> and TM-VO<sub>2meas</sub> was observed in the 20 TM-VO<sub>2meas</sub> studies with only a 0.4% difference (p = 0.84) between group means. By comparison, a highly significant 21.1% (p<0.001) over-prediction of VO<sub>2peak</sub> was observed when comparing HRI-VO<sub>2</sub> against TM-VO<sub>2pred</sub> in the 20 TM-VO<sub>2pred</sub> studies. The



**Fig 3. Comparison of pooled data from 20 studies for TM-VO<sub>2meas</sub> and TM-VO<sub>2pred</sub> against HRI-VO<sub>2</sub>.** A. Comparison of group mean data for 20 TM-VO<sub>2meas</sub> and TM-VO<sub>2pred</sub> studies against HRI-VO<sub>2</sub> (mean ± SE), B. Comparison of cardiorespiratory fitness tertiles from 20 studies for TM-VO<sub>2meas</sub> against HRI-VO<sub>2</sub> (mean ± SE). Percentage difference between TM-VO<sub>2meas</sub> and HRI-VO<sub>2</sub> shown within figure and C. Comparison of cardiorespiratory fitness tertiles from 20 studies for TM-VO<sub>2pred</sub> against HRI-VO<sub>2</sub> (mean ± SE). Percentage difference between TM-VO<sub>2pred</sub> and HRI-VO<sub>2</sub> shown within figure.

doi:10.1371/journal.pone.0166608.g003



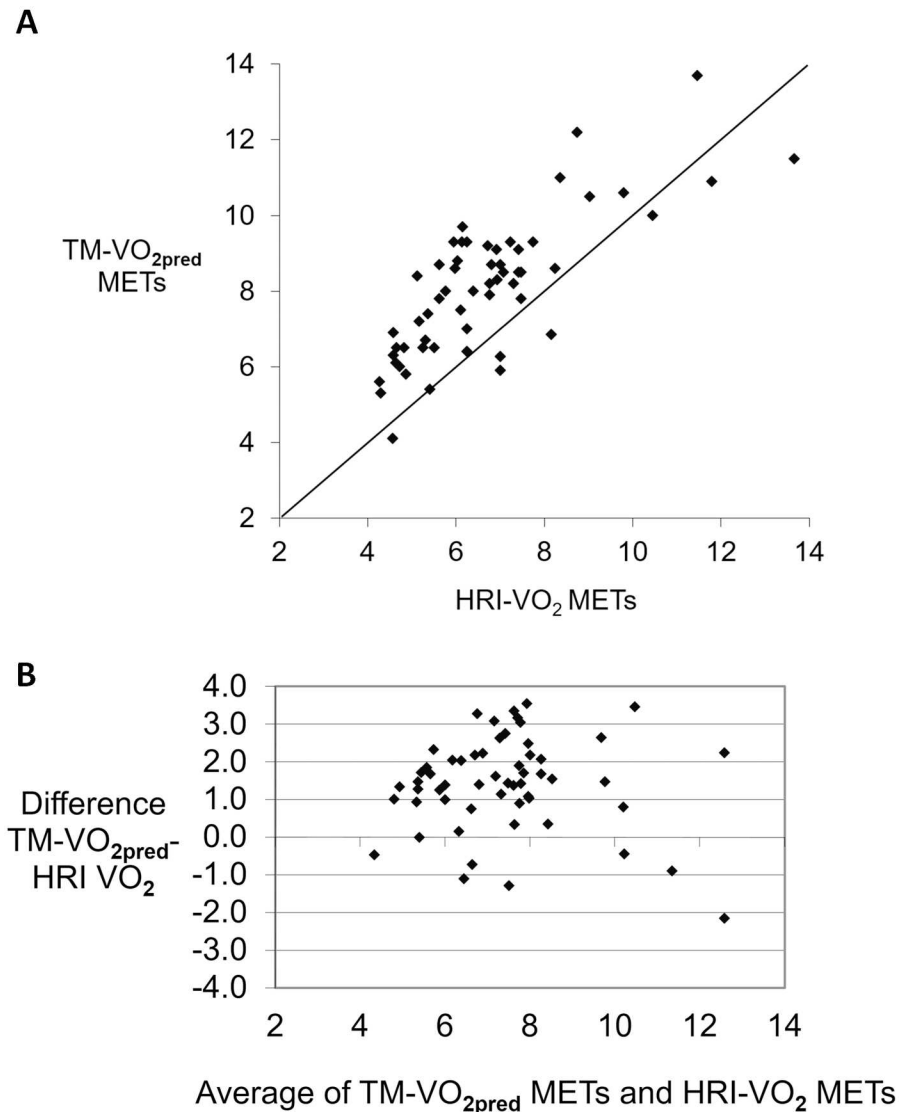
**Fig 4. Line of identity and Bland Altman plot for TM-VO<sub>2meas</sub>.** A. Line of identity for TM-VO<sub>2meas</sub> and B. against Bland Altman plot for TM-VO<sub>2meas</sub> against HRI-VO<sub>2</sub>.

doi:10.1371/journal.pone.0166608.g004

magnitude of the potential error using TM-VO<sub>2pred</sub> challenges the current methods of treadmill prediction of CRF which appear to lead to overestimation of CRF and potentially to false prognostic classification.

If the magnitude of the disparity between HRI-VO<sub>2</sub> and TM-VO<sub>2pred</sub> as shown in this study is, for example, applied to the outcome data of CRF as expressed in METs in the meta-analysis by Kodama [2], there is a strong likelihood of a false classification based on the over-prediction of CRF. For example, in treadmill studies investigating the effect of handrail support, a practice that lengthens treadmill time, VO<sub>2peak</sub> is over-predicted by 20% to 30% [9–13,17] which would lead to a potentially false prognostic classification of CRF. To correct for the consistently observed over-prediction of VO<sub>2peak</sub> of around 20% resulting from the use of





**Fig 5. Line of identity and Bland Altman plot for TM-VO<sub>2pred</sub>.** A. Line of identity for TM-VO<sub>2pred</sub> and B. against Bland Altman plot for TM-VO<sub>2pred</sub> against HRI-VO<sub>2</sub>.

doi:10.1371/journal.pone.0166608.g005

handrail support, Foster has developed simple modifications of the ACSM equations for use when handrail support is observed during treadmill testing [17]. None of the 20 TM-VO<sub>2pred</sub> studies used in this analysis referenced use of the Foster or similar equations to correct for observed handrail support. This prediction error could potentially apply to other published studies that express results in the form of survival tables and Kaplan-Meier curves. The measurement of CRF is not only limited to CVD. CRF also defines long-term risk in both healthy subjects and other common medical conditions, such as stroke [38], dementia [39] and diabetes mellitus [40]. In the TM-VO<sub>2pred</sub> group of studies, the smallest difference (9.1%) between HRI-VO<sub>2</sub> and TM-VO<sub>2pred</sub> was observed in the highest CRF tertile. Presumably, the fittest subjects find less difficulty with treadmill walking and so have less need for handrail support. Conversely, the least fit, i.e., the lowest tertile, are most likely to utilize handrail support, even when instructed otherwise, and, in the present study, they demonstrated a 31.2% difference

between HRI-VO<sub>2</sub> and TM-VO<sub>2pred</sub>. Results from the HUNT 3 Fitness Study also noted the greatest overestimation of VO<sub>2peak</sub> in the least fit subjects [18].

Collectively the 20 TM-VO<sub>2pred</sub> studies used in this analysis involve a tenfold greater number of subjects when compared with the 20 TM-VO<sub>2meas</sub> studies, whether considering the total number of subjects (105,044 TM-VO<sub>2pred</sub> versus 11,477 TM-VO<sub>2meas</sub>) or the median number (3,736 TM-VO<sub>2pred</sub> versus 337 TM-VO<sub>2meas</sub>). This observation indicates an inherent bias in using predicted VO<sub>2</sub> studies for epidemiological purposes. In recognizing the need for high quality population CRF data, the Fitness Registry and the Importance of Exercise: A National Database (FRIEND) was established in 2014 [41]. A recent publication from this group has provided age-related reference standards of CRF from 7783 tests in which VO<sub>2max</sub> was determined by gas analysis, the authors highlighting the shortcomings of using TM-VO<sub>2pred</sub> largely because of over-prediction of VO<sub>2max</sub> associated with hand rail support [42]. Their statement together with the observations in the present review suggest that, for the continued use of TM-VO<sub>2pred</sub> data, a reappraisal of current methods used for prediction of VO<sub>2peak</sub> warrants consideration.

One important question arising from this analysis is the value of using maximal HRI to predict VO<sub>2peak</sub> from HR derived values (rest and peak) as opposed to treadmill parameters (speed, incline or treadmill time). When calculating maximal HRI, two independent predictors of future CVD risk, namely an estimated VO<sub>2peak</sub> [2,43] and HR<sub>rest</sub> [44] are incorporated within the HRI. The maximal HRI is based on two measured values of HR and, when used as an index, there is minimal predictive error especially when compared to VO<sub>2pred</sub> using equations based on speed, incline or treadmill time. As a 1.0 MET increment corresponds to a HRI increment of 0.167, Kaplan-Meier curves ranging from <5 to >10 METs have a corresponding HRI range from <1.67 to >2.50 (e.g., 5 METs = Rest [HRI = 1] + 4 METs [HRI = 4 x 0.167] = 1.67). In considering a range of activity from rest (1.0 MET) to the maximum aerobic performance of an elite athlete (e.g. 19 METs), the corresponding range of HRI would be from 1 to 4. The simplicity of calculating HRI together with the range of index used for clinical evaluation suggests that it could provide a useful addition to the assessment of CRF. To illustrate this, a range of 5, 10 and 15 MET levels have corresponding HRIs of 1.67, 2.5 and 3.33.

## Study Limitations

This review has used the simple concept of HRI as a surrogate measure of VO<sub>2</sub>. The equation was established from aggregate data acquired from 60 studies. In applying the HR index to this analysis, we have compared aggregate data from TM-VO<sub>2pred</sub> and TM-VO<sub>2meas</sub> against HRI-VO<sub>2</sub> with no intention of indicating the individual predictive accuracy of the equation. Ideally the use of individual, as opposed to aggregate data would have been preferable but it was beyond the capability of this analysis.

## Conclusions

The usefulness of CRF is well established for assessing CV risk with treadmill testing providing a simple and convenient method of assessing CRF. The aggregate analysis used in this study shows a close relationship, i.e., a non-significant 0.4% difference, between HRI-VO<sub>2</sub> and TM-VO<sub>2meas</sub> but a large and highly significant 21.1% difference between HRI-VO<sub>2</sub> and TM-VO<sub>2pred</sub>. This overestimation of TM-VO<sub>2pred</sub>, and so CRF, challenges the validity of predicting VO<sub>2peak</sub> from equations based on treadmill speed, incline or protocol time when attempting to document a link between CRF and long-term morbidity/mortality.

## Supporting Information

**S1 File. Supplementary Reference List– 40 treadmill studies.** File listing the 40 treadmill studies used for analysis.  
(RTF)

## Acknowledgments

Neither author has any conflict of interest nor was any financial assistance provided for this study. The authors thank the staff of the Gold Coast Hospital Library for their invaluable help in sourcing reference materials and assistance offered in developing search strategies provided by Ms Sarah Thorning, senior librarian, Gold Coast University Hospital Library and Professor Paul Glasziou, Department of Evidence Based Medicine, Faculty of Health Sciences and Medicine, Bond University, Queensland Australia.

## Author Contributions

**Conceptualization:** JRW NBO.

**Formal analysis:** JRW NBO.

**Investigation:** JRW.

**Methodology:** JRW NBO.

**Visualization:** JRW NBO.

**Writing – original draft:** JRW NBO.

**Writing – review & editing:** JRW NBO.

## References

1. ACSM (2000) ACSM's Guidelines for Exercise Testing and Prescription. 6th edition. Philadelphia, PA: Lippincott Williams & Wilkins.
2. Kodama S, Saito K, Tanaka S, Maki M, Yachi Y, Asumi M, et al. (2009) Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women. *JAMA* 301: 2024–2035. doi: [10.1001/jama.2009.681](https://doi.org/10.1001/jama.2009.681) PMID: [19454641](https://pubmed.ncbi.nlm.nih.gov/19454641/)
3. Echouffo-Tcheugui JB, Butler J, Yancy CW, Fonarow GC (2015) Association of physical activity or fitness with incident heart failure: A systematic review and meta-analysis. *Circ Heart Fail* 8: 853–861. doi: [10.1161/CIRCHEARTFAILURE.115.002070](https://doi.org/10.1161/CIRCHEARTFAILURE.115.002070) PMID: [26175539](https://pubmed.ncbi.nlm.nih.gov/26175539/)
4. Keteyian SJ, Brawner CA, Savage PD, Ehrman JK, Schairer J, Divine G, et al. (2008) Peak aerobic capacity predicts prognosis in patients with coronary heart disease. *Am Heart J* 156: 292–300. doi: [10.1016/j.ahj.2008.03.017](https://doi.org/10.1016/j.ahj.2008.03.017) PMID: [18657659](https://pubmed.ncbi.nlm.nih.gov/18657659/)
5. McNeer JF, Margolis JR, Lee KL, Kisslo JA, Peter RH, Kong Y, et al. (1978) The role of the exercise test in the evaluation of patients for ischemic heart disease. *Circulation* 57: 64–70. PMID: [618399](https://pubmed.ncbi.nlm.nih.gov/618399/)
6. Berry JD, Pandey A, Gao A, Leonard D, Farzaneh-Far R, Ayers C, et al. (2013) Physical fitness and risk for heart failure and coronary artery disease. *Circ Heart Fail* 6: 627–634. PMID: [23677924](https://pubmed.ncbi.nlm.nih.gov/23677924/)
7. Gander JC, Sui X, Hébert JR, Hazlett LJ, Cai B, Lavie CJ, et al. (2015) Association of cardiorespiratory fitness with coronary heart disease in asymptomatic men. *Mayo Clin Proc* 90: 1372–1379. doi: [10.1016/j.mayocp.2015.07.017](https://doi.org/10.1016/j.mayocp.2015.07.017) PMID: [26434963](https://pubmed.ncbi.nlm.nih.gov/26434963/)
8. Khan H, Kunutsor S, Rauramaa R, Savonen K, Kalogeropoulos AP, Georgiopoulos VV, et al. (2014) Cardiorespiratory fitness and risk of heart failure: a population-based follow-up study. *Eur J Heart Fail* 16: 180–188. doi: [10.1111/ehhf.37](https://doi.org/10.1111/ehhf.37) PMID: [24464981](https://pubmed.ncbi.nlm.nih.gov/24464981/)
9. Pinkstaff S, Peberdy MA, Kontos MC, Fabiato A, Finucane S, Arena R (2011) Overestimation of aerobic capacity with the bruce treadmill protocol in patients being assessed for suspected myocardial ischemia. *J Cardiopulm Rehabil Prev* 31: 254–260. doi: [10.1097/HCR.0b013e318211e3ed](https://doi.org/10.1097/HCR.0b013e318211e3ed) PMID: [21427601](https://pubmed.ncbi.nlm.nih.gov/21427601/)

10. Berling JM, Foster C, Gibson M, Doberstein S, Porcari JP (2006) The effect of handrail support on oxygen uptake during steady state treadmill exercise. *J Cardiopulm Rehabil Prev* 26: 250.
11. Zeimet G, McNeil J, Hall J, Moss R (1985) Quantifiable changes in oxygen uptake, heart rate, and time to target heart rate when hand support is allowed during treadmill exercise. *J Cardiopulm Rehabil* 5: 525–530.
12. Haskell WL, Savin W, Oldridge N, DeBusk R (1982) Factors influencing estimated oxygen uptake during exercise testing soon after myocardial infarction. *Am J Cardiol* 50: 299–304. PMID: [7102560](#)
13. Manfre MJ, Yu GH, Varmá AA, Mallis GI, Kearney K, Karageorgis MA (1994) The effect of limited handrail support on total treadmill time and the prediction of VO<sub>2</sub> max. *Clin Cardiol* 17: 445–450. PMID: [7955592](#)
14. Foster C, Hare J, Taylor M, Goldstein T, Anholm J, Pollock M (1984) Prediction of oxygen uptake during exercise testing in cardiac patients and healthy volunteers. *J Cardiac Rehabil* 4: 537–542.
15. Foster C, Jackson AS, Pollock ML, Taylor MM, Hare J, Sennett SM, et al. (1984) Generalized equations for predicting functional capacity from treadmill performance. *Am Heart J* 107: 1229–1234. PMID: [6720550](#)
16. McConnell TR, Foster C, Conlin NC, Thompson NN (1991) Prediction of functional capacity during treadmill testing: effect of handrail support. *J Cardiopulm Rehabil Prev* 11: 255–260.
17. Foster C, Crowe A, Daines E, Dumit M, Green MA, Lettau S, et al. (1996) Predicting functional capacity during treadmill testing independent of exercise protocol. *Med Sci Sports Exerc* 28: 752–756. PMID: [8784763](#)
18. Loe H, Nes BM, Wisløff U (2016) Predicting VO<sub>2</sub> peak from Submaximal and Peak Exercise Models: The HUNT 3 Fitness Study, Norway. *PLoS ONE* 11: e0144873. doi: [10.1371/journal.pone.0144873](https://doi.org/10.1371/journal.pone.0144873) PMID: [26794677](#)
19. Buchfuhrer MJ, Hansen JE, Robinson TE, Sue DY, Wasserman K, Whipp BJ (1983) Optimizing the exercise protocol for cardiopulmonary assessment. *J Appl Physiol* 55: 1558–1564. PMID: [6643191](#)
20. Myers J, Buchanan N, Walsh D, Kraemer M, McAuley P, Hamilton-Wessler M, et al. (1991) Comparison of the ramp versus standard exercise protocols. *J Am Coll Cardiol* 17: 1334–1342. PMID: [2016451](#)
21. Opasich C, Myers J (2001) Working Group Report. Recommendations for exercise testing in chronic heart failure patients. *Eur Heart J* 22: 37–45. PMID: [11133208](#)
22. Roberts JM, Sullivan M, Froelicher VF, Genter F, Myers J (1984) Predicting oxygen uptake from treadmill testing in normal subjects and coronary artery disease patients. *Am Heart J* 108: 1454–1460. PMID: [6507241](#)
23. Alexander NB, Dengel DR, Olson RJ, Krajewski KM (2003) Oxygen-uptake (VO<sub>2</sub>) kinetics and functional mobility performance in impaired older adults. *J Gerontol A Biol Sci Med Sci* 58: M734–M739.
24. Mezzani A, Agostoni P, Cohen-Solal A, Corra U, Jegier A, Kouidi E, et al. (2009) Standards for the use of cardiopulmonary exercise testing for the functional evaluation of cardiac patients: a report from the Exercise Physiology Section of the European Association for Cardiovascular Prevention and Rehabilitation. *Eur J Cardiovasc Prev Rehabil* 16: 249–267. doi: [10.1097/HJR.0b013e32832914c8](https://doi.org/10.1097/HJR.0b013e32832914c8) PMID: [19440156](#)
25. Sullivan M, Genter F, Savvides M, Roberts M, Myers J, Froelicher V (1984) The reproducibility of hemodynamic, electrocardiographic, and gas exchange data during treadmill exercise in patients with stable angina pectoris. *Chest* 86: 375–382. PMID: [6467998](#)
26. Elborn J, Stanford C, Nicholls D (1990) Reproducibility of cardiopulmonary parameters during exercise in patients with chronic cardiac failure. The need for a preliminary test. *Eur Heart J* 11: 75–81. PMID: [2106439](#)
27. Levy WC, Maichel BA, Steele NP, Leclerc KM, Stratton JR (2004) Biomechanical efficiency is decreased in heart failure during low-level steady state and maximal ramp exercise. *Eur J Heart Fail* 6: 917–926. PMID: [15556054](#)
28. Hiatt WR, Cox L, Greenwalt M, Griffin A, Schechter C (2005) Quality of the assessment of primary and secondary endpoints in claudication and critical leg ischemia trials. *Vasc Med* 10: 207–213. PMID: [16235774](#)
29. Gagge AP, Burton AC, Bazett HC (1941) A practical system of units for the description of the heat exchange of man with his environment. *Science* 94: 428–430. PMID: [17758307](#)
30. Myers J, Prakash M, Froelicher V, Do D, Partington S, Atwood JE (2002) Exercise capacity and mortality among men referred for exercise testing. *N Engl J Med* 346: 793–801. PMID: [11893790](#)
31. Kubrychtova V, Olson TP, Bailey KR, Thapa P, Allison TG, Johnson BD (2009) Heart rate recovery and prognosis in heart failure patients. *Eur J Appl Physiol* 105: 37–45. doi: [10.1007/s00421-008-0870-z](https://doi.org/10.1007/s00421-008-0870-z) PMID: [18797918](#)

32. Wicks JR, Oldridge NB, Nielsen LK, Vickers CE (2011) HR index—a simple method for the prediction of oxygen uptake. *Med Sci Sports Exerc* 43: 2005–2012. doi: [10.1249/MSS.0b013e318217276e](https://doi.org/10.1249/MSS.0b013e318217276e) PMID: [21364476](https://pubmed.ncbi.nlm.nih.gov/21364476/)
33. Robinson KA, Dunn AG, Tsafnat G, Glasziou P (2014) Citation networks of related trials are often disconnected: implications for bidirectional citation searches. *J Clin Epidemiol* 67: 793–799. doi: [10.1016/j.jclinepi.2013.11.015](https://doi.org/10.1016/j.jclinepi.2013.11.015) PMID: [24725642](https://pubmed.ncbi.nlm.nih.gov/24725642/)
34. Bland JM, Altman D (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 327: 307–310.
35. Bruce RA, Kusumi F, Hosmer D (1973) Maximal oxygen intake and nomographic assessment of functional aerobic impairment in cardiovascular disease. *Am Heart J* 85: 546–562. PMID: [4632004](https://pubmed.ncbi.nlm.nih.gov/4632004/)
36. Dressendorfer R, Franklin B, Gordon S, Timmis G (1993) Resting oxygen uptake in coronary artery disease. Influence of chronic beta-blockade. *Chest* 104: 1269–1272. PMID: [8104769](https://pubmed.ncbi.nlm.nih.gov/8104769/)
37. Oliveira RB, Myers J, Araujo CGS, Abella J, Mandic S, Froelicher V (2009) Maximal exercise oxygen pulse as a predictor of mortality among male veterans referred for exercise testing. *Eur J Cardiovasc Prev Rehabil* 16: 358–364. doi: [10.1097/HJR.0b013e3283292fe8](https://doi.org/10.1097/HJR.0b013e3283292fe8) PMID: [19357518](https://pubmed.ncbi.nlm.nih.gov/19357518/)
38. Do Lee C, Folsom AR, Blair SN (2003) Physical activity and stroke risk a meta-analysis. *Stroke* 34: 2475–2481. doi: [10.1161/01.STR.0000091843.02517.9D](https://doi.org/10.1161/01.STR.0000091843.02517.9D) PMID: [14500932](https://pubmed.ncbi.nlm.nih.gov/14500932/)
39. DeFina LF, Willis BL, Radford NB, Gao A, Leonard D, Haskell WL, et al. (2013) The association between midlife cardiorespiratory fitness levels and later-life dementia: a cohort study. *Ann Intern Med* 158: 162–168. doi: [10.7326/0003-4819-158-3-201302050-00005](https://doi.org/10.7326/0003-4819-158-3-201302050-00005) PMID: [23381040](https://pubmed.ncbi.nlm.nih.gov/23381040/)
40. Helmrich SP, Ragland DR, Leung RW, Paffenbarger RS Jr (1991) Physical activity and reduced occurrence of non-insulin-dependent diabetes mellitus. *N Engl J Med* 325: 147–152. doi: [10.1056/NEJM199107183250302](https://doi.org/10.1056/NEJM199107183250302) PMID: [2052059](https://pubmed.ncbi.nlm.nih.gov/2052059/)
41. Kaminsky LA, Arena R, Beckie TM, Brubaker PH, Church TS, Forman DE, et al. (2013) The importance of cardiorespiratory fitness in the United States: the need for a national registry a policy statement from the American Heart Association. *Circulation* 127: 652–662. doi: [10.1161/CIR.0b013e31827ee100](https://doi.org/10.1161/CIR.0b013e31827ee100) PMID: [23295916](https://pubmed.ncbi.nlm.nih.gov/23295916/)
42. Kaminsky LA, Arena R, Myers J (2015) Reference standards for cardiorespiratory fitness measured with cardiopulmonary exercise testing: Data from the Fitness Registry and the Importance of Exercise National Database. *Mayo Clin Proc* 90: 1515–1523. doi: [10.1016/j.mayocp.2015.07.026](https://doi.org/10.1016/j.mayocp.2015.07.026) PMID: [26455884](https://pubmed.ncbi.nlm.nih.gov/26455884/)
43. Blair SN, Kohl HW 3rd, Paffenbarger RS Jr., Clark DG, Cooper KH, Gibbons LW (1989) Physical fitness and all-cause mortality. A prospective study of healthy men and women. *JAMA* 262: 2395–2401. PMID: [2795824](https://pubmed.ncbi.nlm.nih.gov/2795824/)
44. Fox K, Borer JS, Camm AJ, Danchin N, Ferrari R, Sendon JLL, et al. (2007) Resting heart rate in cardiovascular disease. *J Am Coll Cardiol* 50: 823–830. doi: [10.1016/j.jacc.2007.04.079](https://doi.org/10.1016/j.jacc.2007.04.079) PMID: [17719466](https://pubmed.ncbi.nlm.nih.gov/17719466/)