Article

Improved VO$_{2max}$ Estimation by Combining a Multiple Regression Model and Linear Extrapolation Method

Tomoaki Matsuo [1,2,*], Rina So [1,2] and Fumiko Murai [2]

1 Ergonomics Research Group, National Institute of Occupational Safety and Health, Japan, Kawasaki 214-8585, Japan
2 Research Center for Overwork-Related Disorders, National Institute of Occupational Safety and Health, Japan, Kawasaki 214-8585, Japan
* Correspondence: matsuo@h.jniosh.johas.go.jp

Abstract: Maximal oxygen consumption (VO$_{2max}$) is an important health indicator that is often estimated using a multiple regression model (MRM) or linear extrapolation method (LEM) with the heart rate (HR) during a step test. Nonetheless, both methods have inherent problems. This study investigated a VO$_{2max}$ estimation method that mitigates the weaknesses of these two methods. A total of 128 adults completed anthropometric measurements, a physical activity questionnaire, a step test with HR measurements, and a VO$_{2max}$ treadmill test. The MRM included step-test HR, age, sex, body mass index, and questionnaire scores, whereas the LEM included step-test HR, predetermined constant VO$_2$ values, and age-predicted maximal HR. Systematic differences between estimated and measured VO$_{2max}$ values were detected using Bland–Altman plots. The standard errors of the estimates of the MRM and LEM were 4.15 and 5.08 mL·kg^{-1}·min^{-1}, respectively. The range of 95% limits of agreement for the LEM was wider than that for the MRM. Fixed biases were not significant for both methods, and a significant proportional bias was observed only in the MRM. MRM bias was eliminated using the LEM application when the MRM-estimated VO$_{2max}$ was $\geq$45 mL·kg^{-1}·min^{-1}. In conclusion, substantial proportional bias in the MRM may be mitigated using the LEM within a limited range.

Keywords: cardiorespiratory fitness; exercise test; occupational health; physical fitness

1. Introduction

Cardiorespiratory fitness (CRF) strongly influences disease incidence and is therefore an important health indicator [1–3]. Nevertheless, CRF is not routinely assessed in clinical practice [4], primarily because of assessment challenges. The measurement of maximal oxygen consumption (VO$_{2max}$), the gold standard metric for CRF evaluation, is time-intensive and requires skilled examiners as well as specialized equipment (e.g., a treadmill/cycling ergometer and a gas analyzer). A field-based exercise test, such as the 20 m shuttle run [5], is an alternative for facilitating efficient mass screening of CRF. This validated shuttle run test is convenient for children because it can be performed in conjunction with a sporting event or physical education class in the school gymnasium or sports field; in contrast, it is much less convenient for adults because such facilities are rarely available in the workplace.

A questionnaire-based, non-exercise VO$_{2max}$ estimation [6–9] has been developed for population-based CRF assessments. However, such non-exercise methods have been shown by some previous studies [10,11] to have insufficient accuracy, suggesting that biological data such as the heart rate (HR) should be included in VO$_{2max}$ estimation because only the participants' subjective responses may not correspond with VO$_{2max}$ variance. Therefore, the use of HR in simple exercise tests such as the step test has the potential to be applied to CRF evaluation in adults, as it can be performed without skilled examiners, expensive equipment, a large exercise space, and participants' maximal exercise effort.

A multiple regression model (MRM) is often used in the development of estimated VO_{2max} (eVO_{2max}) equations, with measured VO_{2max} (mVO_{2max}) as the response variable and clinicodemographic variables and/or other measurement variables, such as HR and physical activity (PA) parameters, as the explanatory variables [6–9,12–14]. Two studies [15,16] showed that the addition of HR during a step test to a questionnaire-based, non-exercise MRM improved the accuracy of VO_{2max} estimation. On the other hand, the Chester step test [17], which is considered an appropriate evaluation tool for CRF in adults owing to its acceptable reliability [18,19], uses the linear extrapolation method (LEM). The LEM is based on the use of HR along with predetermined VO_2 values as constants during a stepping exercise. Subsequently, the best-fit line for HR and VO_2 is drawn, and the line is extrapolated up to eVO_{2max}, which corresponds to the age-predicted maximal HR.

The MRM and LEM play an important role in VO_{2max} estimation; however, each of these two methods has its own inherent problems. In particular, the MRM-predicted values are distorted at regression edges, creating systematic errors [20]. Typically, the MRM either underestimates the VO_{2max} in participants with high CRF or overestimates the VO_{2max} in those with low CRF [7–9,14]. On the other hand, the LEM usually does not introduce systematic errors but yields a larger error variance than the MRM [14], which is likely due to the inaccuracy of the predicted maximal HR [21,22]. A previous study [16] reported that the addition of HR data from a step test to a questionnaire-based, non-exercise MRM could improve the estimation power of eVO_{2max} in cross-sectional analyses. However, VO_{2max} estimation using the developed MRM was problematic in detecting increased VO_{2max} induced by lifestyle changes because the predicted values were distorted at the upper limit, resulting in a VO_{2max} underestimation among participants whose actual VO_{2max} was improved by their exercise lifestyle. Additionally, the study showed that the apparent systematic errors (underestimation of high values) seen in the MRM were absent in the LEM, that the LEM could detect the increased VO_{2max} induced by the exercise intervention, and that the estimated error of the LEM was generally greater than that of the MRM.

Therefore, a comparison of accuracy between VO_{2max} estimation using the MRM and VO_{2max} estimation using the LEM on data from the same participants may facilitate the construction of a combined model with the greatest possible accuracy. We speculated that the most accurate estimate of VO_{2max} could be achieved through the use of a combined model in which the LEM corrected for VO_{2max} underestimation by the MRM in participants with a high level of fitness.

In the present study, we applied the eVO_{2max} MRM equation, including questionnaire scores and step-test HRs, which was developed in a previous study [16], to participants who were different from those in the equation development study and investigated the cross-validity of this equation. Subsequently, using this dataset, we aimed to investigate an accurate VO_{2max} estimation method that combines the advantages of both MRM and LEM and mitigates the weaknesses of these two methods.

2. Materials and Methods

2.1. Participants

Adults aged 30–60 years who were residents of the Tokyo area, were working part-time or full-time at least three days per week, did not use any drugs that influence the autonomic nervous system (e.g., β-blockers), and had no medical conditions precluding VO_{2max} testing were included as study participants. The participants were recruited via website advertisements between June 2016 and March 2022. A total of 137 working adults (65 women and 72 men) participated in this study. The participants visited our laboratory on two occasions, with an interval of 1 week between these visits. On the first visit, the participants underwent anthropometric measurements and a step test and then responded to a questionnaire. On the second visit, the participants underwent a treadmill exercise test for actual VO_{2max} measurement. Nine participants were excluded because of insufficient data for analyses. Consequently, 128 participants (60 women and 68 men) were included in the analysis. This study was conducted in accordance with the principles embodied in

the Declaration of Helsinki. The study protocol was reviewed and approved by the Ethics Committee of the National Institute of Occupational Safety and Health, Japan (approval ID: H2744; date of approval: 31 March 2016). Written informed consent was obtained from all participants after providing a full explanation of the study aims and research protocols.

2.2. Measures

2.2.1. Anthropometric Measurements

Height was measured once to the nearest 0.1 cm, whereas body weight was measured once to the nearest 0.1 kg. The body mass index (BMI) was calculated as weight in kilograms divided by the square of height in meters.

2.2.2. Questionnaire

The questionnaire used in this study was previously validated for VO_{2max} estimation [7]. It consists of several questions regarding the frequency, duration, and intensity of PA, yielding a total PA score of 0–44 points. A higher PA score suggested a higher VO_{2max}. The questions and scores assigned to each are presented in Table S1.

2.2.3. Step Test

The step test used in this study, named the National Institute of Occupational Safety and Health, Japan (J-NIOSH) step test (JST), was previously validated for VO_{2max} estimation [14]. It comprises three 1 min stepping exercise stages followed by two 1 min recovery stages. In the exercise stages, the participants were required to step up and down a 30 cm step in time with a metronome beating at four times the step rate. The initial step rate (stage 1) was 15 steps/min (60 beats per min [bpm]) and was increased by 5 steps/min (20 bpm) for each subsequent stage (to a final step rate of 25 steps/min or 100 bpm on the metronome during stage 3). At the end of stage 3, the participants rested in the sitting position, and HR recordings were obtained at 1 and 2 min during the recovery stage. The protocol for the JST allowed the participants to skip stage 3 if the following two criteria were met at the end of stage 2: (i) HR of 80% of the age-predicted maximal HR (i.e., 220—age in years) and (ii) a rating of perceived exertion (RPE) of 17 on the Borg scale. However, none of the participants skipped stage 3 in this study. The HR index was calculated as follows: *HR index = (HR at exercise stage 3 − HR at exercise stage 1) + (HR at recovery stage 1 − HR at recovery stage 2)*.

2.2.4. mVO_{2max}

In order to directly measure the VO_{2max}, the participants performed an exhaustion-limited graded exercise test on a treadmill (AR200; Minato Medical Science, Osaka, Japan) using the Bruce protocol. During the test, the ventilation and expired gases were continuously measured using an open-circuit computerized indirect calorimeter (AE-310S; Minato Medical Science, Osaka, Japan) that was calibrated prior to each trial. The HR was monitored using an electrocardiogram (Life Scope, Nihon Kohden, Tokyo, Japan). The RPE was recorded using the Borg 6–20 scale. The highest 30 s average VO_2 value was defined as the VO_{2max} value when three of the following four criteria were satisfied: (i) the respiratory exchange ratio exceeded 1.10; (ii) the maximal HR was within 10 bpm of the age-predicted maximum (i.e., 220—age in years); (iii) the RPE exceeded 17; and (iv) the VO_2 reached a plateau despite further increases in workload [23,24].

2.2.5. eVO_{2max}

The MRM eVO_{2max} was calculated using the following equation derived from a previous study [16]: *eVO_{2max} (MRM) = 64.22 − (0.23 × age) + (5.74 × sex; women = 0 and men = 1) − (0.57 × BMI) + (0.19 × questionnaire's PA score) − (0.18 × JST's HR index)*.

The LEM eVO_{2max} was calculated based on a previous study [16]. Briefly, the LEM used a predetermined constant VO_2 value for each stage of the step test. The constant values were 4, 13, 19, 22, 17, and 8 $mL \cdot kg^{-1} \cdot min^{-1}$ for females and 4, 14, 20, 23, 18,

and 8 mL·kg^{-1}·min^{-1} for males at rest, during exercise stages 1–3, and during recovery stages 1 and 2, respectively. A scatter plot was constructed for each participant, with six predetermined constant VO$_2$ values (mL) for each stage of the JST (at rest, during exercise stages 1–3, and during recovery stages 1 and 2) on the *x*-axis and HR (bpm) on the *y*-axis. The best-fit line was calculated for the scatter plot, and the VO$_2$ value (*x*-axis) corresponding to the participant's age-predicted maximal HR (*y*-axis) was extrapolated as eVO$_{2max}$. The relation "208 − 0.7 × age" [25] was applied to the age-predicted maximal HR for calculation.

2.3. Data Analysis

Systematic differences between mVO$_{2max}$ and eVO$_{2max}$ were detected using Bland–Altman plots and linear regression analyses. Pearson correlation coefficients (*r*) were calculated to evaluate the relationship between mVO$_{2max}$ and eVO$_{2max}$. Error statistics, such as, constant error ($CE = \sum(Y - \hat{Y})/N$), standard error of the estimate ($SEE = SD_Y \sqrt{(1 - r^2)}$), and total error ($TE = \sqrt{\sum (Y - \hat{Y})^2 / N}$), where Y is mVO$_{2max}$ and $\hat{Y}$ is eVO$_{2max}$, were compared among the estimation methods. All statistical analyses were performed using SAS version 9.4 (SAS Institute Japan, Tokyo, Japan) and Prism 9 (GraphPad Software, San Diego, CA, USA), with statistical significance set at a two-tailed *p*-value of <0.05.

3. Results

The participants' average values of the anthropometric variables—mVO$_{2max}$ when using the treadmill, the questionnaire's PA score, and the JST's HR index—are summarized separately for women and men (Table 1). The participants of both sexes generally had normal body sizes and CRF levels.

Table 1. Characteristics of the study participants.

	Females (*N* = 60)			Males (*N* = 68)			Total (*N* = 128)		
Age, years	48.3	±	7.0	48.3	±	6.9	48.3	±	6.9
Height, cm	158.6	±	5.8	170.7	±	5.1	165.0	±	8.1
Body weight, kg	54.4	±	9.1	71.7	±	9.7	63.6	±	12.8
BMI, kg·m^{-2}	21.6	±	3.2	24.6	±	2.9	23.2	±	3.4
VO$_{2max}$, mL·kg^{-1}·min^{-1}	34.4	±	5.0	41.2	±	6.3	38.0	±	6.6
VO$_{2max}$, L·min^{-1}	1.86	±	0.30	2.93	±	0.44	2.42	±	0.66
Questionnaire's PA score, points	8.6	±	7.5	11.9	±	7.9	10.3	±	7.9
HR index, step test score, bpm	46.9	±	15.0	35.4	±	9.3	40.8	±	13.6

Values are presented as mean ± standard deviation. BMI, body mass index; HR, heart rate; PA, physical activity; and VO$_{2max}$, maximal oxygen consumption.

The Bland–Altman plots revealed systematic differences between mVO$_{2max}$ and eVO$_{2max}$. The MRM (Figure 1A) exhibited a non-significant fixed bias (−0.47 mL·kg^{-1}·min^{-1}, *p* = 0.21) and a significant proportional bias (*r* = −0.30, *p* < 0.01), which increased at higher VO$_{2max}$ values. The LEM (Figure 2A) also showed both fixed bias (0.89 mL·kg^{-1}·min^{-1}, *p* = 0.10) and proportional bias (*r* = 0.15, *p* = 0.10); however, neither was significant. The range of 95% limits of agreement (LoA) for the LEM was clearly wider than that for the MRM, suggesting a lower estimation accuracy. Indeed, the regression analysis showed a strong correlation between mVO$_{2max}$ and eVO$_{2max}$ values using both the MRM (*r* = 0.78, Figure 1B) and LEM (*r* = 0.64, Figure 2B); nonetheless, the correlation was stronger when the MRM was used. As shown in Figure 1B, the MRM also produced eVO$_{2max}$ values under the ideal line (*Y* = *X*) for 9 out of 11 participants with the highest fitness (VO$_{2max}$ ≥ 45.0 mL·kg^{-1}·min^{-1}, indicated by black triangles in all figures). On the other hand, several values derived using the LEM were near the ideal line for the same 11 participants (Figure 2B). Thus, the MRM tended to underestimate the VO$_{2max}$ in fit individuals, whereas the LEM showed no proportional bias but generally yielded less accurate VO$_{2max}$ estimates.
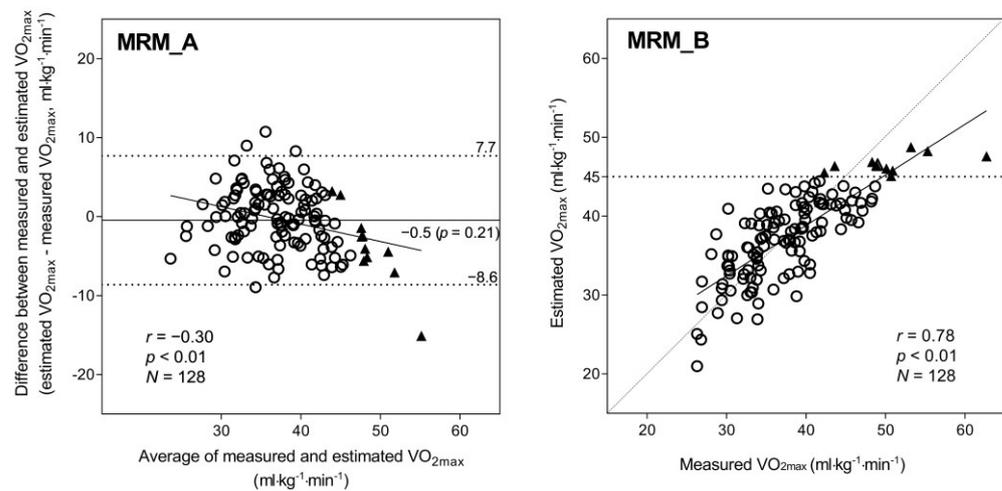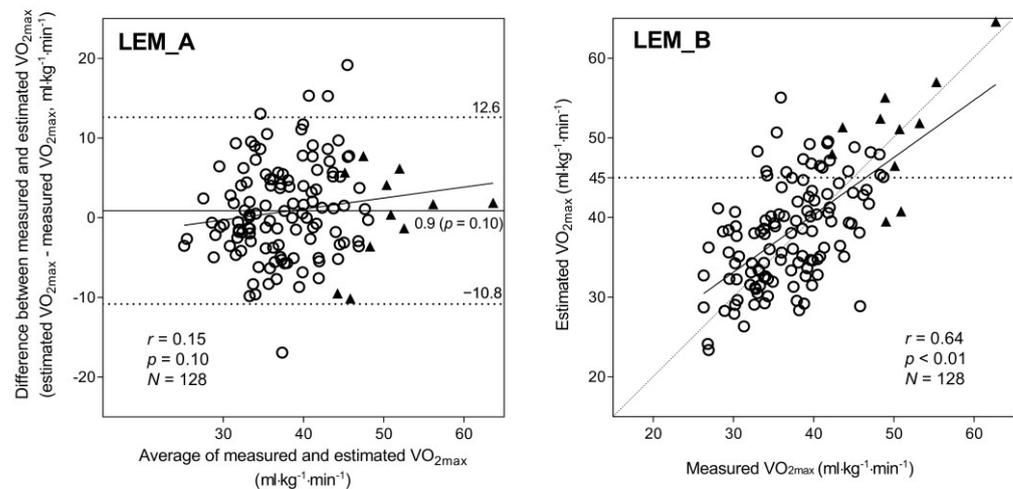
Figure 1. Bland–Altman plot (**A**) and correlation chart (**B**) with mVO_{2max} and eVO_{2max} using the MRM. In the Bland–Altman plots (**A**), the mean difference is shown as a solid line, and the 95% LoA is indicated as dashed lines; the regression line and correlation coefficient are also shown. In the correlation chart (**B**), the regression line and correlation coefficient are shown; the dashed line is the ideal line for the estimated and measured values. A total of 11 participants whose eVO_{2max} with the MRM exceeded 45.0 mL·kg^{-1}·min^{-1} are shown with a black triangle in (**A**,**B**). Abbreviations: eVO_{2max}, estimated VO_{2max}; LoA, limit of agreement; MRM, multiple regression model; mVO_{2max}, measured VO_{2max}; and VO_{2max}, maximal oxygen consumption.

Figure 2. Bland–Altman plot (**A**) and correlation chart (**B**) with mVO_{2max} and eVO_{2max} using the LEM. In the Bland–Altman plots (**A**), the mean difference is shown as a solid line, and the 95% LoA is indicated as dashed lines; the regression line and correlation coefficient are also shown. In the correlation chart (**B**), the regression line and correlation coefficient are shown; the dashed line is the ideal line for the estimated and measured values. A total of 11 participants whose eVO_{2max} with the MRM exceeded 45.0 mL·kg^{-1}·min^{-1} are shown with a black triangle in (**A**,**B**). Abbreviations: eVO_{2max}, estimated VO_{2max}; LEM, linear extrapolation method; LoA, limit of agreement; MRM, multiple regression model; mVO_{2max}, measured VO_{2max}; VO_{2max}, maximal oxygen consumption.

We speculated that the most accurate eVO_{2max} could be obtained through the use of a combined model in which the LEM corrected for VO_{2max} underestimation by the MRM in participants with high fitness. The procedure was as follows: (1) the eVO_{2max} for each participant was calculated using the MRM, and (2) if the eVO_{2max} exceeded 45.0 mL·kg^{-1}·min^{-1}, then it was recalculated using the LEM. Figure 3 shows the Bland–Altman plot (A) and scatter plot (B) constructed using this procedure. With this combined

method, neither the fixed bias (-0.11 mL·kg^{-1}·min^{-1}, $p = 0.76$) nor the proportional bias ($r = -0.05$, $p = 0.55$) were significant, and the correlation between mVO$_{2max}$ and eVO$_{2max}$ was still strong ($r = 0.80$).

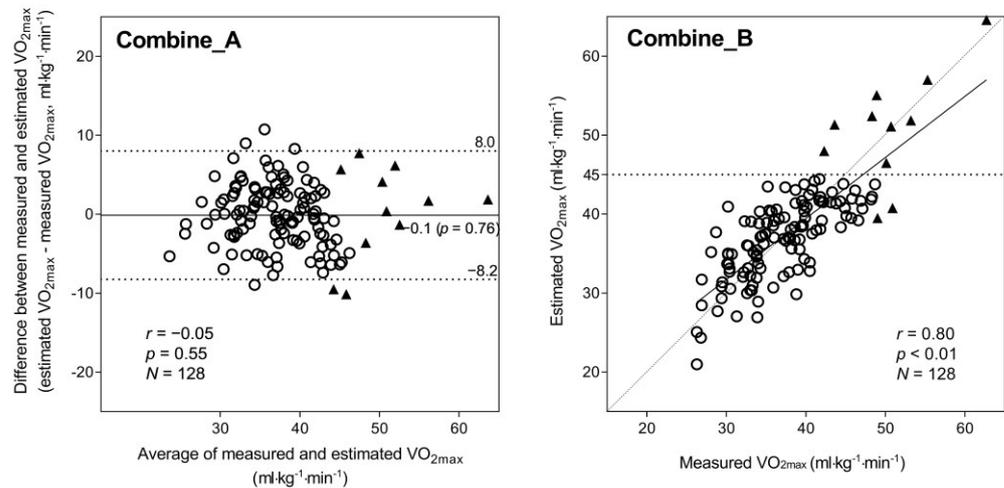**Figure 3.** Bland–Altman plot (**A**) and correlation chart (**B**) with mVO$_{2max}$ and eVO$_{2max}$ using the combined method. In the Bland–Altman plots (**A**), the mean difference is shown as a solid line, and the 95% LoA is indicated as dashed lines; the regression line and correlation coefficient are also shown. In the correlation chart (**B**), the regression line and correlation coefficient are shown; the dashed line is the ideal line for the estimated and measured values. A total of 11 participants whose eVO$_{2max}$ with the MRM exceeded 45.0 mL·kg^{-1}·min^{-1} are shown with a black triangle in (**A**,**B**). Abbreviations: eVO$_{2max}$, estimated VO$_{2max}$; LoA, limit of agreement; MRM, multiple regression model; mVO$_{2max}$, measured VO$_{2max}$; and VO$_{2max}$, maximal oxygen consumption.

Table 2 compares the other error statistics (CE, SEE, and TE) among the three estimation methods. The CE was not significant for all three methods; however, a larger difference between mVO$_{2max}$ and eVO$_{2max}$ was observed with the LEM than with the other two methods. The SEE increased in the order of LEM, MRM, and combined method. This tendency was also observed for the TE. Additionally, similar values were obtained between the SEE and TE for the MRM and combined method, whereas the TE was relatively higher than the SEE in the LEM.

Table 2. Measured versus estimated VO$_{2max}$ and error statistics ($N = 128$).

Measured VO$_{2max}$ (mL·kg^{-1}·min^{-1})				Estimated VO$_{2max}$ (mL·kg^{-1}·min^{-1})			CE (*p*-Value)				SEE	TE
38.0	±	6.6	MRM	37.6	±	5.4	0.5	±	4.2	(0.21)	4.15	4.17
			LEM	38.9	±	7.4	−0.9	±	6.0	(0.10)	5.08	6.02
			Combined	37.9	±	6.4	0.1	±	4.1	(0.76)	3.99	4.13

Values are presented as mean ± standard deviation. CE, constant error; SEE, standard error of the estimate; TE, total error; and VO$_{2max}$, maximal oxygen consumption.

4. Discussion

The high correlation coefficient ($r = 0.78$) between the MRM eVO$_{2max}$ and mVO$_{2max}$ indicated that the equation derived from 173 adults in a previous study [16] could be applied to an independent participant group (i.e., good stability of the equation). The MRM yielded greater estimation accuracy; however, a substantial proportional bias was also detected, resulting in a remarkable VO$_{2max}$ underestimation in participants with high fitness (Figure 1). On the other hand, the LEM demonstrated lower estimation accuracy than the MRM but no proportional bias—that is, the LEM did not underestimate VO$_{2max}$ in participants with high fitness (Figure 2). Interestingly, the combined method (Figure 3),

which replaces the MRM with the LEM for predicted values ≥ 45 mL·kg^{-1}·min^{-1} yields an optimal result, suggesting that the inaccuracies induced by the MRM and LEM could be partly mitigated by optimizing the statistical model. Thus, VO$_{2max}$ could be predicted with a relatively high degree of accuracy from a simple combination of morphometric measurements, questionnaire data, and step-test HR by combining MRM and LEM calculations.

One metabolic equivalent (MET) change in CRF is a meaningful value for disease prevention [26]. Hence, it would be a reasonable policy that the target SEE of VO$_{2max}$ estimation models should be ≤ 1.0 MET (3.5 mL·kg^{-1}·min^{-1}). Peterman et al. [11] reported SEEs ranging from 4.1 to 6.2 mL·kg^{-1} min^{-1} for non-exercise models, indicating substantial variation in the validity of these questionnaires for VO$_{2max}$ estimation. Some studies [7,8] suggested that eVO$_{2max}$ questionnaires should include items assessing the frequency, duration, and intensity of PA and that, owing to its greater influence on eVO$_{2max}$, the intensity score should be more heavily weighted than the frequency and duration scores. One of the previous studies [7] that used the same questionnaire as the present study reported an SEE of 4.29 mL·kg^{-1}·min^{-1} for the non-exercise MRM (including age, sex, BMI, and the questionnaire's PA score), as well as excellent test–retest reliability (intraclass correlation coefficient [ICC], 0.87; 95% confidence interval [CI], 0.82–0.91) for the PA score. Meanwhile, several studies have shown that evaluating "changes" in VO$_{2max}$ is difficult to perform using a questionnaire-based, non-exercise estimation method. For instance, Peterman et al. [27] found limited accuracy for 27 non-exercise prediction equations as compared with that for directly measured values in a cohort of 987 healthy adults. Similarly, Lannoy and Ross [10] reported limited VO$_{2peak}$ estimation accuracy, as compared with that for mVO$_{2peak}$, in 163 adults participating in a 24 week exercise intervention. These studies suggest that biological measures such as the HR may be required for accurate VO$_{2max}$ estimation.

The HR during the step test is often used for VO$_{2max}$ estimation. The increase in HR during a stepping exercise will be lower and will more rapidly return to baseline in individuals with high fitness than in those with low fitness [14]. The Queen's College step test [28], the Astrand–Ryhming step test [29], and the Chester step test [17] assume that the HR during or soon after exercise is lower in individuals with high fitness. On the other hand, others, such as the Harvard step test [30] and YMCA step test [31], assume that the HR decreases more rapidly during recovery in individuals with high fitness. However, the use of recovery HR alone may not be sufficiently sensitive for precisely predicting VO$_{2max}$ [14,32]. The JST's HR index in the present study captures the HR responses for VO$_{2max}$ estimation both during the stepping exercise and during the recovery period. A recent study [14] reported that the SEE for the JST (4.54 mL·kg^{-1}·min^{-1}) was relatively lower than the SEE for the Chester step test (4.99 mL·kg^{-1}·min^{-1}) and that the JST's HR index demonstrated fair-to-good test–retest reliability (ICC, 0.65; 95% CI, 0.53–0.74).

Recently, Webb et al. [15] reported that the addition of both the questionnaire score and step-test HR to the MRM (including sex and body weight) improved the accuracy of eVO$_{2max}$. Similarly, Matsuo et al. [16] showed that the addition of the JST's HR index to a questionnaire-based, non-exercise MRM (including age, sex, BMI, and the questionnaire's PA score) improved the accuracy of eVO$_{2max}$. Nevertheless, the study [16] suggested that this "synergistic effect" of the questionnaire's PA score and step-test HR for MRM eVO$_{2max}$ was not effective in detecting changes in mVO$_{2max}$ in an intervention experiment. That is, the study showed that mVO$_{2max}$, which was measured using a cycling ergometer and gas analyzer, increased by approximately 20% during the exercise training period and decreased by approximately 10% during the subsequent detraining period. The MRM eVO$_{2max}$ apparently underestimated the increase in mVO$_{2max}$ owing to a substantial proportional bias; therefore, the MRM could not detect changes in the CRF along with lifestyle modifications. Additionally, the study also showed that the LEM eVO$_{2max}$ presumably detected the increase in mVO$_{2max}$ while yielding a larger error variance than the MRM.

The results of the present study are consistent with the findings of a previous study [16]. The MRM exhibited good VO$_{2max}$ estimation accuracy in participants with average or

low fitness but significantly underestimated the VO_{2max} in participants with high fitness (Figure 1). On the other hand, with the LEM (Figure 2), the eVO_{2max} values of participants with high fitness were distributed near the ideal line (i.e., no underestimation); however, the LEM yielded larger errors across the VO_{2max} range, likely because of greater errors in the age-predicted maximal HR [21,22], which is heavily weighted in the LEM calculation process. Therefore, in the present study, we propose a method combining the advantages of both the MRM and LEM—that is, the generally higher accuracy of the MRM when eVO_{2max} is below a certain cut-off (in this case, 45 $mL \cdot kg^{-1} \cdot min^{-1}$) and the absence of proportional bias in the LEM estimates when VO_{2max} is above the cut-off. Consequently, the significant proportional bias observed for the MRM (Figure 1A) disappeared, and the wider LoA seen for the LEM (Figure 2A) was improved by the combined method (Figure 3A). Additionally, the combined method produced an improved scatter plot as well as the highest correlation coefficient among the three methods. Based on the principles of cross-validation analyses [33,34], the SEE and TE values should be calculated because the TE reflects the actual difference between measured and estimated values, whereas the SEE reflects only the variation in regression; similar values between the SEE and TE reflect a close approximation between the regression line and the line of identity. From this viewpoint, we compared the error statistics of the three methods (Table 2). Consequently, better CE, SEE, and TE values were observed for the combined method, and the difference between the SEE and TE for the combined method was not large. Thus, the favorable error statistic values demonstrate the superiority of the combined method.

This study has some limitations. First, the participants decided to participate after viewing our research advertisement, which might have likely introduced a selection bias in adults seeking to monitor their own CRF value. Second, mVO_{2max} was measured only once in each participant; hence, there was no estimation of intra-individual variability or possible measurement errors. Third, this study primarily included healthy participants and did not include any individuals having medical conditions precluding VO_{2max} testing or using medical drugs, such as β-blockers. Therefore, the estimation method proposed in this study might not apply to those patients and should be further investigated. Fourth, it should be considered that the favorable results obtained by the combined method may have happened by chance because of the unique dataset used in the present study. Similarly, the cut-point (≥ 45 $mL \cdot kg^{-1} \cdot min^{-1}$) may not be applicable to other populations. Therefore, the combined method should be validated using different subjects' data. Even so, we believe that the combined method presented in this study may be a feasible method for improving the accuracy of VO_{2max} estimation.

5. Conclusions

The present study showed that the MRM yielded higher estimation accuracy than the LEM but demonstrated marked proportional bias (i.e., VO_{2max} underestimation in participants with high fitness). However, this weakness could be mitigated through the use of the LEM within a limited range because the proportional bias was less likely to occur in the LEM. Further research is needed to confirm that this approach is applicable to other populations.

References

1. Myers, J.; Prakash, M.; Froelicher, V.; Do, D.; Partington, S.; Atwood, J.E. Exercise capacity and mortality among men referred for exercise testing. *N. Engl. J. Med.* **2002**, *346*, 793–801. [CrossRef] [PubMed]
2. Laukkanen, J.A.; Rauramaa, R.; Salonen, J.T.; Kurl, S. The predictive value of cardiorespiratory fitness combined with coronary risk evaluation and the risk of cardiovascular and all-cause death. *J. Intern. Med.* **2007**, *262*, 263–272. [CrossRef] [PubMed]
3. Gupta, S.; Rohatgi, A.; Ayers, C.R.; Willis, B.L.; Haskell, W.L.; Khera, A.; Drazner, M.H.; de Lemos, J.A.; Berry, J.D. Cardiorespiratory fitness and classification of risk of cardiovascular disease mortality. *Circulation* **2011**, *123*, 1377–1383. [CrossRef] [PubMed]
4. Ross, R.; Blair, S.N.; Arena, R.; Church, T.S.; Despres, J.P.; Franklin, B.A.; Haskell, W.L.; Kaminsky, L.A.; Levine, B.D.; Lavie, C.J.; et al. Importance of assessing cardiorespiratory fitness in clinical practice: A case for fitness as a clinical vital sign: A scientific statement from the american heart association. *Circulation* **2016**, *134*, e653–e699. [CrossRef]
5. Domone, S.; Mann, S.; Sandercock, G.; Wade, M.; Beedie, C. A method by which to assess the scalability of field-based fitness tests of cardiorespiratory fitness among schoolchildren. *Sports Med.* **2016**, *46*, 1819–1831. [CrossRef]
6. Malek, M.H.; Housh, T.J.; Berger, D.E.; Coburn, J.W.; Beck, T.W. A new nonexercise-based VO_{2max} equation for aerobically trained females. *Med. Sci. Sports Exerc.* **2004**, *36*, 1804–1810. [CrossRef]
7. Matsuo, T.; So, R.; Takahashi, M. Workers' physical activity data contribute to estimating maximal oxygen consumption: A questionnaire study to concurrently assess workers' sedentary behavior and cardiorespiratory fitness. *BMC Public Health* **2020**, *20*, 22. [CrossRef]
8. Nes, B.M.; Janszky, I.; Vatten, L.J.; Nilsen, T.I.; Aspenes, S.T.; Wisløff, U. Estimating VO_{2peak} from a nonexercise prediction model: The HUNT Study, Norway. *Med. Sci. Sports Exerc.* **2011**, *43*, 2024–2030. [CrossRef]
9. Wier, L.T.; Jackson, A.S.; Ayers, G.W.; Arenare, B. Nonexercise models for estimating VO_{2max} with waist girth, percent fat, or BMI. *Med. Sci. Sports Exerc.* **2006**, *38*, 555–561. [CrossRef]
10. DE Lannoy, L.; Ross, R. Nonexercise equations for determining change in cardiorespiratory fitness. *Med. Sci. Sports Exerc.* **2020**, *52*, 1525–1531. [CrossRef]
11. Peterman, J.E.; Whaley, M.H.; Harber, M.P.; Fleenor, B.S.; Imboden, M.T.; Myers, J.; Arena, R.; Kaminsky, L.A. Comparison of non-exercise cardiorespiratory fitness prediction equations in apparently healthy adults. *Eur. J. Prev. Cardiol.* **2021**, *28*, 142–148. [CrossRef] [PubMed]
12. Billinger, S.A.; VAN Swearingen, E.; Mcclain, M.; Lentz, A.A.; Good, M.B. Recumbent stepper submaximal exercise test to predict peak oxygen uptake. *Med. Sci Sports Exerc.* **2012**, *44*, 1539–1544. [CrossRef] [PubMed]
13. Teren, A.; Zachariae, S.; Beutner, F.; Ubrich, R.; Sandri, M.; Engel, C.; Loffler, M.; Gielen, S. Incremental value of Veterans Specific Activity Questionnaire and the YMCA-step test for the assessment of cardiorespiratory fitness in population-based studies. *Eur. J. Prev. Cardiol.* **2016**, *23*, 1221–1227. [CrossRef] [PubMed]
14. Matsuo, T.; So, R.; Takahashi, M. Estimating cardiorespiratory fitness from heart rates both during and after stepping exercise: A validated simple and safe procedure for step tests at worksites. *Eur. J. Appl. Physiol.* **2020**, *120*, 2445–2454. [CrossRef] [PubMed]
15. Webb, C.; Vehrs, P.R.; George, J.D.; Hager, R. Estimating VO_{2max} using a personalized step test. *Meas. Phys. Educ. Exerc. Sci.* **2014**, *18*, 184–197. [CrossRef]
16. Matsuo, T.; So, R.; Murai, F. Estimation methods to detect changes in cardiorespiratory fitness due to exercise training and subsequent detraining. *Eur. J. Appl. Physiol.* **2022**, 1–13. [CrossRef]
17. Buckley, J.P.; Sim, J.; Eston, R.G.; Hession, R.; Fox, R. 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*, 197–205. [CrossRef]

18. Bennett, H.; Parfitt, G.; Davison, K.; Eston, R. Validity of submaximal step tests to estimate maximal oxygen uptake in healthy adults. *Sports Med.* **2016**, *46*, 737–750. [CrossRef]

19. Cuenca-Garcia, M.; Marin-Jimenez, N.; Perez-Bey, A.; Sanchez-Oliva, D.; Camiletti-Moiron, D.; Alvarez-Gallardo, I.C.; Ortega, F.B.; Castro-Pinero, J. Reliability of Field-Based Fitness Tests in Adults: A Systematic Review. *Sports Med.* **2022**, *52*, 1961–1979. [CrossRef]

20. Nakamura, E.; Miyao, K.; Ozeki, T. Assessment of biological age by principal component analysis. *Mech. Ageing Dev.* **1988**, *46*, 1–18. [CrossRef]

21. Arena, R.; Myers, J.; Kaminsky, L.A. Revisiting age-predicted maximal heart rate: Can it be used as a valid measure of effort? *Am. Heart J.* **2016**, *173*, 49–56. [CrossRef] [PubMed]

22. Sarzynski, M.A.; Rankinen, T.; Earnest, C.P.; Leon, A.S.; Rao, D.C.; Skinner, J.S.; Bouchard, C. Measured maximal heart rates compared to commonly used age-based prediction equations in the Heritage Family Study. *Am. J. Hum. Biol.* **2013**, *25*, 695–701. [CrossRef] [PubMed]

23. Howley, E.T.; Bassett, D.R.J.; Welch, H.G. Criteria for maximal oxygen uptake: Review and commentary. *Med. Sci. Sports Exerc.* **1995**, *27*, 1292–1301. [CrossRef] [PubMed]

24. Riebe, D.; Ehrman, J.K.; Liguori, G.; Magal, M. Clinical exercise testing and interpretation. In *ACSM's Guidelines for Exercise Testing and Prescription*, 10th ed.; Wolters Kluwer: Philadelphia, PA, USA, 2017; pp. 111–142.

25. Tanaka, H.; Monahan, K.D.; Seals, D.R. Age-predicted maximal heart rate revisited. *J. Am. Coll. Cardiol.* **2001**, *37*, 153–156. [CrossRef]

26. Kodama, S.; Saito, K.; Tanaka, S.; Maki, M.; Yachi, Y.; Asumi, M.; Sugawara, A.; Totsuka, K.; Shimano, H.; Ohashi, Y.; et al. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: A meta-analysis. *JAMA* **2009**, *301*, 2024–2035. [CrossRef]

27. Peterman, J.E.; Harber, M.P.; Imboden, M.T.; Whaley, M.H.; Fleenor, B.S.; Myers, J.; Arena, R.; Finch, W.H.; Kaminsky, L.A. Accuracy of nonexercise prediction equations for assessing longitudinal changes to cardiorespiratory fitness in apparently healthy adults: BALL ST cohort. *J. Am. Heart Assoc.* **2020**, *9*, e015117. [CrossRef]

28. Chatterjee, S.; Chatterjee, P.; Mukherjee, P.S.; Bandyopadhyay, A. Validity of Queen's College step test for use with young Indian men. *Br. J. Sports Med.* **2004**, *38*, 289–291. [CrossRef]

29. Astrand, P.O.; Ryhming, I. A nomogram for calculation of aerobic capacity (physical fitness) from pulse rate during sub-maximal work. *J. Appl. Physiol.* **1954**, *7*, 218–221. [CrossRef]

30. Gallagher, J.R.; Brouha, L.V. A simple method of evaluating fitness in boys: The step test. *Yale. J. Biol. Med.* **1943**, *15*, 769–779.

31. Lee, O.; Lee, S.; Kang, M.; Mun, J.; Chung, J. Prediction of maximal oxygen consumption using the Young Men's Christian Association-step test in Korean adults. *Eur. J. Appl. Physiol.* **2019**, *119*, 1245–1252. [CrossRef]

32. Santo, A.S.; Golding, L.A. Predicting maximum oxygen uptake from a modified 3-min step test. *Res. Q. Exerc. Sport* **2003**, *74*, 110–115. [CrossRef] [PubMed]

33. Lohman, T.G. Skinfolds and body density and their relation to body fatness: A review. *Hum. Biol.* **1981**, *53*, 181–225. [PubMed]

34. Sinning, W.E.; Dolny, D.G.; Little, K.D.; Cunningham, L.N.; Racaniello, A.; Siconolfi, S.F.; Sholes, J.L. Validity of "generalized" equations for body composition analysis in male athletes. *Med. Sci. Sports Exerc.* **1985**, *17*, 124–130. [CrossRef] [PubMed]