ORIGINAL ARTICLE

Prediction of maximal oxygen consumption using the Young Men's Christian Association-step test in Korean adults

On Lee[1] · Sukho Lee[2] · Minsoo Kang[3] · Junbae Mun[4] · Jinwook Chung[5]

Abstract

Purpose To develop accurate and practical prediction models of maximal oxygen consumption (VO_2max) using the Young Men's Christian Association (YMCA)-step test in South Korean adults.

Methods In total, 568 adults (20–66 years) were included in this study. To develop and cross-validate prediction models of VO_2max, the total sample was divided into 80% training and 20% testing using a simple random sampling method. VO_2max was measured using the maximal-graded exercise treadmill test. Sex, age, 1-min recovery heart rate, body weight, and height were measured as potential predictors. Each test was conducted within a 2- to 3-day interval, ensuring sufficient rest. Preliminary prediction models were developed from training datasets, which were cross-validated using regression analyses and/or repeated-measures analysis of variance. The accuracy of prediction models was evaluated using R^2, standard error of estimate (SEE), and mean difference (MD) against a criterion-measured VO_2max.

Results The average age and VO_2max were 43.5 ± 12.9 years and 39.1 ± 7.5 ml/kg/min, respectively. For model development, three practical models with acceptable accuracy were developed ($R^2 = 0.56$–0.61; SEE $= 4.74$–5.01). For model cross-validation, significant relationships between the criterion-measured and predicted VO_2max were observed in all three models ($R^2 = 0.56$–0.61; SEE $= 4.62$–4.88). The difference between criterion-measured and predicted VO_2max was not significant in the models (MD $= -0.03$ to -0.14).

Conclusions The prediction models included 3–5 variables as significant predictors of VO_2max and had acceptable accuracy in a large sample of South Korean adults. The selected models provide a simple and practical method to estimate VO_2max using the YMCA-step test for South Korean adults.

Keywords VO_2max · Submaximal exercise tests · Prediction model · Cross-validation

Abbreviations

CRF	Cardiorespiratory fitness
CHD	Coronary heart disease
MD	Mean difference
RHR	Recovery heart rate
SEE	Standard error of estimate
YMCA	Young Men's Christian Association

Communicated by Guido Ferretti.

✉ Jinwook Chung
  cjw826@dongguk.edu

[1] Korea Institute of Sport Science, 109, 727 Hwarang-ro, Nowon-gu, Seoul 01794, South Korea

[2] Department of Counseling, Health, and Kinesiology, College of Education and Human Development, Texas A&M University-San Antonio, One University Way, STEM 142A, San Antonio, TX 78224, USA

[3] Health, Exercise Science, and Recreation Management, The University of Mississippi, P.O. Box 1848, 219 Turner Center, University, MS 38677-1848, USA

[4] Department of Physical Education, Korea Military Academy, Seoul, South Korea

[5] Sports Culture Department, Dongguk University, 101-492, 30, Pildong-ro 1 gil, Jung-gu, Seoul 04620, South Korea

Introduction

Cardiorespiratory fitness (CRF) is a health-related component of physical fitness that refers to the ability of circulatory and respiratory systems to supply oxygen to skeletal muscles during sustained physical activity (Physical Activity Guidelines Advisory Committee 2008). CRF exerts a significant influence on the incidence rate of cardiovascular disease and mortality (Gulati et al. 2005; Myers et al. 2002). Moreover, a

meta-analysis indicated an inverse relationship between CRF and coronary heart disease (CHD) or all-cause mortality in healthy participants (Kodama et al. 2009). Thus, accurately measuring CRF is important for diagnosis of health-related risk and evaluation of an intervention to improve health.

In general, measuring a participant's maximum oxygen uptake (VO2max) is considered a "gold standard" for assessing cardiorespiratory endurance. Methods for measuring maximum oxygen intake include the maximal exercise test, which analyzes expiratory gas collected from the participant while he or she performs maximum intensity exercise with a gradually increased load, and the submaximal exercise test, which estimates the maximum oxygen intake through heart rate during submaximal intensity (Santo and Golding 2003). Maximal exercise tests may provide accurate measurements of VO2max reflecting people's cardiorespiratory endurance, but they require full cooperation from the participant because the exercise must be performed until exhaustion, and they are also associated with a relatively high risk of accidents due to the nature of maximal effort exercise. Moreover, these tests require various equipments and continuous monitoring by an experienced assessor, and present difficulties in multiple participants.

For these reasons, many researchers have designed various types of tests that indirectly estimate maximum oxygen intake by using submaximal exercise (Margaria et al. 1965; Kasch et al. 1966; Åstrand 1976; Lakomy and Lakomy 1993; Sartor et al. 2013). Submaximal exercise tests can be divided into laboratory tests, such as treadmill or bicycle ergometer tests, and field tests, such as running, walking, and step tests outside of the lab setting. While submaximal exercise test methods produce less accurate results than maximal exercise tests, they have a lower risk of accidents, require less time and less expensive equipment, and allow assessors to measure multiple participants at the same time.

Among the several submaximal exercise tests, the step test is widely used because it is simple and easy to administer. Margaria step test consists of stepping up and down a 30- to 40-cm bench. They reported that the variability of the data obtained is within ± 7% with those directly determined (Margaria et al. 1965). Since then many step tests have been developed. In particular, the Young Men's Christian Association (YMCA) step test can be applied more broadly due to its relatively low step height (30.5 cm), compared to other step heights (41.3–50.8 cm). Although previous studies have developed equations that estimate maximum oxygen intake by using the YMCA step test (Santo and Golding 2003; Jacks et al. 2012; Beutner et al. 2015), there are limitations of such findings because these studies included a small number of participants with limited age range, and they did not cross-validate the equations. Moreover, some previous studies used recovery heart rate, sex, and age as explanatory variables (Beutner

et al. 2015), but it is necessary to consider physical characteristics, such as height, weight, and body composition. This is because the same step box height and testing speed may produce huge differences in relative load due to individual's physical characteristics such as height and weight (Shahnawaz 1978). Therefore, considering such characteristics would be appropriate to develop more accurate equations for predicting maximum oxygen intake. In addition, previous studies were carried out in Europe and US, but not in Asia.

Hence, the present study was conducted with a sufficient number of Korean adults to develop a practical and accurate estimation model for predicting maximum oxygen uptake using the YMCA step test.

Materials and methods

Study sample

We recruited 665 adult participants between 20 and 66 years of age from Seoul. We divided the participants based on their sex and further subdivided them into age groups of 5-year intervals, with 20 participants in each group. This study was approved by the Institutional Review Board at the authors' affiliated institution. Before beginning the study, we explained the planned measurements to the participants and received their written consent. To ensure accuracy of the study, we excluded people with cardiac disease, pulmonary disease, any anti-hypertensive medications record in past 6 months or orthopedic disease in the joints of lower limb as these could interfere with physical activity. Due to the exclusion criteria and invalid measures, 97 participants (i.e., heart disease = 8, medication = 18, invalid VO2max measure = 18, and invalid heart rate measure = 53) were excluded from analyses. This resulted in a final dataset consisting of 568 adults. Physical characteristics of participants are shown in Table 1.

Measurements

Physical examination and YMCA step test were performed on the first visit. VO2max was measured on the subsequent visit. All participants had 2–3 days of rest to minimize the influence of one test on the other. In order to minimize errors between measurements, we instructed the participants to refrain from excessive physical activity, smoking, and alcohol consumption on the day before their visit to the laboratory, and not to eat or drink beverages containing stimulants during or around the time of tests.

Table 1 Participants characteristics

	Total	Training dataset	Testing dataset
Sample size (*n*)	568	454	114
Age (years)	43.5 ± 12.9	43.3 ± 12.9	44.4 ± 12.8
Male (*n*)	294 (51.8%)	236 (52.0%)	58 (50.9%)
Height (m)	1.66 ± 0.1	1.67 ± 0.1	1.65 ± 0.1
Weight (kg)	64.9 ± 11.8	65.2 ± 11.8	63.8 ± 11.7
Body mass index (kg/m^2)	23.3 ± 3.0	23.4 ± 3.0	23.2 ± 3.0
Treadmill test			
VO_{2max} (ml/kg/min)	39.1 ± 7.5	39.1 ± 7.5	38.8 ± 7.4
RER	1.2 ± 0.1	1.2 ± 0.1	1.2 ± 0.1
Step test			
RHR (beat/min)	104.1 ± 15.9	104.2 ± 16.0	103.7 ± 15.5

Data are presented as means ± standard deviation

RER respiratory exchange ratios, *RHR* recovery heart rate

Demographic variables: age, sex, weight, height

During their first visits, participants were surveyed on their health-related habits and daily lifestyle, and a basic medical examination was conducted. Participants whose survey results did not meet the conditions required for the study were excluded. We measured the participants' height, bodyweight and calculated BMI.

VO_{2max}

VO_2max was measured through a maximal-graded exercise test using treadmills, and the detailed protocol is outlined below. Participants wore a wireless heart rate monitor (S610i; Polar, USA) and relaxed for at least 10 min before starting the test. While participants performed the exercise, their VO_2max was measured using an expired gas analyzer (K4b2; Cosmed, USA). In the maximal-graded exercise test, we used the Bruce protocol designed for adults (Bruce et al. 2004). Bruce protocol has shown to produce valid VO2max value irrespective of age, sex or underlying fitness levels (Hamlin et al. 2012). It is the most common protocol used during treadmill exercise stress testing and also used for fitness evaluation. Also, measurement time of Bruce protocol was relatively short (10:41 min by Buchfuhrer et al. 1983). So, we used Bruce protocol in present study because of large number of subjects (*n* = 586) in this study. The participant's condition was continuously monitored during the maximal-graded exercise test by perceived exertion (Borg RPE 10-scale) and heart rate. For accurate measurements of VO_2max, participants were considered exhausted from maximal exercise when they met two out of the following three criteria: oxygen uptake leveling-off despite the increasing exercise load, a change in respiratory ratio (CO_2/O_2) of at least 1.15, or heart rate approaching 95% of the predicted maximum heart rate (220 − age) (Howley et al. 1995). After the completion of tests, invalid cases in which reversal of steady state heart rate at successive phases was observed and cases in which the maximal values could not be measured due to withdrawal were excluded from analysis.

YMCA step test

Participants walked up and down for 3 min on a 30.5-cm-high step box at a speed of 96 steps per minute (Golding 2000). After the test, participants were immediately seated on chairs, and recovery heart rate was measured using the palpation method from the radial artery for 1 min. Metronomes were used to maintain a steady speed for 3 min. In participants who had difficulty maintaining a steady speed, the tester guided them to maintain an accurate pace. All participants completed the YMCA step test without withdrawal.

Statistical analyses

In order to develop and cross-validate a practical VO_2max prediction model, the total dataset (*n* = 568) was divided into 80% training (*n* = 454) and 20% testing (*n* = 114) using a simple random sampling method. Descriptive statistics of the main variables were calculated for the total, training, and testing datasets.

Prediction model development procedures

The outcome variable was VO_2max measured by the maximal-graded exercise treadmill test, following the protocol established by Bruce et al. (2004). Sex, age, 1-min recovery heart rate (RHR), body weight and height were chosen as potential predictor variables based on the results of previous studies (Shahnawaz 1978; Francis and Feinstein 1991; Francis and Brasher 1992; Santo and Golding 2003; Jacks et al. 2012; Beutner et al. 2015) and ease of measurement. Pearson correlation coefficients were computed to examine the relationships between the outcome variable and potential predictor variables.

Multiple regression analysis was used to develop a VO_2max prediction model using the training dataset. A backwards-selection regression approach was used to identify the most parsimonious prediction VO_2max model. The initial model included all candidate predictor variables such as dummy coded sex (i.e., male = 1, female = 0), age, RHR, body weight, and height. At each step, the least significant variable was removed from the model. The least significant variable was determined by comparing the *t* value for each variable. The smallest *t* value was considered the least significant. To determine the accuracy of each model, R^2 was

calculated. Our goal was to create a simple prediction model that maintained R^2 similar to the original model. Therefore, for each successive model, the least significant variable was removed and the consequent change in R^2 was determined. The best model was determined with the fewest variables and no significant change in R^2 (i.e., R^2 change < 0.01).

In order to develop a valid regression model, assumptions for regression analysis were verified. Scatter plots examined the linear relationships between the outcome and predictor variables. Kolmogorov–Smirnov test evaluated the normality of residuals (i.e., criterion-measured VO_2max − predicted VO_2max by the VO_2max prediction models). The homoscedasticity and independence of the residuals were evaluated graphically using scatter plots between the predicted VO_2max and residuals. Lastly, multicollinearity was checked using a variation inflation factor (VIF).

Cross-validation procedures

For cross-validation purposes, the VO_2max prediction models were compared to the actual VO_2max using the testing dataset. The developed VO_2max prediction models were also compared to a previous VO_2max prediction model, which comprised sex, age, and RHR, developed by Beutner et al. (2015). Repeated-measures (RM) ANOVA and simple regression analyses were conducted to evaluate the prediction models. The RM ANOVA examined if the averages of predicted VO_2max from the developed models and the Beutner et al. (2015) model were significantly different when compared to the criterion-measured VO_2max. A simple regression analysis examined the accuracy of predicted VO_2max by the two models against criterion-measured VO_2max. Standard error of estimate (SEE) was calculated for evaluating the accuracy of VO_2max prediction models. SEE was calculated as $S_Y\sqrt{1 - R^2 YY'}$, where S_Y is standard deviation of the criterion-measured VO_2max, Y is the criterion-measured VO_2max, and Y' is predicted VO_2max; a smaller SEE is preferred in the comparison.

All statistical analyses were performed using SPSS ver. 20 (IBM, USA). Statistical significance was set at an alpha level of 0.05.

Results

Participant characteristics are presented in Table 1. Average age was 43.5 ± 12.9 years (range 20–66 years). Approximately 52% of the participants were men. The average VO_2max was 39.1 ± 7.5 ml/kg/min. Overall, no significant differences were identified in participant characteristics and measured variables between training and testing datasets (see Table 1, $p < .05$).

Development of a VO_2max prediction model

In the training dataset, all predictor variables had a linear relationship with criterion-measured VO_2max, except for RHR. A slight curvilinear relationship was identified between RHR and criterion-measured VO_2max. Thus, the inverse of RHR (i.e., 1/RHR), which shows a linear relationship with VO_2max, was considered as another potential predictor variable.

All potential predictor variables were significantly correlated with the criterion-measured VO_2max ($r > 0.20$ or $r < -0.37$, $p < .001$, see Table 2). The highest correlation was found between the criterion-measured VO_2max and sex ($r = 0.56$), and the lowest between the criterion-measured VO_2max and weight ($r = 0.20$). Because the correlation between inversed RHR and the criterion-measured VO_2max ($r = 0.46$) was higher than that of RHR ($r = -0.44$), the inversed RHR was included in the VO_2max prediction model instead of RHR.

Model 1 included all potential prediction variables (i.e., sex, age, inversed RHR, weight, and height). All variables in the model were significant predictors of VO_2max, $F(5, 448) = 139.31$, MSE $= 22.49$, $p < .001$ (see Table 3). Model 1 was the most accurate with $R^2 = 0.61$, SEE $= 4.74$. Although

Table 2 Correlations between VO_{2max} and predictor variables in training dataset ($n = 454$)

Variables	1	2	3	4	5	6	7
1. VO_{2max}	1.000						
2. Sex	0.558[a]	1.000					
3. Age	− 0.366[a]	− 0.016	1.000				
4. RHR	− 0.444[a]	− 0.292[a]	− 0.094[b]	1.000			
5. Inversed RHR	0.463[a]	0.301[a]	0.096[b]	− 0.978[a]	1.000		
6. Weight	0.200[a]	0.683[a]	− 0.029[a]	− 0.099[b]	0.106[b]	1.000	
7. Height	0.485[a]	0.723[a]	− 0.311[a]	− 0.176[a]	0.176[a]	0.699[a]	1.000

RHR recovery heart rate, *BF%* body fat percentage

[a]Correlation coefficient is significant ($p < 0.001$)

[b]Correlation coefficient is significant ($p < 0.05$)

Table 3 Summary of developed VO_{2max} prediction models in training dataset ($n = 454$)

	Model 1	Model 2	Model 3
Intercept	25.22[b]	40.31[a]	28.80[a]
Sex	8.97[a]	9.77[a]	6.48[a]
Age (years)	− 0.21[a]	− 0.23[a]	− 0.24[a]
Inversed RHR (beat/min)	1592.32[a]	1611.60[a]	1747.45[a]
Weight (kg)	− 0.21[a]	− 0.19[a]	
Height (m)	9.99[b]		
R	0.78	0.78	0.75
R^2	0.61	0.60	0.56
R^2 change		− 0.004	− 0.044
p values for R^2 change		0.028	<0.001
SEE	4.74	4.76	5.01

RHR recovery heart rate, *SEE* standard error of estimate

[a]Coefficient is significant ($p < 0.001$)

[b]Coefficient is significant ($p < 0.05$)

all variables were significant predictors, height and weight were removed from model 1 in order of their respective significance to identify the model with most parsimony and good accuracy. Height ($t = 2.21$, $p = .028$) was removed first (model 2) and weight ($t = − 7.05$, $p < .001$) removed second (model 3). Model 2 was comparably accurate compared to model 1 (R^2 change $= − 0.004$) while model 3 was not when compared to model 2 (R^2 change $= − 0.044$). Model 2, which included sex, age, inversed RHR and weight variables, was chosen as the best model due to its accuracy and simplicity, $F(4, 449) = 171.45$, MSE $= 22.68$, $p < .001$, $R^2 = 0.60$, SEE $= 4.76$.

For the developed VO_2max prediction models, all regression assumptions were satisfied. The Kolmogorov–Smirnov test indicated that residual was normally distributed ($p > 0.05$). No pattern and shape were found in the scatter plot between predicted VO_2max and residuals. There was no multicollinearity among the predictor variables (VIF ranged from 1.01 to 3.20; desirable VIF ≤ 10) (O'Brien 2007).

Cross-validation of the chosen VO₂max prediction models

The averages of criterion-measured VO_2max and predicted VO_2max from varying models were significantly different, $F(1.94, 219.26) = 40.77$, MSE $= 25.22$, G–G $p < .001$, $\eta^2 = 0.27$ (see Table 4). The differences between criterion-measured and predicted VO_2max among the three models were not significant (i.e., 95% CI includes zero). However, there was a significant difference in the predicted VO_2max when using the Beutner et al. (2015) model (model 4: mean difference $= − 4.69$, 95% CI $= − 6.66$ to $− 2.72$).

Table 4 The results of repeated-measures ANOVA in testing dataset ($n = 114$)

	Mean	SD of mean	MD	95% CI[f] Lower	95% CI[f] Upper
Criterion (measured VO2max)	38.95	7.34			
Model 1[a]	38.98	6.09	− 0.03	− 1.27	1.21
Model 2[b]	39.09	6.06	− 0.14	− 1.37	1.10
Model 3[c]	38.83	5.75	− 0.12	− 1.18	1.42
Model 4[d]	43.64	5.65	− 4.69[e]	− 6.66	− 2.72

CI confidence interval for mean difference, *MD* mean difference (criterion-measured VO_{2max}-predicted VO_{2max}), *SD* standard deviation

[a]Model 1 includes sex, age, inversed recovery heart rate, weight, and height

[b]Model 2 includes sex, age, inversed recovery heart rate, and weight

[c]Model 3 includes sex, age, and inversed recovery heart rate

[d]Model 4 is Beutner et al. 8 model, $78.2 – 4.2 \times$ sex $− 0.38 \times$ age $− 0.15 \times$ recovery heart rate

[e]Mean difference is significantly different from zero ($p < .05$)

[f]Sidak adjustment is applied for multiple comparisons

The results of the simple regression analyses indicated significant relationships between criterion-measured and predicted VO_2max by the three developed models, $R^2 = 0.56$–0.61, all p values < 0.001, SEE $= 4.62$ to 4.88 (see Table 5). For the three models, the slopes of predicted VO_2max were not significantly different from one (i.e., 95% CI includes one), and the intercepts were not significantly different from zero (i.e., 95% CI includes zero), which indicate no proportional bias of predicted VO_2max against the criterion-measured VO_2max. Model 4 overestimated VO_2max in participants with lower criterion-measured VO_2max while it under estimated VO_2max in participants with higher criterion-measured VO_2max (slope $= 0.49$, 95% CI $= 0.27$–0.72; intercept $= 17.44$, 95% CI $= 7.54$–27.34). The R^2 and SEE of model 4 were 0.14 and 6.82, respectively.

Discussion

The aim of this study was to develop a practical and accurate estimation model of the YMCA step test among Korean population. This study included a sufficient number of Korean adults and performed validation tests after developing the equation. Also, to our knowledge, this is the first study on prediction of VO_2max using YMCA step test in the Korean population.

The best model from this study includes sex, age, inversed RHR, and weight variables. These predictors are consistent with those used in previous studies (Santo and Golding 2003; Beutner et al. 2015) that developed VO_2max

Table 5 The results of simple regression analyses in testing dataset ($n = 114$)[a]

Model	B	SE	β	95% CI		R	R^2	SEE
				Lower	Upper			
Model 1[b]								
Intercept	2.45	2.82		− 3.15	8.04	0.78	0.60	4.63
Slope	0.94	0.07	0.78	0.80	1.08			
Model 2[c]								
Intercept	2.13	2.84		− 3.49	7.74	0.78	0.61	4.62
Slope	0.94	0.07	0.78	0.80	1.08			
Model 3[d]								
Intercept	1.81	3.13		− 4.40	8.01	0.75	0.56	4.88
Slope	0.95	0.08	0.75	0.80	1.12			
Model 4[e]								
Intercept	17.44[f]	5.00		7.54	27.34	0.38	0.14	6.82
Slope	0.49[g]	0.11	0.38	0.27	0.72			

SE standard error, *SEE* standard error of estimate

[a]Criterion-measured VO_{2max} is regressed by each predicted VO_2max

[b]Model 1 includes sex, age, inversed recovery heart rate, weight, and height

[c]Model 2 includes sex, age, inversed recovery heart rate, and weight

[d]Model 3 includes sex, age, and inversed recovery heart rate

[e]Model 4 is Beutner et al. 8 model, $78.2 - 4.2 \times sex - 0.38 \times age - 0.15 \times 1$ min recovery heart rate

[f]Coefficients are significantly different from zero ($p < .05$)

[g]Coefficients are significantly different from one ($p < .05$)

prediction models using the YMCA step test. These previous studies reported that sex, age, and 1-min resting heart rate from the step test are significant predictors of VO_2max. In addition to these three predictors, weight was identified as a significant predictor in the current study. A similar finding was reported in a previous study (Shephard et al. 1991) that developed a VO_2max prediction model for Canadian Home Fitness Test, which is a type of step test. In that study, weight was one of the significant predictors in the VO_2max prediction model for the step test. However, there are a few limitations in comparing the results of this study with those of previous ones because very few studies have developed the VO_2max prediction model using step tests. Moreover, some of the previous studies (McARDLE et al. 1972; Santo and Golding 2003) only used RHR to estimate VO_2max.

As compared to the previous VO_2max prediction model (Beutner et al. 2015) ($R^2 = 0.74$; SEE = 5.50), which included sex, age, and RHR from YMCA step test, the prediction model from the current study provided better SEE and lower R^2 in the VO_2max prediction for Korean adults ($R^2 = 0.60$; SEE = 4.63). In terms of cross-validation, the prediction model from this study provided better accuracy ($R = 0.78$; SEE = 4.62; mean difference = − 0.14 ml/kg/min) compared to other prediction models based on the maximal exercise tests on field (Noh et al. 2014) (mean difference = 11.63 ml/kg/min) and laboratory (Im and Jeon 2013a, b) ($R = 0.62$–0.79; SEE = 4.54–5.70; mean difference = − 1.61 to 1.74 ml/kg/min) settings. Although our prediction

model was developed based on the submaximal exercise test, its accuracy was either comparable or better than the previously developed VO_2max prediction models using maximal exercise tests. Therefore, the prediction model from this study can be useful to estimate VO_2max in Korean adults.

In addition to model 2, which is the best model from this study, model 1 and model 3 may be used as alternative methods to estimate VO_2max in Korean adults. Although model 1 requires an additional predictor (height), it seems to be comparable to model 2 in VO_2max prediction. The most accurate model in the development stage was model 1, and height and weight variables were typically measured together in the field and research settings. If height data are available without much effort, we can use model 1 to estimate VO_2max. Model 3, which included sex, age, and inversed RHR, had relatively less accuracy, but it was the most parsimonious model as compared to others. The difference in the accuracy indices ($\Delta R^2 = 0.05$; ΔSEE = 0.26) and mean difference (0.26 ml/kg/min) was small between model 3 and 2, and the accuracy of model 3 was much higher than the previous model developed by Beutner et al. (2015) ($R^2 = 0.14$; SEE = 6.82). The accuracy of models 1 and 3 were moderately good when compared to the previously developed VO_2max prediction models (Im and Jeon 2013a, b; Noh et al. 2014) for Korean adults.

The model developed by Beutner et al. (2015) showed good accuracy in a cross-validation study for German adults ($R^2 = 0.74$; SEE = 5.50), but had poor accuracy for Korean

adults in this study ($R^2 = 0.14$; SEE $= 6.82$). This discrepancy in accuracy of the model developed by Beutner et al. (2015) may be due to the difference in BMI between the samples of two studies (25.20 ± 4.3 and 23.35 ± 3.03 kg/m^2 for Beutner et al. (2015) study and this study, respectively). BMI could cause changes in the heart rate during the same step test (Ko and Seo 2007). In other words, those with high BMI may have higher changes in heart rate during the same intensity exercise test than those with low BMI (Ko and Seo 2007). The difference in BMI between the two studies may explain the discrepancy in accuracy of Beutner et al. (2015). prediction model. Another possible reason for the difference in accuracy is a different body size between the populations. Leg length and standing height could influence the physiological response in the same step test (Shahnawaz 1978; Culpepper and Francis 1987; Beutner et al. 2015). For instance, because Asian adults have relatively short leg length and standing height when compared to European adults, (Rush et al. 2007, 2009; Stanfield et al. 2012) the physiological response (i.e., heart rate or reserved heart rate) in the same exercise test (i.e., YMCA step test) could be different between the two populations. It may imply that validation studies should be performed before the application of a VO$_2$max prediction model based on step tests for different populations who have different physical characteristics.

Mitochondrial muscle oxidative capacity and oxygen delivery capabilities, as determined by submaximal oxygen pulse, account for most if not all of the ethnic differences in VO$_2$max (Roy et al. 2006). Although there are studies reporting differences in the maximum oxygen uptake by ethnic (Lai et al. 2012; Mcmurray et al. 2002), research is still lacking to identify ethnic differences. Differences in aerobic capacity according to sex are reported to be due to differences in the ratio of muscle to body weight and hemoglobin levels (Perez-Gomez et al. 2008). Assuming that these physical conditions are equal, there remains considerable differences in the cardiovascular response to exercise according to sex, the reasons for which remain less clear (Charkoudian and Joyner 2004).

There are some limitations to this study. First, old PO Åstrand's maximal heart rate prediction formula was used as a criterion for verifying the measured maximal oxygen uptake in present study. This formula ($220 -$ age) is believed to underestimate maximal heart rate as the age increases. So new formula ($208 - 0.7 *$ age) were reported to be more accurate and recommended people over age 40 (Tanaka et al. 2001). However, we were not able find significant differences in the results when we replaced with new formula in present study. Even though lactate can provide important information as one of the subsidiary criteria (> 8–10 mM) to measure VO$_2$max during the exercise (Åstrand et al. 2003; Ferretti 2014). We were not able to measure lactate concentration at the end of exercise in present study.

Participants of this study were adults aged 20–66 years; therefore, readers need to be cautious about using the equations for other populations such as children and the elderly. In addition, our findings may not be applicable to people who are psychologically sensitive or who take medications that affect heart rate because the YMCA step test is based on heart rate. Further research is required to develop more accurate estimation models by adding a measurable variable such as percent body fat and muscle mass. There is also a need for studies on more diverse populations.

Conclusion

We developed and cross-validated the three prediction models to estimate VO$_2$max from a large sample of Korean adults. The three models are easy and practical to use in the field and laboratory settings. It showed comparable accuracy to the previously developed models for estimating VO$_2$max. These models could be used in large-scale epidemiologic research because they can easily measure many individuals with lower costs and time (i.e., less than 5 min) than other field and laboratory methods. In addition, these models can be embedded in a smart phone health-related app so that individuals with no special knowledge can easily estimate their VO$_2$max at home, school, as well as health-related organizations and facilities. Further research will be required to develop a prediction model of VO$_2$max for Korean children, adolescents, and elderly, as well as models that are more accurate for Korean populations.

References

Åstrand PO (1976) Quantification of exercise capability and evaluation of physical capacity in man. Progress Cardiovasc Disease 19(1):51–67

Åstrand PO, Rodahl K, Dahl HA, Strømme SB (2003) Textbook of work physiology: physiological bases of exercise. Hum Kinet

Beutner F, Ubrich R, Zachariae S, Engel C, Sandri M, Teren A, Gielen S (2015) Validation of a brief step-test protocol for estimation of peak oxygen uptake. Eur J Prev Cardiol 22(4):503–512

Bruce RA, Blackmon JR, Jones JW, Strait G (2004) Exercising testing in adult normal subjects and cardiac patients. Ann Noninvasive Electrocardiol 9(3):291–303

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

Charkoudian N, Joyner MJ (2004) Physiologic considerations for exercise performance in women. Clin Chest Med 25(2):247–255

Culpepper MI, Francis KT (1987) An anatomical model to determine step height in step testing for estimating aerobic capacity. J Theor Biol 129(1):1–8

Ferretti G (2014) Maximal oxygen consumption in healthy humans: theories and facts. Eur J Appl Physiol 114(10):2007–2036

Francis K, Brasher J (1992) A height-adjusted step test for predicting maximal oxygen consumption in males. J Sports Med Phys Fitness 32(3):282–287

Francis K, Feinstein R (1991) A simple height-specific and rate-specific step test for children. South Med J 84(2):169–174

Golding LA (2000) YMCA fitness testing and assessment manual. Human Kinetics, Champaign

Gulati M, Black HR, Shaw LJ, Arnsdorf MF, Merz CNB, Lauer MS et al (2005) The prognostic value of a nomogram for exercise capacity in women. N Engl J Med 353(5):468–475

Hamlin M, Draper N, Blackwell G, Shearman J, Kimber N (2012) Determination of maximal oxygen uptake using the bruce or a novel athlete-led protocol in a mixed population. J Hum Kinet 31:97–104

Howley ET, Bassett DR, Welch HG (1995) Criteria for maximal oxygen uptake: review and commentary. Med Sci Sports Exerc 27(9):1292–1301

Im JH, Jeon YJ (2013a) Prediction and validity of VO2max in Korean men. Korean J Phys Educ 52:461–470 (**In Korean**)

Im JH, Jeon YJ (2013b) Prediction and validity of VO2max with metabolic variables in sub-maximal exercise. Korean J Phys Educ 52:431–440 (**In Korean**)

Jacks DE, Topp R, Moore JB (2012) Prediction of VO2 peak using a sub-maximal bench step test in children (revised). Clin Kinesiol 66(3):74–82

Kasch FW, Phillips WH, Ross WD, Carter JE, Boyer JL (1966) A comparison of maximal oxygen uptake by treadmill and step-test procedures. J Appl Physiol 21(4):1387–1388

Ko YW, Seo CJ (2007) Prediction and establishment of optimal exercise intensity in obese male high school students according to bench step heights for development of cardiorespiratory endurance. J Korean Sport Res 18:391–402 (**In Korean**)

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: a meta-analysis. JAMA 301(19):2024–2035

Lai N, Tolentino-Silva F, Nasca MM, Silva MA, Gladden LB, Cabrera ME (2012) Exercise intensity and oxygen uptake kinetics in African-American and Caucasian women. Eur J Appl Physiol 112(3):973–982

Lakomy HKA, Lakomy J (1993) Estimation of maximum oxygen uptake from submaximal exercise on a Concept II rowing ergometer. J Sports Sci 11(3):227–232

Margaria R, Aghemo P, Rovelli E (1965) Indirect determination of maximal O_2 consumption in man. J Appl Physiol 20(5):1070–1073

Mcardle WD, Katch FI, Pechar GS, Jacobson LONI, Ruck S (1972) Reliability and interrelationships between maximal oxygen intake, physical work capacity and step-test scores in college women. Med Sci Sports 4(4):182–186

Mcmurray RG, Harrell JS, Bradley CB, Deng S, Bangdiwala SI (2002) Predicted maximal aerobic power in youth is related to age, gender, and ethnicity. Med Sci Sports Exerc 34(1):145–151

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(11):793–801

Noh WH, Jin SM, Kim WJ (2014) Research on the relationship between evaluation of cardiorespiratory function and shuttle run test. Tech Educ 32:15–26 (**In Korean**)

O'Brien RM (2007) A caution regarding rules of thumb for variance inflation factors. Qual Quant 41(5):673–690

Perez-Gomez J, Rodriguez GV, Ara I, Olmedillas H, Chavarren J, González-Henriquez JJ et al (2008) Role of muscle mass on sprint performance: gender differences? Eur J Appl Physiol 102(6):685–694

Physical Activity Guidelines Advisory Committee (2008) Physical activity guidelines advisory committee report, 2008. US Department of Health and Human Services, Washington, p G2

Roy JL, Hunter GR, Fernandez JR, McCarthy JP, Larson-Meyer DE, Blaudeau TE et al (2006) Cardiovascular factors explain genetic background differences in VO2max. Am J Hum Biol 18(4):454–460

Rush EC, Goedecke JH, Jennings C, Micklesfield L, Dugas L, Lambert EV, Plank LD (2007) BMI, fat and muscle differences in urban women of five ethnicities from two countries. Int J Obes 31(8):1232

Rush EC, Freitas I, Plank LD (2009) Body size, body composition and fat distribution: comparative analysis of European, Maori, Pacific Island and Asian Indian adults. Br J Nutr 102(4):632–641

Santo AS, Golding LA (2003) Predicting maximum oxygen uptake from a modified 3-minute step test. Res Q Exerc Sport 74(1):110–115

Sartor F, Vernillo G, De Morree HM, Bonomi AG, La Torre A, Kubis HP, Veicsteinas A (2013) Estimation of maximal oxygen uptake via submaximal exercise testing in sports, clinical, and home settings. Sports Med 43(9):865–873

Shahnawaz H (1978) Influence of limb length on a stepping exercise. J Appl Physiol 44(3):346–349

Shephard RJ, Thomas S, Weiler I (1991) The Canadian home fitness test. Sports Med 11(6):358–366

Stanfield KM, Wells JC, Fewtrell MS, Frost C, Leon DA (2012) Differences in body composition between infants of South Asian and European ancestry: the London Mother and Baby Study. Int J Epidemiol 41(5):1409–1418

Tanaka H, Monahan KD, Seals DR (2001) Age-predicted maximal heart rate revisited. J Am Coll Cardiol 37(1):153–156