ASSESSING PEAK AEROBIC CAPACITY IN DUTCH LAW ENFORCEMENT OFFICERS

HARRIET WITTINK[1], **TIM TAKKEN**[2], **JANKE DE GROOT**[1], **MICHIEL RENEMAN**[3], **ROELOF PETERS**[1], **and LUC VANHEES**[1,4]

[1] Utrecht University of Applied Sciences, Utrecht, The Netherlands
Faculty of Health Care, Lifestyle and Health Research Group

[2] University Medical Center Utrecht, Utrecht, The Netherlands
Child Development and Exercise Center, Wilhelmina Children's Hospital

[3] University of Groningen, The Netherlands
Department of Rehabilitation Medicine, University Medical Center Groningen, Center for Rehabilitation

[4] Katholieke Universiteit Leuven, Leuven, Belgium
Department of Rehabilitation Sciences, Cardiovascular Rehabilitation Research Center

Abstract

Objectives: To cross-validate the existing peak rate of oxygen consumption (VO_{2peak}) prediction equations in Dutch law enforcement officers and to determine whether these prediction equations can be used to predict VO_{2peak} for groups and in a single individual. A further objective was to report normative absolute and relative VO_{2peak} values of a sample of law enforcement officers in the Netherlands. **Material and Methods:** The peak rate of oxygen consumption ($ml \times kg^{-1} \times min^{-1}$) was measured using a maximal incremental bicycle test in 1530 subjects, including 1068 male and 461 female police officers. Validity of the prediction equations for groups was assessed by comparing predicted VO_{2peak} with measured VO_{2peak} using paired t-tests. For individual differences limits of agreement (LoA) were calculated. Equations were considered valid for individuals when the difference between measured and predicted VO_{2peak} did not exceed ±1 metabolic equivalent (MET) in 95% of individuals. **Results:** None of the equations met the validity criterion of 95% of individuals having ±1 MET difference or less than the measured value. Limits of agreement (LoAs) were large in all predictions. At the individual level, none of the equations were valid predictors of VO_{2peak} ($ml \times kg^{-1} \times min^{-1}$). Normative values for Dutch law enforcement officers were presented. **Conclusions:** Substantial differences between measured and predicted VO_{2peak} ($ml \times kg^{-1} \times min^{-1}$) were found. Most tested equations were invalid predictors of VO_{2peak} at group level and all were invalid at individual levels.

Key words:
Aerobic capacity, VO_{2peak}, Prediction equations, Maximal exercise test, Bicycle protocol, Law enforcement officers

INTRODUCTION

Physical fitness is essential for highly demanding and risky occupations, especially when public safety is involved. Physical fitness refers to the human ability to exert physical work. It is a multivariate concept that not only concerns aerobic capacity, but also anaerobic capacity, muscular strength, endurance, flexibility and coordination [1]. Maximal aerobic capacity (VO_{2max}) is accepted as the criterion measure of cardiorespiratory fitness [2,3]. Expressed either as an absolute rate in l of oxygen per min ($l \times min^{-1}$)

or as a relative rate in ml of oxygen per kg of bodyweight per min ($ml \times kg^{-1} \times min^{-1}$), it reflects the ability to perform external work and is determined by the performance of the cardiac muscle (oxygen (O_2) delivery) and efficiency of the muscular system in extracting oxygen from blood for use in generating energy. Progressive ramp type tests are typically continued to the limit of subject tolerance, a "maximally achieved" oxygen (O_2) uptake or "peak" oxygen consumption (VO_2) is also standardly determined. This is commonly taken to be equivalent to the VO_{2max} in subjects who give good effort [4].

Peak rate of oxygen consumption (VO_{2peak}) directly affects the amount and intensity of physical activity a healthy person is able to perform. A police officer is often placed in situations that make great demands on his/her physical capacity. In these instances, physical fitness is often the factor that spells the difference between success and failure – even life and death [5]. Law enforcement officers who remain physically fit prove more readily able to cope with the day-to-day stress of the job and are better prepared to handle critical incidents [6]. If the work demand exceeds the worker's capacity for sustained physical work, then the development of fatigue is inevitable.

A physiological limit of 30–40% of VO_{2peak} is thought to be acceptable for an 8 h working day [7] although 50% of a worker's VO_{2peak} [8] has also been recommended. Workers with higher levels of VO_{2peak} will thus have a higher capacity for sustained work, or will experience their workload as less fatiguing than workers who have lower levels of VO_{2peak} and who do the same work. Workers who perform work above their capacity may become unduly fatigued or have insufficient time between their working days to sufficiently recover. For instance, the workload in a number of law enforcement officers during mountain bike patrols was shown to exceed the threshold level for physiological stress demands in professional male cyclists, although these officers possessed much lower average VO_{2peak} values [9]. Although an association could not

be investigated, there was a high sick leave level and number of bikers resigning from the team.

Periodic assessment of aerobic capacity may help to identify and manage certain health conditions and risks at an early stage, that otherwise may progress to impact fitness for duty. In the laboratory, VO_{2peak} is assessed through maximal incremental exercise testing, obtaining direct measures of VO_{2peak} through respiratory gas analysis. With the nose occluded or wearing a mask, the subject breathes through a low resistance valve while pulmonary ventilation and O_2 and carbon dioxide (CO_2) concentrations are measured breath by breath in expired air samples. Because of the costs associated with equipment, space and personnel needed to carry out these tests, the direct measurement of VO_{2peak} is usually reserved for research or specialized clinical settings [2].

In field settings, as well as in occupational medicine, however, the sub-maximal tests are often used [10–16], as they do not require the expertise and costs associated with maximal exercise testing. Results of these submaximal tests are then used to predict VO_{2peak} using regression equations (e.g., Astrand test [7]). Another approach is to estimate VO_{2peak} from the maximal work rate that a subject accomplished during a graded exercise test to exhaustion by using a regression equation [17].

Although there is a number of VO_{2peak} regression equations available for use, there is an ongoing need to improve the predictive validity of these equations. A recent paper on comparative analysis of 3 prediction equations for estimating VO_{2max} using the 1 mile run test in male police officers, showed statistically significant differences between the equations. The authors recommend evaluating subjects using more than 1 equation [18]. Frequently cited equations derived from using cycle ergometry include those by Jones et al. [19], Hansen et al. [20], Wasserman et al. [21], Storer et al. [22] and the American College of Sports Medicine (ACSM) [2]. Except for the ACSM equation, these equations are also reported in several international

guidelines for maximal cardiopulmonary exercise testing [23–25] as giving normal reference values. These equations are based on general populations [2,19–22], as well as university students [19] and shipyard workers [21].

A previous study [26] that cross-validated a number of prediction equations on samples of aerobically trained males and females (93 males and 49 females) found that almost all currently available prediction equations underestimated measured VO$_{2peak}$ (ml×kg^{-1}×min^{-1}). The equations better predicted VO$_{2peak}$ in those subjects with the lowest levels of aerobic fitness. The authors suggested that the equations might be population-specific. It is therefore imperative to cross-validate these equations on various populations, including those that reflect a normal work force.

In contrast to firemen, whose aerobic capacity has been examined extensively, not much is known about aerobic capacity in law enforcement officers. Quantifying the energy demands of law enforcement is difficult as much of the work consists of sedentary tasks [27], with occasional near – or maximal capacity work [28]. To the best of our knowledge, normative data on male and female law enforcement officers has not been reported before.

We therefore have set out:

1. To cross-validate well-known VO$_{2peak}$ prediction equations on a sample of workers in the Utrecht Police force to determine whether these prediction equations can be used to accurately predict VO$_{2peak}$ in a single individual.
2. To report normative values on a sample of law enforcement officers in the Netherlands.

We hypothesized that prediction equations could predict VO$_{2peak}$ accurately on the group level, but did not expect that they would accurately predict VO$_{2peak}$ in a single individual.

MATERIAL AND METHODS
Population
This study is a part of the Utrecht Police Lifestyle Intervention Fitness and Training (UP-LIFT) study, a voluntary fitness and lifestyle test for police employees in Utrecht, the Netherlands. The population comprised 1786 volunteers (1211 men, 574 women), aged 18–62 years, who visited our research department between December 2004 and November 2009. All 1786 participants provided written informed consent and approval for the study was obtained from the Ethical Committee of the Utrecht University Medical Centre. The followed procedures were in accordance with the Helsinki Declaration of 1975, as revised in 2008 [29]. The study protocol is described in detail elsewhere [30].

Assessment of VO$_{2peak}$
All testing was conducted according to the American Thoracic Society (ATS) guidelines [23]. Subjects were tested on a bicycle ergometer (Jaeger ER800®, Würtzburg, Germany), in a laboratory with stabilized room temperature at 18–22°C. Seat height was adjusted so that a subject's legs were near full extension during each pedal revolution. The initial work rate of 20 W was increased every minute by 20 W. Patients were encouraged to give a maximal effort prior to the test, while during the test no further motivational interventions were used.

Test end point was volitional exhaustion or termination by the tester for medical reasons. Subjects were asked to maintain a cycling cadence between 60–70 revolutions/min. Peak workload (W$_{peak}$) was determined as the workload at the last complete 30 s exercise stage. During the test, a 12-lead electrocardiogram (ECG) and respiratory data through breath-by-breath analysis (Oxyxon Pro®, Jaeger, Care Fusion, Houten, The Netherlands) were continuously measured. Heart rate was determined from the ECG.

The gas analyzers and the flow meter were calibrated before each test according to the manufacturer's instructions. The gas analyzers were a paramagnetic O$_2$ analyzer and an infrared CO$_2$ analyzer. Minute ventilation (V$_E$) was measured by means of a turbine flow meter. Oxygen consumption (VO$_2$) and carbon dioxide production (VCO$_2$) were determined from the continuous measurement of oxygen

and carbon dioxide concentration in the inspired and expired air. Respiratory data was calculated using a moving average of 8 breaths to reduce variability and the disturbing effect of breath-by-breath noise. Peak rate of oxygen consumption (VO_{2peak}) was defined as the highest 15-s average of VO_2 obtained at the end of the test and was expressed in $ml\times min^{-1}$ and $ml\times kg^{-1}\times min^{-1}$. The respiratory gas exchange ratio (RER) was calculated as VCO_2/VO_2, the ventilatory equivalent for oxygen (EqO_2) as V_E/VO_2 and the ventilatory equivalent for carbon dioxide ($EqCO_2$) as V_E/CO_2. Subjects' data was included in this analysis if they stopped because of volitional exhaustion and if they met the RER > 1.00 criterion.

Estimated VO_{2peak}

We estimated VO_{2peak} ($ml\times kg^{-1}\times min^{-1}$) for our sample by using 3 non-gender specific (ACSM [2], Jones [19] and Wasserman [21]) and 4 gender specific (Jones [19], Storer [22], Hansen [20] and Wasserman [21]) internationally widely used prediction equations generated from maximal testing bicycle protocols (Table 1).

Physical activity

Participants were asked to report on how many days a week they were physically active for at least 30 min using the Short QUestionnaire to ASsess Health-enhancing physical activity (PA) (SQUASH). The SQUASH is fairly reproducible and reasonably valid; the overall reproducibility amounted to r = 0.58 (95% confidence interval (CI): 0.36–0.74) and concurrent validity with the Computer Science and Applications, Inc. (CSA) activity monitor amounted to r = 0.45 (95% CI: 0.17–0.66) [31].

Statistical analysis

Data was tested for normal distribution by means of examining the skewness and kurtosis of the variables for the sample as a whole and for males and females separately. Means and standard deviations were calculated

Table 1. Prediction equations for estimation of peak rate of oxygen consumption (VO_{2peak})

Author (unit)	Prediction equations		
	sample	males	females
American College of Sports Medicine (ACSM) [$ml\times kg^{-1}\times min^{-1}$]	$(10.8\times W_{max})/weight+7$	–	–
Jones [ml/min]	$(0.046\times height)-(0.021\times age)-(0.62\times gender)-4.31$ F = 1, M = 0	$(0.046\times height)-(0.021\times age)-4.31$	$(0.046\times height)-(0.021\times age)-4.93$
Hansen [ml/min]*	–	$weight\times(50.75-0.372(age))$	$(weight+43)\times(22.78-0.17(age))$
Storer [ml/min]	–	$10.51\times W_{max}+6.35\times weight-10.49\times age+519.3$	$9.39\times W_{max}+7.7\times weight-5.88\times age+136.7$
Wasserman [ml/min]**	$(58\times weight)+151+(10.1\times W_{max})$	$(AW+NPW/2)\times(50.72-(0.372\times age))$ $AW\times50.72-(0.372\times age)$ $AW\times(50.72-(0.372\times age))+6\times(AW-NPW)$	$(AW-NPW+86/2)\times22.78-(0.17\times age)$ $(AW+43)\times22.78-(0.17\times age)$ $(AW+43)\times22.78-(0.17\times age)+6\times(AW-NPW)$

W_{max} – maximal wattage; F – female; M – male; AW – measured actual weight; NPW – normal predicted weight.
* When actual weight > predicted; the predicted weight should be used.
** NPW formula for males: $0.79\times height-60.7$; for females $0.65\times height-42.8$. 1 = AW < NPW; 2 = AW = NPW; 3 = AW > NPW.

for the characteristics of the subjects, for measured and predicted VO$_{2peak}$. The percentage of current smokers was calculated as well as the percentage of subjects in different body mass index (BMI) categories. For PA we calculated the percentage of subjects that reported to have been physically active for 30 min, 5 days per week or more.

Estimated VO$_{2peak}$ (ml×kg^{-1}×min^{-1}) was calculated using the prediction equations. As the Hansen and Wasserman equations take into account whether subjects are normal, under- or over normal weight, we calculated normal predicted weight (NPW) using the formula (0.79×height)−60.7 for males and (0.65×height)−42.8 for females [20,21]. We then calculated the percentage relative weight by dividing actual measured weight by predicted normal weight. We generated 3 Wasserman VO$_{2peak}$ (ml×kg^{-1}×min^{-1}) predictions for males and females; 1 for those weighing less than NPW (Wasserman < NPW), 1 for those weighing equal to NPW (Wasserman = NPW) and 1 for those weighing more than NPW (Wasserman > NPW) as proposed by Wasserman [21]. We generated 2 Hansen VO$_{2peak}$ (ml×kg^{-1}×min^{-1}) predictions for males and females; 1 for those weighing the same or less than NPW (Hansen NPW) and 1 for those weighing more than NPW (Hansen > NPW) as proposed by Hansen [20].

To compare group means we used 2-tailed paired t-testing for the whole sample as well as for each gender to test whether the null-hypothesis (measured VO$_{2peak}$ (ml×kg^{-1}×min^{-1}) = predicted VO$_{2peak}$ (ml×kg^{-1}×min^{-1})) should be rejected for the various prediction equations. Alpha was set at 0.05.

To investigate differences in measured and predicted values between individuals, we calculated the mean difference and limits of agreement (LoA) as proposed by Bland and Altman [32]. A Bland-Altman plot is a graphic representation of the individual subject differences between the tests plotted against the respective individual means. Using this plot, rather than the conventional test-retest scatter gram, a rough indication of systematic bias and random error is provided by examining the direction and magnitude of the scatter around the zero line, respectively. The limits of agreement were defined as the mean difference ±2 standard deviations of the difference between the actual and predicted measurements. The closer the mean difference was to 0 and the smaller the standard deviations in the Bland Altman plots, the better we considered the agreement.

In the absence of a gold standard we considered equations clinically valid for individuals when the difference between measured and predicted VO$_{2peak}$ did not exceed ±3.5 ml O$_2$×kg^{-1}×min^{-1} or ±1 metabolic equivalent (MET) in 95% of individuals. To compare all prediction equations, VO$_{2peak}$ was standardized to ml×kg^{-1}×min^{-1}. For the normative VO$_{2peak}$ data, we calculated percentiles VO$_{2peak}$ in ml×min^{-1} and VO$_{2peak}$ in ml×kg^{-1}×min^{-1} by decade for males and females separately.

All analyses were performed with Stata (StataCorp version 7).

RESULTS

Population

Using the inclusion criterion of volitional exhaustion and RER > 1.00 yielded data for 1529 subjects; 1068 males and 461 females. Females and smokers were less likely to reach RER > 1.00 (N = 245), but they did not differ significantly in age, height or weight from those who did. For those who reached RER > 1.0, reasons for stopping the bicycle ergometry test were shortness of breath (41%), leg fatigue (42%) or both (17%).

Characteristics of the subjects included in the sample are described in the Table 2. Peak rate of oxygen consumption, weight, height and age had skewness near 0 and kurtosis near 3 for the sample, males and females, indicating normal distribution of these variables.

Mean RER was 1.10, indicating a (near) maximal effort in our sample.

The mean (M) (± standard deviation (SD)) percentage relative weight (actual weight/NPW) was 102.8±11.6%

Table 2. Characteristics of the study group

Variable	Males (N = 1 068)	Females (N = 461)
Age (M±SD) [years]	40.9±10.3	35.3±9.5
Height (M±SD) [cm]	181.9±6.4	169.5±6.4
Weight (M±SD) [kg]	85.2±10.6	68.7±11.1
Current smoking – yes [%]	18.8	20.3
Normal predicted weight (M±SD) [kg] [20,21]	83.0±5.1	67.4±4.2
BMI (M±SD)	25.8±2.9	23.8±3.7
< 18.5 [% of subjects]	0	2
18.5–25.0 [% of subjects]	43	66
25.0–30.0 [% of subjects]	49	2
> 30.0 [% of subjects]	8	7
W_{peak} (M±SD)	268.9±47.3	189.5±36.5
HR_{peak} (M±SD) [bpm]	170.6±13.5	170.1±13.6
VO_{2peak} (M±SD) [ml×min^{-1}]	3 204.9±597.9	2 136.7±428.1
VO_{2peak} (M±SD) [ml×kg^{-1}×min^{-1}]	38.0±7.5	31.6±6.8
20–29 years	43.3±6.2	34.6±5.9
30–39 years	39.4±6.5	32.3±6.4
40–49 years	37.5±7.2	29.5±6.1
50–59 years	33.2±6.9	23.8±4.4
RER (M±SD)	1.10±0.1	1.10±0.1
EqO_2 (M±SD) [ml/min]	31.5±4.8	31.0±4.6
PA 30 min, 5 days/week or more [%]	70	70

M – mean; SD – standard deviation; BMI – body mass index; W_{peak} – peak wattage; VO_{2peak} – peak rate of oxygen consumption; HR_{peak} – peak heart rate; RER – respiratory exchange ratio; EqO_2 – equivalent for oxygen; PA – physical activity.

for males and 102±16% for females. According to the formula for NPW, 56% of the males and 46% of the females weighed more than NPW. According to BMI measurements, 56% of males and 32% of females were overweight (BMI > 25) or obese (BMI > 30). In none of the subjects, NPW equalled measured weight.

Non-gender specific equations

For the non-gender specific equations, statistically significant differences were found at the group level between measured and predicted VO_{2peak}, except for the Jones equation. The largest mean difference of 4.2 ml×kg^{-1} ×min^{-1} was found for the ACSM prediction. The ACSM and Wasserman prediction equations overestimated measured VO_{2peak} (ml×kg^{-1}×min^{-1}). Although the Jones equation did not demonstrate statistically significant differences between the means of measured and predicted VO_{2peak} (ml×kg^{-1}×min^{-1}), the LoA were very large, ranging 12.6–12.4 ml×kg^{-1}×min^{-1}. This means that this equation can under- or over-predict by as much as 3.6 METs in an individual (Table 3, Figure 1). None of the equations met the validity criterion of 95% of individuals having ±1 MET difference or less than the measured value.

The table 3 shows the mean predicted VO_{2peak} (ml× kg^{-1}×min^{-1}) for each prediction equation, the mean (SD) difference between measured and predicted VO_{2peak}

Table 3. Non-gender specific prediction equations

Prediction equation	Predicted VO$_{2peak}$ (M±SD) [ml×kg^{-1}×min^{-1}]	Group difference (M±SD) [ml×kg^{-1}×min^{-1}]	t-test [p]	Bland and Altman LoA [ml×kg^{-1}×min^{-1}]	±1 MET difference or less [%]
ACSM	40.3±7.0	4.2±2.8	p = 0.0001	−1.5–9.8	39.4
Jones	36.1±6.8	−0.1±6.2	p = 0.5900[ns]	−12.6–12.4	45.3
Storer	37.4±7.1	1.5±2.8	p = 0.0001	−4.2–7.1	76.5
Wasserman	38.9±6.6	2.7±3.0	p = 0.0001	−3.3–8.8	71.0

ns – not significant; LoA – limits of agreement; MET – metabolic equivalent.
Other abbreviations as in Tables 1 and 2.
Negative values indicate underestimation: lower predicted VO$_{2peak}$ [ml×kg^{-1}×min^{-1}] than measured VO$_{2peak}$ [ml×kg^{-1}×min^{-1}].
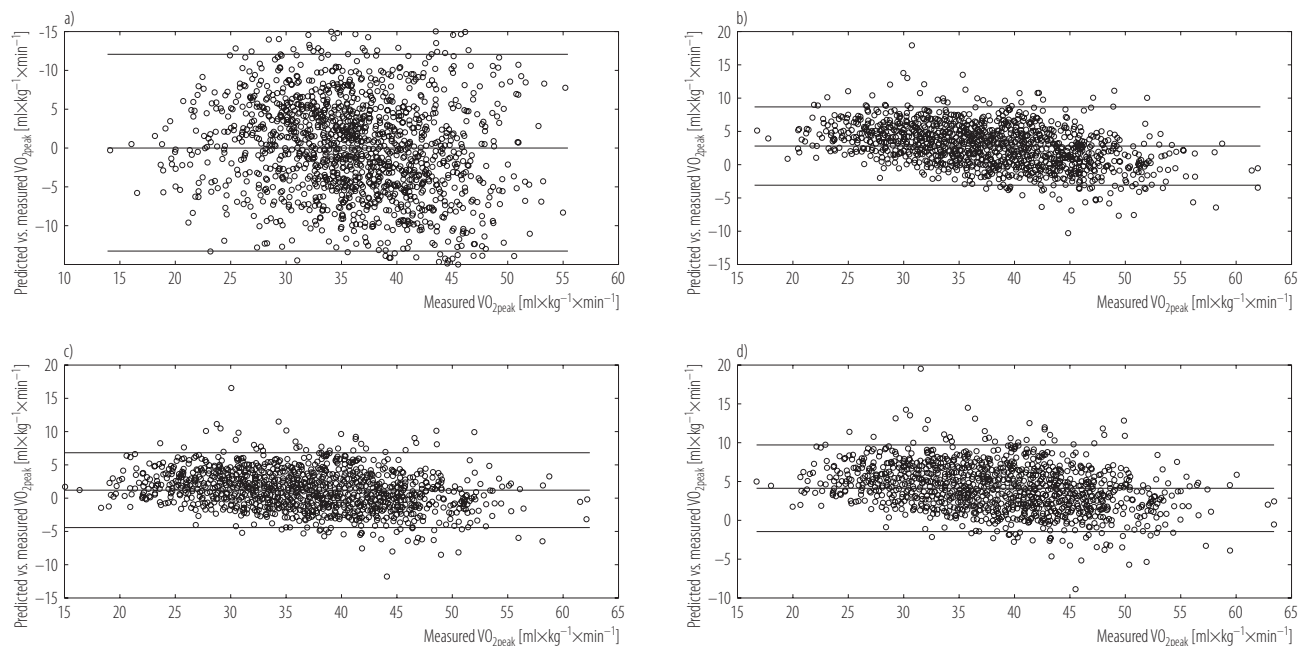
VO$_{2peak}$ – peak rate of oxygen consumption.

Fig. 1. Bland and Altman plots lines for 95% agreement have disappeared for a) Jones, b) Wasserman, c) Storer, and d) American College of Sports Medicine equations

(ml×kg^{-1}×min^{-1}), the p-value of the paired t-testing, 95% LoA and percentage of individuals with a difference ±3.5 ml×kg^{-1}×min^{-1} (1 MET) from the measured value.

Gender specific equations

Paired t-tests equations revealed significant (p < 0.0001) group differences between measured and predicted VO$_{2peak}$ (ml×kg^{-1}×min^{-1}) for both males and females

except for the Jones prediction for males and females and the Hansen < NPW prediction in males.

The Storer and Wasserman > NPW predictions significantly overestimated VO$_{2peak}$ (ml×kg^{-1}×min^{-1}) in males and females. The Wasserman > NPW equation showed the largest average systematic overprediction of 11.1 ml×kg^{-1}×min^{-1} (3.2 METs) in males and 7.7 ml×kg^{-1}×min^{-1} (2.2 METs) in females. All other equations

significantly underestimated VO_{2peak} $(ml \times kg^{-1} \times min^{-1})$ in males and females on average. Most equations showed a tendency to underpredict VO_{2peak} at higher VO_{2peak} levels for individuals. Limits of agreement (LoA) for individual differences between measured and predicted VO_{2peak} were very large for all equations, with the Storer equation for males and females showing the smallest LoA. None of the equations met the validity criterion of 95% of individuals having ±1 MET difference or less than the measured value.

Tables 4 (males) and 5 (females) show the mean estimated VO_{2peak} $(ml \times kg^{-1} \times min^{-1})$ for each of the genders for all prediction equations, the mean (SD) difference between measured and predicted VO_{2peak} $(ml \times kg^{-1} \times min^{-1})$, 95% LoA and percentage of individuals with a difference ±3.5 $ml \times kg^{-1} \times min^{-1}$ (1 MET) from the measured value.

Normative data

As expected, VO_{2peak} values declined by decade and were lower in females than in males.

Percentiles VO_{2peak} in $ml \times min^{-1}$ and $ml \times kg^{-1} \times min^{-1}$ for males and females grouped by decade are reported in the Table 6.

Table 4. Prediction equations for males

Prediction equation	Males VO_{2peak} (M±SD) [$ml \times kg^{-1} \times min^{-1}$]	Difference (M±SD) [$ml \times kg^{-1} \times min^{-1}$]	t-test p	95% LoA [$ml \times kg^{-1} \times min^{-1}$]	±1 MET difference or less [%]
Jones	38.0±5.9	–0.1±6.4	p = 0.6900[ns]	–12.9–12.7	44.1
Storer	41.1±7.3	3.0±2.8	p < 0.0001	–2.5–8.6	69.0
Hansen < NPW	36.6±4.0	–5.0±6.8	p < 0.0001	–19.2–8.9	31.0
Hansen > NPW	34.8±3.5	–0.5±6.1	p < 0.0300	–12.7–11.6	51.5
Wasserman < NPW	38.1±4.7	–3.5±6.9	p < 0.0001	–17.2–10.2	28.7
Wasserman > NPW	46.5±2.9	11.2±6.2	p < 0.0001	–1.48–23.7	8.1

Negative values as in Table 3.
Abbreviations as in Tables 1–3.

Table 5. Prediction equations for females

Prediction equation	Females VO_{2peak} (M±SD) [$ml \times kg^{-1} \times min^{-1}$]	Difference (M±SD) [$ml \times kg^{-1} \times min^{-1}$]	t-test p	95% LoA [$ml \times kg^{-1} \times min^{-1}$]	±1 MET difference or less [%]
Jones	31.6±6.6	–0.1±5.9	p = 0.7000[ns]	–11.9–11.7	50.8
Storer	33.1±6.1	1.4±2.4	p < 0.0001	–3.4–6.2	79.1
Hansen < NPW	29.3±2.9	–5.2±5.7	p = 0.0001	–16.5–6.2	31.5
Hansen > NPW	25.6±2.8	–3.0±5.0	p < 0.0001	–13.0–7.0	45.8
Wasserman < NPW	30.3±3.3	–4.2±5.8	p < 0.0001	–15.7–7.3	37.5
Wasserman > NPW	36.3±1.0	7.7±6.1	p < 0.0001	–4.4–19.9	19.6

Negative values as in Table 3.
Abbreviations as in Tables 1–3.

Table 6. Percentiles of peak rate of oxygen consumption – VO_{2peak} for males and females

Age [years]	Percentiles of mean VO_{2peak} [ml×min⁻¹] (mean VO_{2peak} [ml×kg⁻¹×min⁻¹])										
	95th	90th	80th	70th	60th	50th	40th	30th	20th	10th	5th
Males											
20–29 (N = 129)	4 344.8 (53.4)	4 138.7 (51.4)	3 862.4 (48.5)	3 672.2 (46.6)	3 550.6 (44.4)	3 389.5 (43.5)	3 287.0 (42.0)	3 135.0 (40.2)	2 960.0 (38.2)	2 830.6 (35.6)	2 668.7 (33.6)
30–39 (N = 234)	4 429.5 (51.2)	4 140.5 (47.6)	3 869.0 (45.3)	3 726.0 (42.8)	3 554.0 (40.5)	3 395.5 (38.7)	3 213.0 (37.4)	2 978.5 (35.7)	2 812.0 (34.3)	2 628.0 (31.0)	2 413.5 (29.3)
40–49 (N = 391)	4 108.2 (50.1)	3 898.8 (46.4)	3 689.2 (43.8)	3 531.8 (40.9)	3 361.0 (39.2)	3 210.0 (37.1)	3 017.2 (35.0)	2 894.2 (33.6)	2 712.4 (31.3)	2 429.0 (28.4)	2 261.8 (26.3)
50–59 (N = 243)	3 834.2 (46.1)	3 537.0 (42.6)	3 293.2 (38.6)	3 143.4 (36.3)	2 940.4 (34.8)	2 806.0 (32.4)	2 705.2 (30.7)	2 573.6 (29.1)	2 411.4 (27.5)	2 239.2 (24.7)	2 068.2 (22.7)
Females											
20–29 (N = 131)	3 063.8 (44.4)	2 823.2 (42.9)	2 592.8 (39.8)	2 499.2 (37.4)	2 396.0 (36.0)	2 261.0 (34.7)	2 147.2 (33.4)	2 077.6 (32.3)	1 946.8 (30.4)	1 729.0 (26.9)	1 551.0 (24.6)
30–39 (N = 171)	2 748.5 (42.3)	2 613.0 (40.9)	2 459.4 (37.5)	2 404.0 (35.9)	2 301.6 (34.5)	2 168.5 (32.4)	2 067.2 (30.7)	1 961.9 (28.7)	1 828.6 (26.1)	1 622.6 (23.6)	1 458.3 (21.6)
40–49 (N = 108)	2 722.3 (39.2)	2 541.2 (37.9)	2 436.2 (34.5)	2 310.2 (32.7)	2 165.8 (30.7)	2 048.5 (28.6)	1 938.8 (27.6)	1 881.9 (26.7)	1 766.0 (24.6)	1 628.8 (21.4)	1 481.5 (19.9)
50–59 (N = 44)	2 265.5 (31.2)	2 218.0 (29.9)	2 004.0 (28.5)	1 900.5 (25.8)	1 747.0 (25.1)	1 651.5 (24.0)	1 543.0 (22.3)	1 525.0 (21.5)	1 422.0 (19.7)	1 313.0 (18.3)	1 064.0 (16.2)

DISCUSSION

The main results of this study are that substantial differences were observed between predicted and measured VO_{2peak} (ml×kg^{-1}×min^{-1}) both at group and individual level. Therefore, we feel the prediction equations should not be used to generate normal reference values, at least not in law enforcement officers.

For the group comparisons, only the Jones equation yielded no statistically significant differences between the measured and the predicted VO_{2peak} (ml×kg^{-1}×min^{-1}) for the sample and males and females separately. The LoA for the Jones equation was, however, very large indicating substantial prediction error for individuals, under- or overestimating VO_{2peak} (ml×kg^{-1}×min^{-1}) by as much as 12.6 ml×kg^{-1}×min^{-1} (3.6 METs) in a single individual.

Limits of agreement were (very) large for all equations and none of the equations met our pre-set clinical validity criterion of no more than ±1 MET difference in 95% of the cases between measured and predicted VO_{2peak} (ml×kg^{-1}×min^{-1}).

The relevance of these findings may be substantial because submaximal testing and prediction equations are commonly used all over the world [33,34]. Consequently, individuals are classified based on invalid methods, which may often lead to incorrect classifications for fitness and/or fitness for duty.

To the best of our knowledge, this is the 1st study that has attempted to set an *a priori* criterion for a clinically acceptable measurement difference between measured and predicted VO_{2peak} (ml×kg^{-1}×min^{-1}) for individuals. In the absence of any guidance from the existing literature, we set this criterion at ±3.5 ml×kg^{-1}×min^{-1} or ±1 MET or less in 95% of the subjects. Although this can be criticized for being too small a difference in too many subjects, it represents almost 2 deciles in the normative values table (Table 6) by each decade for males and females.

Setting an acceptable measurement difference *a priori* in validation studies is helpful in determining how "good"

a prediction equation is. For instance, even though the Storer equations showed statistically significant differences between measured and predicted VO_{2peak} for the non-gender and male and female specific equations, they yielded the largest percentages of subjects with a difference of ±1 MET or less (76.5%, 69% and 79.1%, respectively). In contrast, the Jones equations, while yielding statistical non-significance between measured and predicted VO_{2peak}, achieved less than 51% of subjects with a difference of ±1 MET or less. We recommend future researchers to determine an acceptable measurement difference before validating an equation.

When comparing the percentiles of VO_{2peak} (ml×kg^{-1}×min^{-1}) for males and females with the published values [2], VO_{2peak} (ml×kg^{-1}×min^{-1}) values in our law enforcement officers were 2–3 ml×kg^{-1}×min^{-1} lower across all percentiles and decades for males and 3–6 ml×kg^{-1}×min^{-1} lower across all percentiles and decades for females. Seventy percent of our sample reported to have been physically active 30 min. on 5 days/week or more, but not all law enforcement officers are out on the streets. Many have office jobs, which might explain a lower VO_{2peak} (ml×kg^{-1}×min^{-1}) as compared to normative values. Another explanation might be that subjects were given no verbal encouragement while being tested, which may have lead to suboptimal VO_{2peak} values. However, the average RER ratio of 1.10 indicates near maximal effort.

Based on our findings and on those of other researchers [17,26], it appears that prediction equations are population specific and depend on the age range, gender, mode of exercise testing and cardiovascular fitness of the sample that the prediction equation was derived from.

The participants of the UP-LIFT study freely volunteered for the study. This may have resulted in bias, with healthier than average subjects, limiting generalizability to the general working population. Our sample was less obese, smoked less and was more physically active than the general Dutch population.

Normal predicted weight and BMI showed similar results in males, with the cut-off for > NPW appearing to lie around a BMI of 25. There was a discrepancy between the NPW formula (46% of women overweight) and BMI (32% of women overweight) for females. The NPW cut-off for women yields a BMI of 23.5, suggesting the NPW formula for women needs some adjusting. This may also explain why the predictions for Hansen and Wasserman < NPW underestimated VO$_{2peak}$ (ml×kg^{-1}×min^{-1}) in females more than in males and why the Wasserman > NPW prediction overestimated less in females than in males.

The large individual differences between measured VO$_{2peak}$ (ml×kg^{-1}×min^{-1}) and predicted VO$_{2peak}$ (ml×kg^{-1}×min^{-1}) support the notion that prediction equations are not precise (valid) enough for determination of VO$_{2peak}$ (ml×kg^{-1}×min^{-1}) in an individual subject. Van Ooij et al. [17] equally found that none of the predictions were able to accurately predict at an individual level.

CONCLUSIONS

To conclude, we cannot confirm the validity of any of the equations for estimating VO$_{2peak}$ (ml×kg^{-1}×min^{-1}) for the sample and males and females separately at a group level. None of these equations are precise enough to predict VO$_{2peak}$ (ml×kg^{-1}×min^{-1}) for individual subjects. Using prediction equations to estimate cardiorespiratory capacity leads to unacceptably large under- or overestimations of VO$_{2peak}$ in single individuals and should not be used in occupational health settings. We therefore recommend measuring VO$_{2peak}$ by maximal exercise testing using respiratory gas exchange analysis in occupational health settings when personnel is tested for preventative and evaluative purposes within the workplace. Consequently, we recommend not to use the prediction equations analyzed in this paper for individuals, as this would lead to invalid results and clinical recommendations.

REFERENCES

1. Bouchard C, Shephard R. Physical activity, fitness and health: The model and key concepts. In: Bouchard C, Shephard R, Stephens T, editors. Physical activity, fitness and health. International proceedings and consensus statement. Champaign (IL): Human Kinetics; 1994. p. 77–88.

2. American College of Sports Medicine. ACSM's guidelines for exercise testing and prescription. 6th ed. Philadelphia: Lippincott Williams & Wilkins; 2000. p. 73.

3. Vanhees L, Lefevre J, Philippaerts R, Martens M, Huygens W, Troosters T, et al. How to assess physical activity? How to assess physical fitness? Eur J Cardiovasc Prev Rehabil. 2005;12(2):102–14, http://dx.doi.org/10.1097/01.hjr.0000161551.73095.9c.

4. Day JR, Rossiter HB, Coats EM, Skasick A, Whipp BJ. The maximally attainable VO$_2$ during exercise in humans: The peak vs. maximum issue. J Appl Physiol. 2003;95(5):1901–7.

5. Federal Bureau of Investigation. FBI police officers physical requirements [cited 2014 May 29]. Available from: https://www.fbijobs.gov/explore-careers/ps-fbi-police.asp.

6. Ebeling P. Physical fitness in law enforcement. FBI Law Enforce Bull. 2002;71(10):1–7.

7. Åstrand P, Rodahl K, Dahl H, Strømme S. Textbook of work physiology. Physiological bases of exercise. 4th ed. Champaign (IL): Human Kinetics; 2003. p. 279–82.

8. Ilmarinen J. Job design for the aged with regard to decline in their maximal aerobic capacity. Part I – Guidelines for the practitioner. Part II – The scientific base for the guide. Int J Ind Ergon. 1992;10:53–77, http://dx.doi.org/10.1016/0169-8141(92)90048-5, http://dx.doi.org/10.1016/0169-8141(92)90049-6.

9. Takken T, Ribbink A, Heneweer H, Moolenaar H, Wittink H. Workload demand in police officers during mountain

bike patrols. Ergonomics. 2009;52(2):245–50, http://dx.doi.org/10.1080/00140130802334553.

10. Yoopat P, Toicharoen P, Boontong S, Glinsukon T, Vanwonterghem K, Louhevaara V. Cardiorespiratory capacity of Thai workers in different age and job categories. J Physiol Anthropol Appl Human Sci. 2002;21(2):121–8, http://dx.doi.org/10.2114/jpa.21.121.

11. Dey NC, Samanta A, Saha R. A study of the workload of underground trammers in the Ranigang coal field area of West Bengal, India. Int J Occup Saf Ergon. 2006;12(4):399–407.

12. Garver JN, Jankovitz KZ, Danks JM, Fittz AA, Smith HS, Davis SC. Physical fitness of an industrial fire department vs. a municipal fire department. J Strength Cond Res. 2005;19(2):310–7, http://dx.doi.org/10.1519/00124278-200505000-00013.

13. Hammer RL, Heath EM. Comparison of aerobic capacity in annually certified and uncertified volunteer firefighters. J Strength Cond Res. 2013 May;27(5):1435–40, http://dx.doi.org/10.1519/JSC.0b013e318265aaf7.

14. Giovannetti JM, Bemben M, Bemben D, Cramer J. Relationship between estimated aerobic fitness and injury rates among active duty at an Air Force base based upon two separate measures of estimated cardiovascular fitness. Mil Med. 2012;177(1):36–40, http://dx.doi.org/10.7205/MILMED-D-11-00225.

15. Hoffman JR, Kahana A, Chapnik L, Shamiss A, Davidson B. The relationship of physical fitness on pilot candidate selection in the Israel Air Force. Aviat Space Environ Med. 1999;70(2):131–4.

16. Kurumatani N, Yamaguchi B, Dejima M, Enomoto Y, Moriyama T. Aerobic capacity of forestry workers and physical demands of forestry operations. Eur J Appl Physiol Occup Physiol. 1992;64(6):546–51, http://dx.doi.org/10.1007/BF00843766.

17. Van Ooij PJAM, Takken T, Houtkoper A, van Hulst RA. [Measured versus predicted maximal oxygen uptake: A world of difference?]. TBV. 2009;17(10):441–6. Dutch.

18. Kayihan G, Özkan A, Köklü Y, Eyuboğlu E, Akça F, Koz M, et al. Comparative analysis of the 1-mile run test evaluation formulae: Assessment of aerobic capacity in male law enforcement officers aged 20–23 years. Int J Occup Med Environ Health. 2014;27(2):165–74, http://dx.doi.org/10.2478/s13382-014-0237-0.

19. Jones NL, Makrides L, Hitchcock C, Chypchar T, McCartney N. Normal standards for an incremental progressive cycle ergometer test. Am Rev Respir Dis. 1985;131(5):700–8.

20. Hansen JE, Sue DY, Wasserman K. Predicted values for clinical exercise testing. Am Rev Respir Dis. 1984;129 (2 Pt 2):S49–55.

21. Wasserman K, Hansen JE, Sue DY, Stringer WW, Whipp BJ. Principles of exercise testing and interpretation. 4th ed. Philadelphia: Lippincott Williams & Wilkins; 2005.

22. Storer TW, Davis JA, Caiozzo VJ. Accurate prediction of VO_{2max} in cycle ergometry. Med Sci Sports Exerc. 1990;22(5):704–12, http://dx.doi.org/10.1249/00005768-199010000-00024.

23. Ross RM. ATS/ACCP statement on cardiopulmonary exercise testing. Am J Respir Crit Care Med. 2003;167(10):1451, http://dx.doi.org/10.1164/ajrccm.167.10.950.

24. Balady GJ, Arena R, Sietsema K, Myers J, Coke L, Fletcher GF, et al. Clinician's guide to cardiopulmonary exercise testing in adults: A scientific statement from the American Heart Association. Circulation. 2010;122(2):191–225, http://dx.doi.org/10.1161/CIR.0b013e3181e52e69.

25. Guazzi M, Adams V, Conraads V, Halle M, Mezzani A, Vanhees L, et al. EACPR/AHA Scientific Statement. Clinical recommendations for cardiopulmonary exercise testing data assessment in specific patient populations. Circulation. 2012;126(18):2261–74, http://dx.doi.org/10.1161/CIR.0b013e31826fb946.

26. Malek MH, Berger DE, Housh TJ, Coburn JW, Beck TW. Validity of VO_{2max} equations for aerobically trained males and females. Med Sci Sports Exerc. 2004;36(8):1427–32, http://dx.doi.org/10.1249/01.MSS.0000135795.60449.CE.

27. Sörensen L, Smolander J, Louhevaara V, Korhonen O, Oja P. Physical activity, fitness and body composition of Finnish police officers: A 15-year follow-up study. Occup Med. (Lond) 2000;50(1):3–10, http://dx.doi.org/10.1093/occmed/50.1.3.

28. Soininen H. The feasibility of worksite fitness programs and their effects on the health, physical capacity and work ability of aging police officers. Kuopio: Kuopio University Publications D. Medical Sciences; 1995.

29. World Medical Association. Ethical principles for medical research involving human subjects. 2008 [cited 2014 Jun 2]. Available from: http://www.wma.net/en/30 publications/10policies/b3/17c.pdf.

30. Sassen B, Cornelissen VA, Kiers H, Wittink H, Kok G, Vanhees L. Physical fitness matters more than physical activity in controlling cardiovascular disease risk factors. Eur J Cardiovasc Prev Rehabil. 2009;16(6):677–83, http://dx.doi.org/10.1097/HJR.0b013e3283312e94.

31. Wendel-Vos GC, Schuit AJ, Saris WH, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. J Clin Epidemiol. 2003;56(12):1163–9, http://dx.doi.org/10.1016/S0895-4356(03)00220-8.

32. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307–10, http://dx.doi.org/10.1016/S0140-6736(86)90837-8.

33. Sartor F, Vernillo G, de Morree HM, Bonomi AG, la Torre A, Kubis HP, et al. Estimation of maximal oxygen uptake via submaximal exercise testing in sports, clinical, and home settings. Sports Med. 2013;43(9):865–73, http://dx.doi.org/10.1007/s40279-013-0068-3.

34. Tierney MT, Lenar D, Stanforth PR, Craig JN, Farrar RP. Prediction of aerobic capacity in firefighters using submaximal treadmill and stairmill protocols. J Strength Cond Res. 2010;24(3):757–64, http://dx.doi.org/10.1519/JSC.0b013e3181c7c282.