Validation of the firefighter WFI treadmill protocol for predicting VO$_2$ max

B. A. Dolezal$^1$, D. Barr$^2$, D. M. Boland$^1$, D. L. Smith$^2$ and C. B. Cooper$^1$

$^1$Department of Medicine and Physiology, Exercise Physiology Research Laboratory, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA, $^2$Department of Health and Exercise Sciences, First Responder Health and Safety Laboratory, Skidmore College, Saratoga Springs, NY 12866, USA.

Correspondence to: B. A. Dolezal, Department of Medicine and Physiology, Exercise Physiology Research Laboratory, David Geffen School of Medicine, 10833 Le Conte Avenue, CHS 37–131, University of California, Los Angeles, CA 90095, USA.
Tel: +1 310 741 8954; fax: +1 310 206 8211; e-mail: drbducla@gmail.com

Background The Wellness-Fitness Initiative submaximal treadmill exercise test (WFI-TM) is recommended by the US National Fire Protection Agency to assess aerobic capacity (VO$_2$ max) in firefighters. However, predicting VO$_2$ max from submaximal tests can result in errors leading to erroneous conclusions about fitness.

Aims To investigate the level of agreement between VO$_2$ max predicted from the WFI-TM against its direct measurement using exhaled gas analysis.

Methods The WFI-TM was performed to volitional fatigue. Differences between estimated VO$_2$ max (derived from the WFI-TM equation) and direct measurement (exhaled gas analysis) were compared by paired t-test and agreement was determined using Pearson Product-Moment correlation and Bland–Altman analysis. Statistical significance was set at $P < 0.05$.

Results Fifty-nine men performed the WFI-TM. Mean (standard deviation) values for estimated and measured VO$_2$ max were 44.6 (3.4) and 43.6 (7.9) ml/kg/min, respectively ($P < 0.01$). The mean bias by which WFI-TM overestimated VO$_2$ max was 0.9 ml/kg/min with a 95% prediction interval of ±13.1. Prediction errors for 22% of subjects were within ±5%; 36% had errors greater than or equal to ±15% and 7% had greater than ±30% errors. The correlation between predicted and measured VO$_2$ max was $r = 0.55$ (standard error of the estimate = 2.8 ml/kg/min).

Conclusions WFI-TM predicts VO$_2$ max with 11% error. There is a tendency to overestimate aerobic capacity in less fit individuals and to underestimate it in more fit individuals leading to a clustering of values around 42 ml/kg/min, a criterion used by some fire departments to assess fitness for duty.

Key words Firefighters; fitness tests; physical fitness.

Introduction

Firefighting requires strenuous lifting and meticulous manoeuvring while wearing cumbersome personal protective equipment (often weighing >25 kg) in high ambient temperatures (100°C is considered routine) under stressful conditions [1]. These physical and psychological factors probably contribute to high rates of injury and cardiovascular events [2–4]. Furthermore, lack of physical fitness and deconditioning that may lead to overexertion, coupled with unrecognized cardiovascular disease risk factors, greatly increase the risk of on-duty injuries or death amongst firefighters [2,5,6]. In fact, sudden cardiac events account for the largest proportion of firefighter deaths on duty [7].

In view of this, a joint task force of the International Association of Firefighters and International Association of Fire Chiefs developed the Fire Service Joint Labor Management Wellness-Fitness Initiative (WFI) [8]. The WFI, most recently revised in 2008, aims to improve the quality of life for safety personnel and to produce a working environment conducive to remaining safe, healthy and physically fit. Inherent in this programme is the need for an accurate assessment of a firefighter’s cardiovascular fitness.

Aerobic capacity (VO$_2$ max) is a measure that defines the limits of cardiovascular function. Research data...
suggest that for safety and optimal job performance, firefighters should develop and maintain levels of aerobic capacity at ~41.5 ml/kg/min [5,9]. Based on this research, the US National Fire Protection Agency recommends that a firefighter should have a minimum VO$_2$ max of 42 ml/kg. It is therefore essential that the testing protocol used to determine VO$_2$ max is accurate.

VO$_2$ max can be determined directly or indirectly using maximal or submaximal protocols, respectively. Submaximal tests are less expensive and easier to administer, making them more feasible for fire departments to implement. Previous submaximal tests endorsed by the WFI to predict VO$_2$ max have shown substantial error, increasing the likelihood that erroneous conclusions about fitness may be made. To improve the accuracy of the submaximal protocol used to predict VO$_2$ max, a refined estimation equation was developed to be used in conjunction with the WFI treadmill test (WFI-TM) [10]. One study has previously attempted to assess the reliability of this protocol and deemed it to be valid [10]. However, the statistical approach employed did not allow for comparisons across a range of fitness levels. As firefighters have been shown to be a heterogeneous group in terms of aerobic performance, validations that are applicable to all firefighters are important. The purpose of this study was therefore to establish the level of agreement between aerobic capacity (VO$_2$ max) estimated from the WFI treadmill protocol (WFI-TM) and actual measurement of VO$_2$ max using exhaled gas analysis.

**Methods**

Subjects were recruited at two sites; the Exercise Physiology Research Laboratory at the University of California, Los Angeles and the Human Performance Laboratory at Skidmore College, Saratoga Springs, NY. The study was approved by the Offices for the Protection of Human Subjects in Research at both institutions. All subjects gave written informed consent and completed a medical and exercise history questionnaire. Inclusion criteria included absence of musculoskeletal conditions or cardiovascular, pulmonary, metabolic or other disorders that would preclude high-intensity exercise testing.

All testing was conducted by trained and experienced personnel in accordance with established guidelines for cardiopulmonary exercise testing [11]. A calibrated metabolic measurement system (Oxycon Mobile, CareFusion, Yorba Linda, CA) was used to measure oxygen uptake (VO$_2$) breath-by-breath. Heart rate was recorded using a wearable monitor (RS400, Polar-Electro, Kempele, Finland). Physiological variables were continuously monitored and recorded during warm up, exercise and recovery. Rating of perceived exertion was recorded at intervals during the test and at maximal exercise using the Borg 6–20 scale [12].

Maximum oxygen uptake was estimated from the WFI-TM protocol using the equation modified in 2008: 

$$\text{VO}_2\text{ max} = 56.981 + (1.242 \times \text{TT}) – (0.805 \times \text{BMI}),$$

where test time (TT) is the time required to achieve target heart rate, determined as 85% of maximal heart rate (208 – (0.7 × age) × 0.85) and BMI is body mass index (kg/m$^2$). The WFI-TM is a modified ramp protocol made-up of a 3 min warm up of 3 m.p.h. at 0% gradient followed by an increase in speed to 4.5 m.p.h. The rest of the test involves 1 min intervals of alternate increases in speed (0.5 m.p.h.) and gradient (2%). Usually this protocol is purposely terminated at the TT when subjects have reached 85% of their predicted age-predicted maximal heart rate for at least 15 seconds. However, for this study, subjects continued the test until volitional fatigue in order to measure VO$_2$ directly via exhaled gas analysis.

Maximum oxygen uptake was determined from the highest 15 second average and accepted as maximal in the presence of a plateau in VO$_2$ despite increasing work rate and a maximum heart rate within 16 beats/min of the age-predicted maximum (208 – (0.7 × age)) [13].

Differences between WFI-TM estimated and directly measured VO$_2$ max were assessed by the limits of agreement and compared with a paired t-test. Agreement was also assessed using Pearson Product-Moment correlation and Bland–Altman analysis. Statistical significance was set at $P < 0.05$.

**Results**

Fifty-nine men were recruited at the two test sites. Their average (standard deviation) age was 33 (8) with a BMI of 25.8 (2.6) kg/m$^2$. Mean values for estimated and measured VO$_2$ max were 44.6 (3.4) and 43.6 (7.9) ml/kg/min, respectively ($P < 0.01$), with the WFI equation tending to overestimate VO$_2$ max. Correlation analysis between estimated and measured VO$_2$ max revealed $r = 0.55$ and standard error of the estimate = 2.8 ml/kg/min (Figure 1).

Bland–Altman analysis yielded a mean bias of 0.94 ml/kg/min (11%) with a 95% confidence interval of ±13.1 (Figure 2). Twenty-two per cent of subjects had prediction errors within ±5%; 36% had errors greater than or...
The diagonal line represents the slope of the relationship indicating a significant bias, $P<0.0001$.

equal to $±15\%$ and 7% had errors greater than $±30\%$. There was a tendency to overestimate aerobic capacity in less fit individuals and to underestimate it in more fit individuals, leading to a clustering of values for this study population around 42 ml/kg/min.

**Discussion**

This study shows that the WFI-TM protocol generally predicts true VO$_{2}$ max within an 11% error margin. As such, it could play an important part in assessment of cardiovascular fitness in groups of firefighters. However in individuals, it has a tendency to overestimate aerobic capacity in those less fit and to underestimate it in the more fit. The result is a clustering of estimated values around 42 ml/kg/min, a criterion used in some fire departments to assess fitness for duty, creating an impression of more consistent and adequate physical fitness than might actually be true. VO$_{2}$ max is most accurately measured by performing exhaled gas analysis during treadmill or cycle ergometry to volitional fatigue. Submaximal tests have been developed to minimize the risks associated with strenuous exercise [14] and facilitate estimation of aerobic capacity without the expense and expertise required by exhaled gas analysis. Nonetheless, there are inherent limitations in the accuracy of VO$_{2}$ max estimation from submaximal tests. These methods necessarily assume a linear relationship between heart rate and oxygen uptake (VO$_{2}$) during incremental work and rely on a prediction of maximal heart rate, which is subsequently used to derive VO$_{2}$ max [14–16]. Unfortunately, age-predicted maximum heart rate is largely imprecise, with values falling into a normal distribution and a standard deviation of ~10 beats/min. This means that only 65% of normal subjects will have a true maximum attainable heart rate within $±10$ beats/min of the age-predicted value. Alternatively, only 5% of normal subjects will have a true maximal heart rate $<16$ beats/min below the age-predicted reference value for maximal heart rate. Furthermore, the relationship between heart rate and oxygen uptake may become increasingly non-linear towards the end of a maximal effort. Another common issue with submaximal derivations is that the prediction equations tend to be less accurate for individuals whose fitness levels deviate from the mean [17].

A limitation of this study is that our subjects were leaner (mean BMI = 26) than the average career US firefighter (BMI = 28–29) [5,9]. Although this may limit the extent to which our results can be generalized, our subjects’ average aerobic capacities were similar to the population norm.

According to our data, the WFI protocol appears to be subject to the same inaccuracies observed in other estimations, both overestimating the values of less fit individuals and underestimated the values of more fit ones. Our results lead to similar conclusions to those from an evaluation of a submaximal treadmill protocol (Gerkin Treadmill) previously used in the fire service. In both studies, VO$_{2}$ max could be predicted fairly well for individuals with average fitness levels, but predictions became increasingly inaccurate for those who were significantly above or below the average level [18]. Similarly, in an early study by Fitchett et al. [17], submaximal testing was found to be adequate for analysis of populations (based on averages) but too variable to be considered accurate when testing individuals. For use in the fire service, these inaccuracies are especially problematic in assessing fitness for duty of those of below-average fitness levels and should be considered unacceptable as an overestimation of fitness places firefighters at increased risk of cardiovascular events, a risk treadmill testing aims to reduce.

In contrast to our findings, a recent study by Drew-Nord et al. [10] found more favourable agreement between the revised WFI-TM and directly measured VO$_{2}$ max. We believe two differences in experimental design may explain the discordance of our data with this study. Firstly, in the Drew-Nord et al. [10] study, estimated and measured VO$_{2}$ max values were obtained at separate times, up to 8 weeks apart. Our measures were obtained simultaneously within a single VO$_{2}$ max test, which we believe provides increased reliability as subjects continued to volitional fatigue in the same time course under the same conditions. Secondly, the previous study utilized a t-test to detect the difference between estimated and measured values, whereas we employed a Bland–Altman analysis. Because the t-test compares the means of the two data sets, it is ill-suited to detect the error observed in our study where significant differences occur not so much in mean values, but more as values deviate above or below the mean. With the Bland–Altman plot, we can better visualize that the WFI equation either overestimates or underestimates in a tendency to bias values toward the observed mean of 43.6 ml/kg/min.
The results of our study suggest that the current policy for determining fitness for duty in emergency responders may be inadequate and should be re-evaluated. Overestimating aerobic capacity in those with a lower level of cardiovascular function may lead to serious errors in fitness classification, placing firefighters at risk of less effective work performance, injury or cardiovascular events on duty. These pitfalls of the WFI protocol could, however, be overcome by using a symptom-limited, maximal incremental protocol with estimation of aerobic capacity using an equation such as the Foster equation [19].

Key points

- The Wellness-Fitness Initiative submaximal treadmill exercise test used to assess aerobic capacity in US firefighters was found to predict VO$_2$ max with 11% error and tended to overestimate aerobic capacity in less fit individuals.
- Overestimates of fitness levels could lead to erroneous conclusions about fitness for duty, resulting in increased cardiovascular risk among some emergency responders and thereby jeopardizing public safety.
- These results suggest that the current policy for measuring fitness in emergency responders in the United States should be re-evaluated and the implementation of a more robust approach considered.

Funding

Department of Homeland Security—Science and Technology Directorate.

Acknowledgements

We wish to express our thanks and appreciation to Jeannie Haller, Wes Lefferts and Eric Hultquist for assistance with data collection and management. We are also grateful of the willingness of participants who so ably and enthusiastically participated in this study.

Conflicts of interest

None declared.

References