Prediction of Maximal Aerobic Power From the 20-m Multi-Stage Shuttle Run Test

Michael K. Stickland, Stewart R. Petersen, and Marcel Bouffard

Catalogue Data

Key words: aerobic fitness, maximal oxygen consumption, field test, test validity
Mots-clés: condition physique aérobie, consommation maximale d’oxygène, test de champ, validité d’un test

Abstract/Résumé
The purpose of this study was to test the accuracy of the 20-m multi-stage shuttle run (SR) test to predict VO₂max in young adults. VO₂max was measured during a graded treadmill test in 60 men and 62 women (mean age 25.3 and 25.1 years, respectively). Each subject was familiarized with the SR procedure and then completed the SR test to predict VO₂max on a separate day. The mean terminal SR stage was 9.5 for men and 7.8 for women. The regression equations of Léger et al. (1988) and Léger and Gadoury (1989) systematically underpredicted VO₂max for both males and females (p < 0.05). New regression equations were developed from present data to predict VO₂max for males: \( \hat{Y} = 2.75X + 28.8 \) (\( r^2 = 0.77 \), \( SEE = 4.07 \text{ ml·kg}^{-1}·\text{min}^{-1} \)); and for females: \( \hat{Y} = 2.85X + 25.1 \) (\( r^2 = 0.66 \), \( SEE = 3.64 \text{ ml·kg}^{-1}·\text{min}^{-1} \)), where \( X \) equals the last half-stage of the SR completed. We suggest that these gender-distinct equations provide more accurate predictions of VO₂max from the SR.

Le but de la présente étude était d’évaluer la validité du test progressif de course navette de 20 m avec paliers de Léger et Gadoury (1989) pour prédire la consommation maximale d’oxygène. Soixante hommes et 62 femmes (age moyen 25.3 et 25.1 ans, respectivement) complétèrent une épreuve progressive de consommation maximale d’oxygène sur tapis roulant et la course navette pendant deux journées différentes. Tous les sujets pratiquèrent la course navette une fois avant cette étude. Le nombre moyen de paliers complétés lors de la course navette fût 9.5 pour les hommes et 7.8 pour les femmes. Les résultats indiquèrent...
que les équations de régression linéaire employée par Léger et al. (1988) et Léger et Gadoury (1989) sont biaisées et sous-estime la consommation maximale d’oxygène pour les hommes et les femmes (p < 0.05). De nouvelles équations de régression linéaire furent développées à partir des résultats de la présente étude pour prédire la consommation maximale d’oxygène. Ces équations sont \( \hat{Y} = 2.75X + 28.8 \) pour les hommes (\( r^2 = 0.77, S_{yx} = 4.07 \text{ ml·kg}^{-1} \text{·min}^{-1} \)) et \( \hat{Y} = 2.85X + 25.1 \) pour les femmes (\( r^2 = 0.66, S_{yx} = 3.64 \text{ ml·kg}^{-1} \text{·min}^{-1} \)). En conclusion, nous suggérons que les nouvelles équations de régression linéaire permettent une meilleure prédiction de la consommation maximale d’oxygène.

**Introduction**

The maximal rate of oxygen consumption (\( \dot{V}O_2\text{max} \)) is considered the gold standard for measurement of aerobic fitness (Sutton, 1992). The direct measurement of \( \dot{V}O_2\text{max} \) demands sophisticated instrumentation, laboratory time, and trained personnel, and it may not be appropriate for some applications. For these reasons there is interest in predictive tests that can serve as convenient alternatives to direct \( \dot{V}O_2\text{max} \) measurement.

The 20-m multi-stage shuttle run test (SR), originally developed by Léger and Lambert (1982), has been widely used as a predictive test for \( \dot{V}O_2\text{max} \). The SR involves graded exercise to exhaustion with a typical duration of 8–12 minutes. Paliczka et al. (1987) reasoned that the SR is an ideal field test because judgment about pace is eliminated by using audio signals, the incremental test ensures a gradual rise in work rate and heart rate, the test is highly reliable (\( r = 0.975; \) Léger and Lambert, 1982), and large numbers of participants can be tested at a time.

The SR was originally designed to predict fitness in healthy adults attending fitness classes and in athletes participating in sports characterized by frequent starts and stops (Léger et al., 1988). The SR test has also been used to screen for aerobic fitness in occupational settings. Physical fitness standards for employment are normally based on the understanding that a certain level of aerobic fitness is required to work safely and effectively. While direct measurement of physiological attributes such as \( \dot{V}O_2\text{max} \) would be ideal, for many reasons employers frequently find that field tests or predictive tests are more suitable. In these situations the accuracy of decisions on suitability for employment is at least partly dependent on the accuracy of the test utilized.

Léger et al. (1988) noted that the preferred equation for prediction of \( \dot{V}O_2\text{max} \) (\( \hat{Y} \)) for men and women from the maximal aerobic speed (MAS) attained during the shuttle run test was \( \hat{Y} = -24.4 + 6.0 \text{ MAS} \) (Equation 1). Interestingly, a table (Table 1 in Léger et al., 1988) of predicted \( \dot{V}O_2\text{max} \) in the same article yields values that are inconsistent with the equation. For example, if Stage 10 in the SR is completed, the MAS entered into the equation is 13.0 km·h\(^{-1}\), which then predicts \( \dot{V}O_2\text{max} \) as 53.6 ml·kg\(^{-1}\)·min\(^{-1}\), yet the table predicts \( \dot{V}O_2\text{max} \) at 50.6 ml·kg\(^{-1}\)·min\(^{-1}\). No explanation can be found for this apparent discrepancy. In another publication on validation of the SR test as a predictor of \( \dot{V}O_2\text{max} \) in adults, Léger and Gadoury (1989) referred to an earlier published abstract (Mercier et al., 1983) where the formula originally developed for children was modified for use with adults. The formula \( \hat{Y} = -27.4 + 6.0 \text{ MAS} \) predicts \( \dot{V}O_2\text{max} \) values that are consistent with those in Table 1 in Léger et al. (1988).
In the article by Léger and Gadoury (1989), a third predictive equation was reported. The formula \( \hat{Y} = -32.678 + 6.592 \text{MAS} \) (Equation 2) describes the relationship between the peak \( \dot{V}_{\text{O}_2} \) obtained at the end of a graded treadmill test and MAS. Substituting the same MAS as used above for illustrative purposes yields a predicted \( \dot{V}_{\text{O}_2}\text{max} \) of 53.0 ml·kg\(^{-1}\)·min\(^{-1}\), which is very consistent with Equation 1 in Leger et al. (1988) but is not consistent with Table 1 from the same article.

Several concerns arise from this brief review. First, it is apparent that one sample of adults (\( N = 77 \)) has been referred to in both the Léger et al. (1988) and Léger and Gadoury (1989) articles that purport to validate the prediction of \( \dot{V}_{\text{O}_2}\text{max} \) from the SR test. Yet, as noted above, several predictive formulas have been published, and surprisingly, an inconsistency is found within one validation article. There is little to guide the practitioner toward the most appropriate formula.

Second, there is the significant concern that \( \dot{V}_{\text{O}_2}\text{max} \) was not actually measured in either study. In Léger et al. (1988), \( \dot{V}_{\text{O}_2}\text{max} \) was estimated using the backward extrapolation technique (Léger et al., 1980). Léger and Gadoury (1989) used as their criterion measure the results of gas analysis from a Douglas bag sample collected during the final minute of a graded treadmill test to exhaustion. Notwithstanding the authors\' apparent confidence in both techniques, the results remain as only estimates of \( \dot{V}_{\text{O}_2}\text{max} \), not direct measurements.

Third, the sample size (53 M, 24 F) is probably inadequate to conclude with great confidence that one formula accurately predicts \( \dot{V}_{\text{O}_2}\text{max} \) for both men and women. Given what is generally accepted about gender differences in physiological attributes, this hypothesis warrants review.

Despite the high correlations reported between \( \dot{V}_{\text{O}_2}\text{max} \) and SR performance (Léger and Gadoury, 1989; Paliczka et al., 1987; Ramsbottom et al., 1988), there have been other reports that the published predictive equations systematically underestimate the \( \dot{V}_{\text{O}_2}\text{max} \). In a study of male runners and squash players, St Clair Gibson et al. (1998) found that Equation 1 (above) underpredicted \( \dot{V}_{\text{O}_2}\text{max} \) by an average of 5.2 ml·kg\(^{-1}\)·min\(^{-1}\), or 6.8%. Using the equation \( \hat{Y} = -27.4 + 6.0 \text{MAS} \), Berthoin et al. (1994) found that predicted \( \dot{V}_{\text{O}_2}\text{max} \) for male and female physical education students was 5.7 ml·kg\(^{-1}\)·min\(^{-1}\), or 9.3% lower than \( \dot{V}_{\text{O}_2}\text{max} \) measured in the laboratory. While these researchers have identified a potential problem with the accuracy of predicting \( \dot{V}_{\text{O}_2}\text{max} \) from the existing methods, they have not offered an alternative.

Given the widespread use of the 20-m SR test to make potentially important decisions such as suitability for certain occupations, it was of interest to investigate these concerns. The main purpose of the present study was to further explore the hypothesis that the published equations systematically underpredict \( \dot{V}_{\text{O}_2}\text{max} \). A secondary purpose was to re-examine the appropriateness of using a single regression equation to predict \( \dot{V}_{\text{O}_2}\text{max} \) for both males and females.

**Methodology**

**SUBJECTS**

Participating in the present study were 60 men (mean age 25.3 ± 5.0 yrs [range 18–38]; weight = 77.7 ± 11.3 kg; height = 176.9 ± 6.2 cm) and 62 women (mean age 25.1 ± 5.2 yrs [range 18–38]; weight = 63.6 ± 8.7 kg; height = 167.4 ± 5.9 cm).
were active adults and came from a variety of sports backgrounds including basketball, volleyball, soccer, and running. Ethnicity included Caucasian, First-Nation Canadian, and Asian-Canadian. The subjects were informed of the purpose of the study and all gave written consent before participating. The study was reviewed and approved by the appropriate institutional ethics review board.

GRADED EXERCISE TEST

Subjects began the graded exercise test (GXT) by walking on a motor-driven treadmill (Quinton model 18-60) at a constant speed (1.81 m·s⁻¹) at 0% grade. Subsequent increases in workload were characterized by 2% increases in grade every 2 minutes. After the third workload (1.81 m·s⁻¹ at 4%), the second phase of the test began whereby speed was increased to a comfortable running pace (range = 2.47 to 3.58 m·s⁻¹). Grade was then increased by 2% each minute until the subject could no longer continue to exercise. Test duration was typically in the range of 10 to 15 minutes. The criterion for \( V\text{O}_2 \text{max} \) was a plateau in \( V\text{O}_2 \) despite an increase in workload prior to volitional exhaustion. Alternately, the peak \( V\text{O}_2 \) value was accepted as \( V\text{O}_2 \text{max} \) if the respiratory exchange ratio was greater than 1.1 and peak heart rate was similar to the age-predicted maximum at volitional exhaustion. Each subject was encouraged to continue as long as possible.

Respiratory gas exchange data were collected continuously using a Rudolf 2700 valve and were averaged for each 20-second time interval with a Sensormedics 2900z metabolic measurement system (Yorba Linda, CA). The metabolic cart was calibrated with known gas standards immediately before each test, and calibration was verified immediately following each test. Heart rate (HR) was recorded at the end of each minute using a Polar telemetry system (Kempele, Finland).

PRACTICE SESSION

On a separate day, prior to the shuttle run test, all subjects received full instruction and practiced the test. They began the SR test and normally completed up to Stage 4. A few subjects were stopped before reaching Stage 4 as it was evident that they were approaching maximal effort. The goal of the practice session was to familiarize the subjects with the graded effort and pacing required to perform the SR test correctly.

SHUTTLE RUN TEST

The SR test was administered indoors on a nonslip surface in accordance with the procedures described by Léger et al. (1988), in which running velocity was increased by 0.5 km·hr⁻¹ after every 1-minute stage. Heart rate was recorded throughout the test using a Polar telemetry system. The test was stopped if the subject failed to reach the interior pylon prior to the “beep” on two successive occasions. Otherwise the test ended when the subject stopped because of fatigue. Each subject was given encouragement to keep running as long as possible. \( \hat{V}\text{O}_2\text{max} \) was predicted from two equations: Equation 1 (\( \hat{Y} = –24.4 + 6.0 \text{MAS} \); Léger et al., 1988) and Equation 2 (\( \hat{Y} = –32.678 + 6.592 \text{MAS} \); Léger and Gadoury, 1989), where MAS is the maximal aerobic speed achieved during the SR and \( \hat{Y} \) is the predicted \( \hat{V}\text{O}_2\text{max} \) (ml·kg⁻¹·min⁻¹).
The data analysis procedure was divided into five major phases. During the first phase, t-tests for correlated samples were completed for each gender, to compare peak HR during the shuttle run and the graded exercise test. During the second phase, the accuracy of predictions of both equations for both genders was evaluated independently for each equation. At this stage the design of the study was viewed as a Gender × Test design with repeated measures on the test factor. The gender factor had two levels, females and males; the test factor also had two levels, the Léger test and the measured VO2 max test. We wanted to determine whether the differences between predicted and measured VO2 max were different for each gender. These differences were calculated for each gender and analysed using independent sample t-tests. The t-tests for independent samples were in fact testing for the presence of a Gender × Test interaction, as follows:

Let δF represent the difference between predicted and measured VO2 max for females and δM the same difference for males. Conceptually, the null hypothesis stated that δF = δM was tested for statistical significance. However, δF = μFM − μFL and δM = μMM − μML where the first subscript under μ represents the sex (Female or Male) and the second subscript represents the test (Measured VO2 max and Léger test). Consequently, we were testing the null hypothesis stating that μFM − μFL = μMM − μML. Stated differently, a t-test conducted on difference scores is equivalent to a two-way Gender × Test interaction test (for details, see Woodward et al., 1990).

Because Equations 1 and 2 underestimated VO2 max differently for both genders, our goal during the third phase became the development of new prediction equations for each sex. Least-squares linear regression equations were produced to predict VO2 max from the SR test stage where the test was terminated. In a fourth phase, the new regression equations were cross-validated using a procedure described later in this paper. Finally, in the fifth phase, Pearson product-moment correlation coefficients were calculated to estimate the strength of the relationship between SR stage attained and measured VO2 max. For all inferential analyses, the probability of Type I error was set at .05.

### Results

No differences were found between peak heart rate responses to the GXT and the SR (see Table 1). The magnitude of the difference between measured and predicted maximal oxygen consumption, using both the equation of Léger et al. (1988)

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Peak Heart Rate During:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Graded exercise test</td>
<td>Shuttle run</td>
</tr>
<tr>
<td>Females (n = 62)</td>
<td>191 ± 8.8</td>
<td>190 ± 10.9</td>
</tr>
<tr>
<td>Males (n = 60)</td>
<td>195 ± 9.6</td>
<td>194 ± 11.4</td>
</tr>
</tbody>
</table>
Table 2  Comparison of Shuttle Run Level Attained, Predicted and Measured \( \text{VO}_2\text{max} (M \pm SD) \)

<table>
<thead>
<tr>
<th></th>
<th>SR level completed</th>
<th>Predicted ( \text{VO}_2\text{max} ) (Léger et al., 1988)</th>
<th>Predicted ( \text{VO}_2\text{max} ) (Léger &amp; Gadoury, 1989)</th>
<th>Measured ( \text{VO}_2\text{max} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females ((n = 62))</td>
<td>7.8 ± 1.8</td>
<td>46.4 ± 5.3</td>
<td>45.1 ± 5.8</td>
<td>47.4 ± 6.2</td>
</tr>
<tr>
<td>Range</td>
<td>4.0 – 10.5</td>
<td>35.6 – 53.6</td>
<td>33.2 – 53.0</td>
<td>32.1 – 61.0</td>
</tr>
<tr>
<td>Males ((n = 60))</td>
<td>9.5 ± 2.7</td>
<td>51.6 ± 8.1</td>
<td>50.7 ± 8.9</td>
<td>54.9 ± 8.4</td>
</tr>
<tr>
<td>Range</td>
<td>2.0 – 14.0</td>
<td>29.6 – 65.6</td>
<td>26.6 – 66.2</td>
<td>36.1 – 74.0</td>
</tr>
</tbody>
</table>

Figure 1. Scatterplot of predicted \( \text{VO}_2\text{max} (\hat{Y} = 6.0X - 24.4; \text{Léger et al.}, 1988) \) and measured \( \text{VO}_2\text{max} (n = 122) \).

and that of Léger and Gadoury (1989), was significantly larger in men than in women \((p < .05)\) (see Table 2). These results indicated (a) the presence of a Gender \( \times \) Test interaction for both equations and (b) showed that the magnitude of the underestimation was higher for males for both equations. The \( \text{VO}_2\text{max} \) predicted from both equations were significantly lower than the measured \( \text{VO}_2\text{max} \) for both males and females \((p < .05)\) (see Table 2 and Figure 1).
CROSS-VALIDATION

The total sample regression equations were cross-validated indirectly using a computer-intensive method. For each gender, the data were randomly divided into two samples of equal size and least-squares regression solutions were obtained for both samples. The regression weights of the first sample (derivation sample) were also applied to the second sample (cross-validation sample). The same procedure was repeated 10,000 times. This process yielded distributions of regression weights and $r^2$ values. The means of these distributions, along with their standard deviations, are reported in Table 3.

Table 3 Mean of Intercept, Slope, and Squared Multiple Correlation of Computer-Intensive Studies ($M \pm SD$)

<table>
<thead>
<tr>
<th>Variable</th>
<th>1st Sample</th>
<th>2nd Sample</th>
<th>Cross-validation study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>25.15 ± 2.30</td>
<td>25.07 ± 2.29</td>
<td>—</td>
</tr>
<tr>
<td>Slope</td>
<td>2.84 ± 0.27</td>
<td>2.85 ± 0.27</td>
<td>—</td>
</tr>
<tr>
<td>$r^2$</td>
<td>0.66 ± 0.07</td>
<td>0.66 ± 0.07</td>
<td>0.61 ± 0.09</td>
</tr>
<tr>
<td>Males</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>28.78 ± 1.50</td>
<td>28.73 ± 1.51</td>
<td>—</td>
</tr>
<tr>
<td>Slope</td>
<td>2.75 ± 0.17</td>
<td>2.75 ± 0.17</td>
<td>—</td>
</tr>
<tr>
<td>$r^2$</td>
<td>0.77 ± 0.05</td>
<td>0.77 ± 0.05</td>
<td>0.74 ± 0.06</td>
</tr>
</tbody>
</table>

In brief, a comparison of the results obtained by using the first and the second random sample show that the average regression weights are fairly similar to each other. In addition, the standard deviations of each regression coefficient are relatively small. Most important, a relatively small shrinkage in $r^2$ values was obtained by applying the regression weights from the first sample to the data of the second sample. The average shrinkage was 0.05 for women and 0.03 for men. Therefore it seems reasonable to submit that, for each gender, all the data could be safely combined to derive a total sample regression equation. It is important to understand that the total sample regression equation was indirectly validated by this procedure. However, on theoretical grounds, it is to be expected that the total sample regression equation is more valid than a regression equation based on only half of the sample. As noted by Darlington (1990), the average of numerous cross-validity figures is “taken as a conservative estimate of the validity of the final regression, which is developed in the total sample” (p. 161).

Using the data from this study, new least-squares linear regression equations were developed for males and females. The regression equation for females was $\hat{Y} = 2.85 X + 25.1$ ($r^2 = 0.66, p < .05$) and the one for males was $\hat{Y} = 2.75 X + 28.8$ ($r^2 = 0.77, p < .05$), where $\hat{Y}$ is the predicted $\text{VO}_2\text{max}$ and $X$ is the last half-stage of
the SR completed (see Figure 2). The standard errors of the slopes for the new regression equations were found to be 4.07 ml·kg⁻¹·min⁻¹ for males and 3.64 ml·kg⁻¹·min⁻¹ for females.

The homogeneity of regression slopes was not rejected, \( F(1, 118) = .09, p > .05 \), indicating that both sexes have the same predicted increment in oxygen consumption given the same increment on the SR test. Pearson product-moment correlation coefficients between \( \dot{V}O_2\max \) and SR level were 0.88 for men and 0.81 for women. Both correlation coefficients were significant (\( p < .01 \)), although the differences between these correlation coefficients were not (\( z = 1.33, p > .05 \)).

**Discussion**

Both regression Equation 1 (Léger et al., 1988) and Equation 2 (Léger and Gadoury, 1989) systematically underpredicted \( \dot{V}O_2\max \) for both females and males. The magnitude of the underprediction was larger (\( p < 0.05 \)) for males. The finding of the present study, using substantially larger sample sizes, is congruent with previous research detailing the systematic underprediction of \( \dot{V}O_2\max \) by the SR (Berthoin et al., 1994; St Clair Gibson et al., 1998).

No differences were found between peak heart rate from the GXT and the SR test for either gender (see Table 1). We did note that the mean peak HR for men was 4 beats·min⁻¹ higher than for women (\( p < .05 \)). Mean peak HR for men was very consistent with the predicted maximum HR (220 – age), while the actual peak HR for women was slightly lower than the predicted value. No explanation can be
offered for this observation. Both the GXT and SR involved progressive running exercise to exhaustion. Accordingly, it is presumed that both tests had similar physiological demands. Furthermore, all subjects met the criteria for $\dot{V}O_2\text{max}$ on the treadmill and all terminated the SR because they were unable to maintain the required pace. Therefore, it is unlikely that differences in predicted and measured $\dot{V}O_2\text{max}$ were the result of dissimilar physiological end-points.

There are several possible explanations for the systematic underprediction of $\dot{V}O_2\text{max}$ using the two Léger regression equations. Small differences are to be expected simply due to the different groups of subjects being studied. However, it seems unlikely that this could account for the magnitude of the underprediction found in the present study. Other possibilities for the underprediction need to be explored.

The sample in the present investigation was larger and the participants were younger, and apparently more homogeneous, than those in the sample described by Léger et al. (1988) and Léger and Gadoury (1989). We submit that our sample is representative of the population who regularly use the SR test. Léger and Gadoury stated that no age effect was noticed. However, we suggest that homogeneity of running economy should not be assumed throughout adulthood (Sidney and Shepard, 1977; Waters et al., 1983). Consequently, the regression equations presented in the present study are most applicable to the age group used to develop them.

Differences between the present results and those of Léger et al. (1988) and Léger and Gadoury (1989) may be partly explained by the test protocols and gas analysis procedures. Léger et al. (1988) recorded $\dot{V}O_2\text{max}$ by the backward extrapolation technique (Léger et al., 1980). While this technique has been validated, it can only be considered an estimate of actual $\dot{V}O_2\text{max}$. Léger and Gadoury (1989) estimated $\dot{V}O_2\text{max}$ during the last minute of the treadmill test using the Douglas bag method. The results obtained by this procedure cannot be used to infer $\dot{V}O_2\text{max}$ with great confidence, since only one $\dot{V}O_2$ value is obtained. The present protocol, in which gas exchange data were averaged every 20 seconds, is a more sensitive method that can detect a plateau in oxygen consumption over the final workloads. It is possible that in some tests the shorter sampling-time window could result in a higher peak value. However, in this study we found that the criteria for $\dot{V}O_2\text{max}$ (Taylor et al., 1955) were satisfied whether either 20-second or 1-minute averaging was used.

The accuracy of the metabolic measurement system used in this study was verified on a regular basis using a Tissot gasometer and discrete gas analyzers. As stated in the methods, the metabolic cart was calibrated with primary gas standards immediately before each test and calibration was verified immediately following each test. No significant drift (>0.0002 in $F_{E}O_2$ and $F_{E}CO_2$) in the gas analyzers was noted following any test. It is therefore unlikely that the measured $\dot{V}O_2\text{max}$ values reported here could result from spurious data from the metabolic measurement system.

Léger and Gadoury (1989) used a treadmill protocol that involved predominantly walking exercise. The test protocol was described as “The initial load of the test was set at 4.83 km·hr$^{-1}$ (3 miles·hr$^{-1}$) and 0% slope, the load increment was a 5% increase in slope every 3 min till the twelfth minute and 2.5% increase every 2 min till the sixteenth minute (20%), followed by a load increment of 0.4 km·hr$^{-1}$
(0.25 mph) every 2 min” (p. 22). It would appear from the test description that most subjects would have achieved \( \dot{V}O_2 \)max while walking at less than 4 mph and at approximately 20% grade. The present study used a progressive walk/run protocol in which all subjects reached \( \dot{V}O_2 \)max while running. The present protocol typically elicited \( \dot{V}O_2 \)max within 10 to 15 minutes of exercise. This time frame may be optimal for two reasons. First, Thoden (1991) has suggested that the ideal test duration is 8 to 12 minutes; this allows achievement of a plateau in \( \dot{V}O_2 \) while avoiding cessation due to other factors. Second, the present \( \dot{V}O_2 \)max protocol provides a good physiological comparison with the SR test because of similarity in duration and the progression in exercise intensity.

As expected, there was a significant gender difference in measured \( \dot{V}O_2 \)max values, consistent with normative data showing \( \dot{V}O_2 \)max in females to be slightly lower than in males at similar activity levels (McArdle et al., 1994). It is important to note that the men and women in the present study were of similar relative fitness level. The group means for \( \dot{V}O_2 \)max are consistent with normative data for active to very active North American adults as described by McArdle et al. (1994).

While the \( \dot{V}O_2 \)max values for the men and women were different, the slopes of the regression equations developed were not. The similarity between slopes implies similar running economy between groups with increasing speed. Shephard (1991) reported that the energy cost of performing heavy work may be up to 25% lower in women. If so, this may offer a reasonable explanation for the slightly lower Y-intercept value for women (25.1 vs. 28.8 for men, see Figure 2). Léger and Gadoury (1989) found no gender effect in their sample. However, the relatively small number of women (\( n = 24 \)) in their investigation may have made it difficult to detect a gender difference.

The standard errors of measurement for the present regression equations (4.07 for males and 3.64 for females) appear to be somewhat better than the value of 4.7 reported by both Léger et al. (1988) and Léger and Gadoury (1989). Previously, \( \dot{V}O_2 \)max was predicted using the last full 1-minute stage completed (Léger et al., 1988; Léger and Gadoury, 1989). Subjects falling just short of completing a full 1-minute stage would be rounded down to the previous stage completed. The equations developed from the present study allow credit for 30 seconds when participants fall short of completing a full stage. This increased precision should help to detect changes in fitness when assessing athletes before and after a period of training. Consequently, the present regression equations also allow for greater sensitivity in the prediction of \( \dot{V}O_2 \)max.

The 20-m multi-stage shuttle run test is a well-designed evaluation of maximal work capacity that allows prediction of \( \dot{V}O_2 \)max from several regression equations. We found that these equations systematically underpredicted \( \dot{V}O_2 \)max for young adults of both sexes, which is consistent with other reports. The relationship between SR test performance and \( \dot{V}O_2 \)max is different between males and females, and consequently the use of gender-distinct equations is strongly advised.

**Acknowledgments**

The authors express their gratitude to the subjects who volunteered for this study. This research was funded by Alberta Environment, Land and Forest Service. M.K. Stickland is supported by a scholarship from the Natural Science and Engineering Research Council of Canada.
References


Received March 27, 2001; accepted in final form May 27, 2002.