



## ORIGINAL ARTICLE

# $\dot{V}O_{2\max}$ assessment in athletes: A thorough method comparison study between Yo-Yo test and direct measurement<sup>☆</sup>



Aldo F. Longo<sup>\*</sup>, Gustavo D. Aquilino, Marcelo L. Cardey, Néstor A. Lentini

Exercise Physiology Laboratory, National Center of High Performance Athletics (CeNARD), Buenos Aires, Argentina

Received 12 April 2016; accepted 11 July 2016

Available online 15 September 2016

## KEYWORDS

Maximal oxygen uptake;  
Direct measurement and Yo-Yo endurance test;  
Precision of the limits of agreement

## Abstract

**Introduction:** Although different studies have reported limits of agreement in assessing  $\dot{V}O_{2\max}$  between the Yo-Yo test and the direct measurement, the precision of these limits in general has not been considered. The aim of this study was to examine the extent of agreement in the assessment of  $\dot{V}O_{2\max}$  in athletes between the Yo-Yo endurance test (YET) and the direct measurement (DM), and to quantify the precision of the estimated limits of agreement.

**Material and methods:** Data were obtained from a group of 11 male field hockey players (Age =  $22.2 \pm 3.6$  yrs, BMI =  $22.1 \pm 2.4$  kg m<sup>-2</sup>). DM was performed using an incremental treadmill running test. YET level 1 was used for indirect estimation of  $\dot{V}O_{2\max}$ . Bland–Altman analysis was employed for assessing agreement between the two methods. The acceptable 95% limits of agreement were set *a priori* at  $\pm 5$  ml kg<sup>-1</sup> min<sup>-1</sup>.

**Results:** A non-statistically significant bias was observed between YET and DM (50.78 vs. 51.09 ml kg<sup>-1</sup> min<sup>-1</sup>,  $P > .05$ ). The estimates of the 95% limits of agreement were  $-4.34$  and  $3.72$  ml kg<sup>-1</sup> min<sup>-1</sup>. And the 95% confidence intervals for these limits were from  $-6.78$  to  $-1.90$  ml kg<sup>-1</sup> min<sup>-1</sup>, and from  $1.29$  to  $6.16$  ml kg<sup>-1</sup> min<sup>-1</sup>, respectively. The difference between methods did not appear to be correlated to the magnitude of measurement.

**Conclusions:** A reasonably good agreement was found between YET and DM. However, the large variance of the limits of agreement due to the small sample size means these results should be treated with caution.

© 2016 Consell Català de l'Esport. Generalitat de Catalunya. Published by Elsevier España, S.L.U. All rights reserved.

<sup>☆</sup> *Note.* It is reported that a preliminary summary of this work was published given its presentation at “57th Annual Meeting and inaugural World Congress on Exercise is Medicine of the American College of Sports Medicine”. The mentioned event took place in “Baltimore Convention Center” in Baltimore, Maryland, between 1 and 5 June 2010. The summary was published in “Medicine and Science in Sports and Exercise, Volume 42: 5 Supplement”.

<sup>\*</sup> Corresponding author at: Exercise Physiology Laboratory, National Center of High Performance Athletics (CeNARD) Miguel B. Sánchez 1050, Buenos Aires (Postal Code: 1429) Argentina.

E-mail address: [alongo@deportes.gob.ar](mailto:alongo@deportes.gob.ar) (A.F. Longo).

**PALABRAS CLAVE**

Consumo máximo de oxígeno; Medición directa y Yo-Yo endurance test; Precisión de los límites de concordancia

## Valoración del $\dot{V}O_{2\max}$ en atletas: Un riguroso estudio de comparación de métodos entre el Yo-Yo test y la medición directa

**Resumen**

**Introducción:** Si bien diferentes estudios han reportado límites de concordancia en la valoración del  $\dot{V}O_{2\max}$  entre el Yo-Yo test y la medición directa, la precisión de estos límites en general no ha sido considerada. El objetivo de este estudio fue examinar el grado de concordancia en la valoración del  $\dot{V}O_{2\max}$  en atletas entre el Yo-Yo endurance test (YET) y la medición directa (DM), y cuantificar la precisión de los límites de concordancia estimados.

**Material y métodos:** Los datos fueron obtenidos de un grupo de 11 jugadores varones de hockey sobre césped (edad =  $22,2 \pm 3,6$  años, BMI =  $22,1 \pm 2,4$  kg m<sup>-2</sup>). La DM fue realizada usando un test de carrera incremental en cinta rodante. El YET nivel 1 fue usado para la estimación indirecta del  $\dot{V}O_{2\max}$ . El análisis de Bland-Altman fue empleado para valorar la concordancia entre los 2 métodos. Los límites aceptables de concordancia del 95% fueron fijados *a priori* en  $\pm 5$  ml kg<sup>-1</sup> min<sup>-1</sup>.

**Resultados:** Un sesgo estadísticamente no significativo fue observado entre YET y DM (50,78 vs. 51,09 ml kg<sup>-1</sup> min<sup>-1</sup>,  $p > 0,05$ ). Las estimaciones de los límites de concordancia del 95% fueron  $-4,34$  y  $3,72$  ml kg<sup>-1</sup> min<sup>-1</sup>. Y los intervalos de confianza del 95% para estos límites fueron, respectivamente, desde  $-6,78$  hasta  $-1,90$  ml kg<sup>-1</sup> min<sup>-1</sup>, y desde  $1,29$  hasta  $6,16$  ml kg<sup>-1</sup> min<sup>-1</sup>. La diferencia entre métodos pareció no estar correlacionada con la magnitud de medición.

**Conclusiones:** Fue encontrada una razonable buena concordancia entre YET y DM. Sin embargo, la gran varianza de los límites de concordancia debido al pequeño tamaño muestral hace necesario considerar este resultado con precaución.

© 2016 Consell Català de l'Esport. Generalitat de Catalunya. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

**Introduction**

The maximal oxygen uptake ( $\dot{V}O_{2\max}$ ) is a basic measure of physical fitness for athletes, mainly in the cases where the performance is influenced by the aerobic power. The direct measurement is considered the gold standard to assess it, but this is rather complex and expensive. In consequence, a variety of indirect tests have been developed to estimate  $\dot{V}O_{2\max}$ , such as Åstrand-Rhyming 6-minute cycle ergometer test,<sup>1</sup> Balke 15-minute run,<sup>2</sup> Cooper 12-minute run,<sup>3</sup> Bruce treadmill test,<sup>4</sup> the multistage 20-metre shuttle run tests of Léger and Lambert<sup>5</sup> and Léger et al.,<sup>6</sup> 1-mile track jog<sup>7</sup> and Yo-Yo endurance test.<sup>8</sup> The Yo-Yo endurance test (YET) is a continuous multistage field test that is widely implemented to estimate  $\dot{V}O_{2\max}$  as an alternative to the direct laboratory measurement, because of its specificity, its practical and easy implementation and the simple testing environment required. It is one of the three Yo-Yo tests.<sup>8</sup> The other ones are the Yo-Yo intermittent endurance test (YIET) and the Yo-Yo intermittent recovery test (YIRT), the last one also providing an estimation of  $\dot{V}O_{2\max}$ .<sup>9</sup>

The Bland-Altman approach<sup>10-12</sup> has been extensively applied to compare methods of measurement in diverse research areas. It allows to assess the degree of agreement between two measurement techniques, to determine if they can be used interchangeably. However, to method comparison studies are sometimes analysed inappropriately, by comparing the mean responses, by using correlation coefficients or by testing the slope of the linear regression between the methods.<sup>10</sup>

Different studies have compared the direct  $\dot{V}O_{2\max}$  measurement obtained through a treadmill exercise test with the outcomes of the Yo-Yo tests in athletes.<sup>13-26</sup> In most cases the population under study was football players. In these works, the direct assessment of  $\dot{V}O_{2\max}$  was compared with the performance (distance covered) in the Yo-Yo tests, or with the indirect estimation of  $\dot{V}O_{2\max}$  obtained from YET or YIRT. Linear correlations were examined or mean responses were contrasted. Some of the studies that compared the  $\dot{V}O_{2\max}$  estimation given by the Yo-Yo test (YET or YIRT) with the direct measurement also reported Bland-Altman limits of agreement.<sup>15,21,25</sup> However, confidence intervals for these limits were not included. Furthermore, it would be ideal to define the acceptable limits of agreement *a priori*. Although it may be difficult for physiologic variables, an attempt should be made; the opinion of experts may be used. Without *a priori* setting of limits of agreement, widely discrepant limits may be selected.<sup>27</sup>

The uncertainty due to sampling error should be considered not only when estimating the difference between methods (bias), but also when estimating the limits of agreement.<sup>12</sup> This issue is of crucial importance when the sample size is small. The Bland-Altman approach provides statistical methodology to quantify the precision of the estimated limits of agreement. Nevertheless, no method comparison studies were found between direct measurement of  $\dot{V}O_{2\max}$  and the indirect estimation by the Yo-Yo test (YET or YIRT) that include confidence intervals for the limits of agreement. This study aimed to examine the extent of agreement in the assessment of  $\dot{V}O_{2\max}$  in athletes between

**Table 1** Kinanthropometric characteristics of the athletes.

	Mean $\pm$ SD
Age (yrs)	22.19 $\pm$ 3.60
Weight (kg)	68.79 $\pm$ 9.38
Height (m)	1.76 $\pm$ 0.08
BMI ( $\text{kg m}^{-2}$ )	22.11 $\pm$ 2.44
Fat mass (%)	11.03 $\pm$ 5.34
Muscle mass (%)	47.43 $\pm$ 4.07

the Yo-Yo endurance test and the direct measurement performed in an incremental treadmill running test, and to quantify the precision of the estimated limits of agreement.

## Material and methods

### Participants

Data were obtained from a group of 11 competitive male field hockey players. All subjects or their guardians gave consent to participate in the study after being informed of the aims and procedures. The study was conducted based on the ethical principles of the Declaration of Helsinki of the World Medical Association. Each participant underwent a physical examination before the testing sessions. Table 1 displays a statistical summary of the kinanthropometric characteristics of the athletes. Age was calculated in decimal years, by subtracting the date of birth from the date of assessment. Percentages of fat mass and muscle mass were estimated using the four-compartment model based on the strategy of De Rose and Guimaraes.<sup>28</sup> This model was adapted using the simple regression equation for male athletes developed by Withers et al., which was cited by Norton,<sup>29</sup> to estimate body density, and Siri formula<sup>30</sup> to calculate the percentage of fat mass.

### Study design

The present research is a study of simple cross-over design, in order to evaluate agreement in assessing  $\dot{V}O_{2\max}$  in  $\text{ml kg}^{-1} \text{min}^{-1}$  between the Yo-Yo endurance test and the direct measurement; a special focus is on the interval estimation of the limits of agreement. The direct measurement of  $\dot{V}O_{2\max}$  (DM) was conducted in the Exercise Physiology Laboratory of the National Center of High Performance Athletics (CeNARD, Buenos Aires, Argentina), using an incremental running test on a motorized treadmill (Technogym Excite Run 700i; Technogym SpA, Gambetola, Italy). Breath-by-breath data collection was obtained by means of a computerized open-circuit metabolic system (Medgraphics Cardiopulmonary Exercise System CPX/D, Breeze Ex v3.06 software; Medical Graphics Corporation, St. Paul, MN, USA). Increments in treadmill speed of  $1 \text{ km h}^{-1}$  every minute were stipulated until volitional exhaustion (initial speed =  $9 \text{ km h}^{-1}$ ). Heart rate was controlled with a heart rate monitor (Polar 610i; Polar Electro Oy, Kempele, Finland).  $\dot{V}O_2$  plateau (change in  $\dot{V}O_2$  difference less than  $2.1 \text{ ml kg}^{-1} \text{min}^{-1}$  with a further increase in workload) was the primary criterion for the attainment of  $\dot{V}O_{2\max}$ ;

secondary criteria were: respiratory exchange ratio greater than 1.1 and heart rate within  $10 \text{ beats min}^{-1}$  of age-predicted maximum heart rate.<sup>31,32</sup> The Yo-Yo endurance test level 1 was performed indoor at CeNARD, in a sport hall with wooden sprung floor. It consisted of repeated 20-m shuttle runs at a progressively increased speed (initial speed =  $8 \text{ km h}^{-1}$ ), which was controlled by audio beeps from a portable audio player. The test was considered ended when the participant was unable to maintain the currently indicated speed. The final speed level and the number of performed 20-metre distances at this speed level were recorded. The estimation of  $\dot{V}O_{2\max}$  was then obtained using the corresponding nomogram. Both YET and DM took place at the same time of day ( $\pm 2$  hours), to minimize the effects of diurnal biological variation on the results, and were carried out at least 72 hours apart from each other to avoid residual fatigue. The order of the tests was randomly assigned for each subject. Prior to the evaluations, the participants had a 20-minute related warm-up.

### Statistical analysis

The kinanthropometric characteristics of the athletes and the heart rate and respiratory exchange ratio values reached when achieving  $\dot{V}O_{2\max}$  were reported as mean  $\pm$  standard deviation (SD). Summary statistics were produced to describe the  $\dot{V}O_{2\max}$  values obtained by means of YET and DM. Bland-Altman analysis was employed for assessing agreement between the two methods of measurement.<sup>10-12</sup> In accordance to Bland,<sup>33</sup> the acceptable 95% limits of agreement were established on the basis of experience, and were set *a priori* at  $\pm 5 \text{ ml kg}^{-1} \text{min}^{-1}$ . The 95% confidence intervals for the limits of agreement were also computed. In view of the small sample size, they were estimated using the exact expression for the estimator of the variance of the limits of agreement instead of the approximate one<sup>12</sup>:

$$\text{Var}(\bar{d} \pm 1.96s_d) = \left( \frac{1}{n} + \frac{1.96^2}{2(n-1)} \right) s_d^2, \quad (1)$$

where  $\bar{d}$  is the mean difference between methods and  $s_d$  is the standard deviation of the differences. Pearson's and Spearman's correlation tests were implemented to evaluate possible associations of the difference between YET and DM with the magnitude of the measurement. Statistical significance was set at the 0.05 probability level. All analyses were performed using the R software environment, version 3.2.0 (R Core Team, Vienna, Austria).<sup>34</sup>

## Results

The  $\dot{V}O_{2\max}$  values obtained for each individual using the two methods of measurement were quite similar. The  $\dot{V}O_{2\max}$  values estimated by YET were, on average, only 0.6% lower than those determined by DM. Table 2 presents a descriptive summary of the outcomes of YET and DM and of the differences between them. The heart rate and respiratory exchange ratio values reached when achieving  $\dot{V}O_{2\max}$  are summarized in Table 3.

A non-statistically significant bias was found between YET and DM ( $-0.31 \text{ ml kg}^{-1} \text{min}^{-1}$ ; 95% confidence interval:

**Table 2**  $\dot{V}O_{2\max}$  assessment: comparative summary of both methods of measurement.

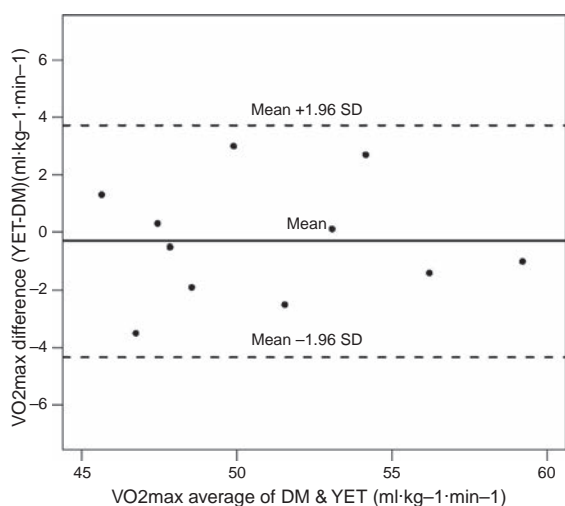
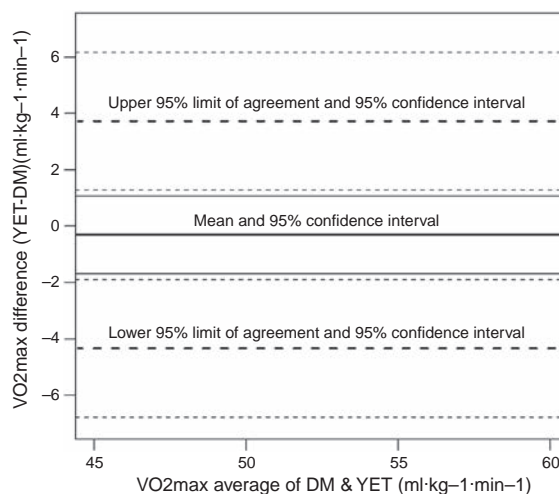
	Mean $\pm$ SD	Minimum	Maximum
Yo-Yo endurance test (YET)	50.78 $\pm$ 4.44	45.00	58.70
Direct measurement (DM)	51.09 $\pm$ 4.41	45.00	59.70
Difference (YET-DM)	-0.31 $\pm$ 2.06	-3.50	3.00

All values expressed in  $\text{ml kg}^{-1} \text{min}^{-1}$ .

**Table 3** Heart rate and respiratory exchange ratio values reached when achieving  $\dot{V}O_{2\max}$ .

	Mean $\pm$ SD
Heart rate ( $\text{beats min}^{-1}$ )	197.9 $\pm$ 8.2
Respiratory exchange ratio	1.1 $\pm$ 0.08

-1.69 to  $1.07 \text{ ml kg}^{-1} \text{min}^{-1}$ ). The estimates of the lower and upper 95% limits of agreement were  $-4.34$  and  $3.72 \text{ ml kg}^{-1} \text{min}^{-1}$ . And the 95% confidence intervals for these limits were, respectively, from  $-6.78$  to  $-1.90 \text{ ml kg}^{-1} \text{min}^{-1}$ , and from  $1.29$  to  $6.16 \text{ ml kg}^{-1} \text{min}^{-1}$ . As previously mentioned, given the small sample size, the confidence intervals for the limits of agreement were constructed using the exact expression for the estimator of the variance of these limits (Eq. (1)). The region determined by the estimated 95% limits of agreement was narrower than the region bounded by the acceptable limits defined *a priori*. However, the total region covered when considering the 95% confidence intervals for the 95% limits of agreement was wider than the *a priori* acceptable limits. The typical Bland-Altman plot with estimated bias and 95% limits of agreement between methods is shown in Fig. 1, while

**Figure 1**  $\dot{V}O_{2\max}$ : Difference between methods against the average measurement with bias and 95% limits of agreement.**Figure 2**  $\dot{V}O_{2\max}$ : Mean and 95% limits of agreement of the difference between methods with 95% confidence intervals.

all the aforementioned results are presented graphically in Fig. 2.

It was observed in the data that neither the difference nor the absolute value of the difference appeared to be correlated to the magnitude of the measurement. In that order: Pearson's  $r=0.02$  (95% confidence interval:  $-0.59$  to  $0.61$ ); Spearman's  $\rho=-0.06$  (95% confidence interval:  $-0.64$  to  $0.56$ ).

## Discussion

The indirect measurement of  $\dot{V}O_{2\max}$  is frequently performed in sport evaluation. Maximal and submaximal effort protocols have been proposed to estimate the maximal aerobic power of an individual.<sup>1-9</sup> Good agreement with the direct measurement should be the main objective of these tests. The Yo-Yo endurance test is a continuous multistage field test broadly used to indirectly assess  $\dot{V}O_{2\max}$ .<sup>15</sup> Its mechanical characteristics make it suitable for athletes participating in sports that involve stop, start and change of direction movement patterns.

Diverse studies have been conducted to compare the outputs provided by the Yo-Yo tests with the  $\dot{V}O_{2\max}$  obtained by direct measurement in an incremental running test on treadmill. The comparisons involved contrasting the  $\dot{V}O_{2\max}$  values or testing the correlation between directly measured  $\dot{V}O_{2\max}$  values and distances covered in the Yo-Yo tests. Some of the studies that compared the  $\dot{V}O_{2\max}$  results of both methods also included the evaluation of agreement. For instance, Metaxas et al.<sup>22</sup> found in male football players ( $n=35$ ) statistically significant differences ( $p<0.05$ ) when comparing the  $\dot{V}O_{2\max}$  estimations given by YET level 1 with the  $\dot{V}O_{2\max}$  values obtained by DM in either a continuous or an intermittent running protocol on a treadmill (YET level 1 values lower than DM values by 11.4% and 13.4%, respectively). Castagna et al.<sup>15</sup> contrasted the  $\dot{V}O_{2\max}$  results of YET level 2 with the ones of DM in male football players ( $n=24$ ), evidencing a non-statistically significant bias ( $p=0.10$ ) of  $1.17 \text{ ml kg}^{-1} \text{min}^{-1}$ , and 95% limits of agreement of  $-5.44$  and  $7.79 \text{ ml kg}^{-1} \text{min}^{-1}$ . Nazarali et al.<sup>23</sup> reported a higher mean for the  $\dot{V}O_{2\max}$  values provided by YIRT level



2 ( $50.8 \text{ ml kg}^{-1} \text{ min}^{-1}$ ) in comparison to the mean of the values determined by DM ( $43.6 \text{ ml kg}^{-1} \text{ min}^{-1}$ ), in a study conducted with female football players ( $n=20$ ). In contrast, Martínez-Lagunas and Hartmann<sup>21</sup> found in female football players ( $n=18$ ) that  $\dot{V}O_{2\max}$  was significantly underestimated ( $p < 0.001$ ) by YIRT level 1 ( $45.2 \text{ ml kg}^{-1} \text{ min}^{-1}$ ) compared to DM ( $55.0 \text{ ml kg}^{-1} \text{ min}^{-1}$ ), expressing the 95% limits of agreement on a percentage scale ( $-31.8\%$  to  $-3.8\%$ ). And Sánchez-Oliva et al.<sup>25</sup> examined the assessments of the two methods in male football players ( $n=15$ ) and also found a statistically significant underestimation of  $\dot{V}O_{2\max}$  using YIRT level 1 ( $51.00 \text{ ml kg}^{-1} \text{ min}^{-1}$ ) in comparison to DM ( $60.85 \text{ ml kg}^{-1} \text{ min}^{-1}$ ), with 95% limits of agreement of 3.37 and  $16.33 \text{ ml kg}^{-1} \text{ min}^{-1}$ .

On the other hand, no research articles were found reporting method or studies by using Bland–Altman analysis between the direct measurement of  $\dot{V}O_{2\max}$  and the indirect estimation by the Yo-Yo test (YET or YIRT) that consider the computation of confidence intervals for the limits of agreement. This is a matter of particular importance in the case of small sample size, which is a frequent situation in studies involving maximal effort testing. The variance of the mean difference and that of the limits of agreement are inversely proportional to the sample size. Therefore, the corresponding confidence intervals are wide when the sample size is small, reflecting the great variation of the differences between methods.

This research presents a thorough method comparison study between the direct assessment of  $\dot{V}O_{2\max}$  in laboratory conditions and the indirect estimation by means of the Yo-Yo endurance test, which is a continuous multistage field test especially useful for athletes of endurance disciplines,<sup>8</sup> but also suitable for sports teams<sup>35</sup>, such as field hockey, football and rugby players. Estimates of bias and 95% limits of agreement between methods are reported. Furthermore, accurate 95% confidence intervals for the limits of agreement are also reported for the current sample size. A small bias was observed between the two techniques, which resulted to be not statistically significant ( $p=0.63$ ). The estimated 95% limits of agreement set at  $\pm 4.03 \text{ ml kg}^{-1} \text{ min}^{-1}$  from the mean difference between them. In the model proposed, for the mean and standard deviation of the differences between methods to be meaningful estimates it must be assumed that they are constant throughout the range of measurement.<sup>12,36,37</sup> It is worth emphasizing that the mean difference and the standard deviation of the differences appeared not to depend on the magnitude of the measurement, which complies with the assumptions underlying the simple 95% limits of agreement method applied. The region determined by the estimated limits of agreement was within the acceptable limits defined *a priori*. However, these estimates have a large variance due to the small sample size, so this last result has to be considered with caution. Taking into account the 95% confidence intervals for the 95% limits of agreement, the total region extended beyond  $\pm 6 \text{ ml kg}^{-1} \text{ min}^{-1}$ , exceeding the acceptable 95% limits of agreement established *a priori*.

The confidence intervals for the mean difference and for the limits of agreement provide a measure of precision and for the point estimates, based on sampling error. The variances of the estimators of these quantities are crucially affected by the sample size. The smaller is the number of observations

for the assessment of the difference between the methods, the wider are the confidence intervals, and the higher is the level of uncertainty. Moreover, although the Bland–Altman approach provides the methodology for constructing limits of agreement, it does not state whether they are acceptable or not acceptable. The acceptable limits should be defined *a priori*, based on specific needs and goals.

## Limitations

The small sample size is a limitation of this study. The 95% confidence intervals for the 95% limits of agreement given the current sample size are  $\pm 1.19$  units of the standard deviation of the differences between measurements of the two methods. Therefore, it would be advisable to reproduce the present analysis with larger sample sizes to increase the precision of the estimates produced. Furthermore, it would be relevant to simultaneously carry out a comparison of the repeatability of each method by collecting replicated data, because the repeatabilities of two methods of measurement restrict the amount of agreement that is possible.<sup>12</sup>

## Conclusions

The point estimates of the 95% limits of agreement were within  $\pm 5 \text{ ml kg}^{-1} \text{ min}^{-1}$ , satisfying the *a priori* requirement based on experience. Thus, a reasonable good agreement was found between the Yo-Yo endurance test and the direct laboratory measurement for the assessment of  $\dot{V}O_{2\max}$  in athletes. According to this fact, the two methods can be used interchangeably. Nevertheless, the need for caution in interpreting this result solely is further emphasized by the finding that the region covered when considering the 95% confidence intervals for the 95% limits of agreement was wider than the acceptable region of agreement defined *a priori*. Hence, although the estimated 95% limits of agreement do not discredit the Yo-Yo endurance test as an alternative method to measure  $\dot{V}O_{2\max}$  in athletes, the present study is limited in precision because of the small sample size and does not allow definitive conclusions.

## Conflict of interests

Authors declare that they do not have any conflict of interests.

## Acknowledgments

The authors wish to express special thanks to Mariela B. Arangio, Enrique D. Balardini, Claudio A. Gillone, Cristina Perez and Enrique O. Prada for their technical assistance.

## References

1. Åstrand PO, Ryhming I. A nomogram for calculation of aerobic capacity from pulse rate during submaximal work. *J Appl Physiol*. 1954;7:218–21.
2. Balke B. A simple field test for the assessment of physical fitness. Oklahoma City, OK: Civil Aeromedical Research Institute, Federal Aviation Agency; 1963; Report No. 63-6.

3. Cooper KH. A means of assessing maximal oxygen intake. Correlation between field and treadmill testing. *J Am Med Assoc.* 1968;203:201–4.
4. Bruce RA. Multi-stage treadmill test of maximal and sub maximal exercise. In: American Heart Association, editor. *Exercise testing and training of apparently healthy individuals: a handbook for physicians.* New York, NY: American Heart Association; 1972. p. 32–4.
5. Léger LA, Lambert J. A maximal multistage 20-m shuttle run test to predict  $\dot{V}O_{2\max}$ . *Eur J Appl Physiol Occup Physiol.* 1982;49:1–12.
6. Léger LA, Mercier D, Gadoury C, Lambert J. The multistage 20 metre shuttle run test for aerobic fitness. *J Sport Sci.* 1988;6:93–101.
7. George JD, Vehrs P, Allsen PE, Fellingham GW, Fisher AG.  $\dot{V}O_{2\max}$  estimation from a submaximal 1-mile track jog for fit college-age individuals. *Med Sci Sports Exerc.* 1993;25:401–6.
8. Bangsbo J. Yo-Yo tests. Copenhagen, Denmark: August Krogh Institute; 1996.
9. Bangsbo J, Marcelllo Iaija F, Krusturup P. The Yo-Yo intermittent recovery test: a useful tool for evaluation of physical performance in intermittent sports. *Sports Med.* 2008;38:37–51.
10. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *The Statistician.* 1983;32:307–17.
11. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;i:307–10.
12. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res.* 1999;8:135–60.
13. Aziz AR, Tan FHY, Teh KC. A pilot study comparing two field tests with the treadmill in soccer players. *J Sports Sci Med.* 2005;4:105–12.
14. Bradley PS, Bendiksen M, Dellal A, Mohr M, Wilkie A, Datson N, et al. The application of the Yo-Yo intermittent endurance level 2 test to elite female soccer populations. *Scand J Med Sci Sports.* 2014;24:43–54.
15. Castagna C, Impellizzeri FM, Chamari K, Carlomagno D, Rampinini E. Aerobic fitness and Yo-Yo continuous and intermittent tests performances in soccer players: a correlation study. *J Strength Cond Res.* 2006;20:320–5.
16. Castagna C, Impellizzeri FM, Rampinini E, D'Ottavio S, Manzi V. The Yo-Yo intermittent recovery test in basketball players. *J Sci Med Sport.* 2008;11:202–8.
17. Higham DG, Pyne DB, Anson JM, Eddy A. Physiological, anthropometric, and performance characteristics of rugby sevens players. *Int J Sports Phys Perf.* 2013;8:19–27.
18. Karakoç B, Akalan C, Alemdaroğlu U, Arslan E. The relationship between the Yo-Yo tests, anaerobic performance and aerobic performance in young soccer players. *J Hum Kinet.* 2012;35:81–8.
19. Krusturup P, Mohr M, Amstrup T, Rysgaard T, Johansen J, Steensberg A, et al. Yo-Yo intermittent recovery test: physiological response, reliability, and validity. *Med Sci Sports Exerc.* 2003;35:697–705.
20. Krusturup P, Mohr M, Nybo L, Majgaard Jensen J, Jung Nielsen J, Bangsbo J. The Yo-Yo IR2 test: physiological response, reliability, and application to elite soccer. *Med Sci Sports Exerc.* 2006;38:1666–73.
21. Martínez-Lagunas V, Hartmann U. Validity of the Yo-Yo intermittent recovery test level 1 for direct measurement or indirect estimation of maximal oxygen uptake in female soccer players. *Int J Sports Physiol Perform.* 2014;9:825–31.
22. Metaxas TI, Koutlianos NA, Kouidi EJ, Deligiannis AP. Comparative study of field and laboratory tests for the evaluation of aerobic capacity in soccer players. *J Strength Cond Res.* 2005;19:79–84.
23. Nazaryo Y, Rajab H, Aliabadi F. The relationship between laboratory, Yo-Yo, and Hoff tests in determining aerobic capacity of players of the National women's soccer team. *Ann Appl Sport Sci.* 2013;1:57–66.
24. Rebelo A, Brito J, Seabra A, Oliveira J, Krusturup P. Physical match performance of youth football players in relation to physical capacity. *Eur J Sport Sci.* 2014;14 Suppl 1: S148–56.
25. Sánchez-Oliva D, Santalla A, Candela JM, Leo FM, García-Calvo T. Analysis of the relationship between Yo-Yo test and maximum oxygen uptake in young football players. *Int J Sport Sci.* 2014;10:180–93.
26. Thomas A, Dawson B, Goodman C. The Yo-Yo test: reliability and association with a 20-m shuttle run and  $VO_{2\max}$ . *Int J Sports Physiol Perform.* 2006;1:137–49.
27. Mantha S, Roizen MF, Fleisher LA, Thisted R, Foss J. Comparing methods of clinical measurement: reporting standards for Bland and Altman analysis. *Anesth Analg.* 2000;90: 593–602.
28. De Rose EH, Guimaraes AGS. A model for optimization of somatotype in young athletes. In: Ostyn M, Beunen G, Simons J, editors. *Kinanthropometry II.* Baltimore, MD: University Park Press; 1980. p. 222.
29. Norton K. Anthropometric estimation of body fat. In: Norton K, Olds K, editors. *Anthropometrica: a textbook of body measurement for sports and health courses.* Sydney: University of New South Wales Press Ltd; 1996. p. 171–98.
30. Siri WE. Body composition from fluid spaces and density: analysis of methods. In: Brozek J, Henzchel A, editors. *Techniques for measuring body composition.* Washington, DC: National Academy of Sciences; 1961. p. 224–44.
31. Howley ET, Bassett DR Jr, Welch HG. Criteria for maximal oxygen uptake: review and commentary. *Med Sci Sports Exerc.* 1995;27:1292–301.
32. O'Connor FG, Kunar MT, Deuster PA. Exercise physiology for graded exercise testing: a primer for the primary care clinician. In: Evans CH, White RD, editors. *Exercise testing for primary care and sports medicine physicians.* New York, NY: Springer; 2009. p. 3–21.
33. Bland M. Interpreting the limits of agreement: do I have good or bad agreement? [Internet page]. [Last updated 20 March 2009; consulted 23 April 2009]. Available at <http://www.ucl.ac.uk/~u293f/medstat/interlim.htm>
34. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015. Available at: <http://www.R-project.org/>
35. Wood RJ. Yo-Yo endurance test [Internet page]. [Consulted 11 May 2009]. Available at <http://www.topendsports.com/testing/tests/yo-yo-endurance.htm>
36. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet.* 1995;346:1085–7.
37. Bland JM, Altman DG. Measurement error proportional to the mean. *Br Med J.* 1996;313:106.