Journal of Science and Medicine in Sport xxx (xxxx)

Contents lists available at ScienceDirect

Journal of Science and Medicine in Sport

journal homepage: www.elsevier.com/locate/jsams

Original research

Criterion-related validity and reliability of the 2-km walk test and the 20-m shuttle run test in adults: The role of sex, age and physical activity level

Nuria Marín-Jiménez [a,b], Sandra Sánchez-Parente [a,b], Pablo Expósito-Carrillo [a,b], José Jiménez-Iglesias [a,b,c], Inmaculada C. Álvarez-Gallardo [a,b], Magdalena Cuenca-García [a,b,*], José Castro-Piñero [a,b]

[a] GALENO Research Group, Department of Physical Education, Faculty of Education Sciences, University of Cadiz, Spain
[b] Instituto de Investigación e Innovación Biomédica de Cádiz (INiBICA), Spain
[c] Sport Science Department Cadiz C.F., Spain

ARTICLE INFO

Article history:
Received 26 September 2022
Received in revised form 30 January 2023
Accepted 7 March 2023
Available online xxxx

Keywords:
Fields tests
Cardiorespiratory fitness test
Validity
Reproducibility
Adults

ABSTRACT

Objectives: To analyze the criterion-related validity and the reliability of fitness field tests for evaluating cardiorespiratory fitness in adults, by sex, age, and physical activity level.
Design: Cross-sectional.
Methods: During 3 weeks, sociodemographic, anthropometric measurements, a treadmill maximal test, the 2-km walk test, and the 20-m SRT were performed in 410 adults aged 18–64 years. Measured and estimated VO_{2max} (by Oja's and Leger's equations) were analyzed.
Results: Measured VO_{2max} was associated with estimated VO_{2max} by the 2-km walk test and 20-m SRT ($r = 0.784$ and $r = 0.875$, respectively; both $p < 0.01$). Bland–Altman analysis showed a mean difference of -0.30 ml* kg^{-1} * min^{-1} ($p < 0.001$, $d = -0.141$) in the 2-km walk test, and 0.86 ml* kg^{-1} * min^{-1} ($p = 0.051$) in the 20-m SRT. Significant mean differences between test and retest were found in the time to complete the 2-km walk test (-1.48 ± 0.51 s, $p = 0.004$, $d = -0.014$) and in the final stage reached in the 20-m SRT (0.04 ± 0.01, $p = 0.002$, $d = 0.015$). Non-significant differences were found between test and retest in the estimated VO_{2max} by Oja's (-0.29 ± 0.20 ml* kg^{-1} * min^{-1}, $p > 0.05$) and Leger's eqs. (0.03 ± 0.04 ml* kg^{-1} * min^{-1}, $p > 0.05$). Moreover, both test results and estimated VO_{2max} equations showed a high test–retest reliability.
Conclusions: Both tests were valid and reliable for evaluating cardiorespiratory fitness in adults aged 18–64 years, regardless of sex, age, and physical activity level.

Practical implications

- The findings of this study suggest that the 20-m SRT has some advantages, such as reduced psychological stress, and the possibility of better regulating the pacing strategy compared with the 2-km walk test.
- Therefore, when existing time or space constraints, the 20-m SRT could be proposed as an ideal tool to evaluate cardiorespiratory fitness in the adult population.
- Alternatively, the 2-km walk test is more suitable for adults who are unable to run.

1. Introduction

Physical fitness is considered a powerful health marker in the adult population, especially cardiorespiratory fitness, and muscular strength.[1,2] Cardiorespiratory fitness has been inversely associated with reduced risk of diseases, such as cardiovascular disease,[1,3] obesity, diabetes,[2] different types of cancer,[4] and is a predictor of all-cause of mortality.[1,3–5] Likewise, high levels of cardiorespiratory fitness have been associated with a decrease in the risk of suffering from mental conditions such as anxiety, panic, and depression.[5] Furthermore, cardiorespiratory fitness seems to be the most determining factor of life expectancy.[6] Consequently, cardiorespiratory fitness assessment is an important tool for prevention and health diagnosis in the adult population.[7]

Laboratory testing is the most objective and accurate method to assess physical fitness. However, due to costly, sophisticated instruments and qualified technicians required, and time constraints, their use is

* Corresponding author.
E-mail address: Magdalena.cuenca@uca.es (M. Cuenca-García).

https://doi.org/10.1016/j.jsams.2023.03.005

limited in sports clubs, schools, or population-based studies. In these settings, field-based fitness tests could be a useful and reasonable alternative, since they are relatively safe and time-efficient, involve minimal, low cost equipment and can be easily administered to a large number of people simultaneously.[8]

The validity and reliability of field-based fitness tests need to be considered when deciding which field-based test to use.[9] Criterion-related validity refers to the extent to which a field-based test of a physical fitness component correlates with the criterion measure (i.e., the gold standard).[9] A test is considered reliable when a participant performs a test on two or more occasions under the same conditions and close proximity in time and obtains similar results.[9]

Concerning its field-based test assessment in adults, the 2-km walk test and 20-m shuttle run test (20-m SRT) are the most used to assess cardiorespiratory fitness.[7,10] A recent systematic review, where the validity of existing field test for physical fitness assessment was evaluated, concluded that these tests are valid in the adult population (using Oja's equation in the 2-km walk test, and Leger's equation in the 20-m SRT, through VO_{2max} calculation).[7] Likewise, another recent systematic review, where the reliability of existing field-test for physical fitness assessment was evaluated, enlightened that the 20-m SRT was strongly reliable in young adults, however, the reliability of the 2-km walk test was limited in adults aged 30–64 years.[10]

Nevertheless, most of the studies that analyzed the criterion-related validity and reliability of these two field tests in adults had a small sample or presented a lack of balanced representation of sex or the full adult age range (i.e., 18–64 years). Furthermore, these studies have not taken into account the physical activity level of the participants, when it is known that the level of physical activity can influence the validity of these tests.[7] Finally, no study has evaluated which of these field-based tests is more valid and reliable taking into account sex, age, and physical activity level.

Therefore, the aim of the present study was to analyze the criterion-related validity and the reliability of the 2-km walk test and the 20-m SRT for evaluating cardiorespiratory fitness in the adult population, according to sex, age, and physical activity level.

2. Materials and methods

The present study is part of a national project: the ADULT-FIT study, whose main aim was to propose a field-based physical fitness-test battery related to health based on their criterion-validity, predictive validity, reliability, feasibility, and safety for use in adults.

Briefly, a total of 410 adults aged 18–64 years were recruited through leaflets, local newspapers, and social media from Cadiz (Spain). The total sample was homogeneously distributed by sex, age (18–34 years, 35–49 years, and 50–64 years), and physical activity level (non-active and active).

The inclusion criteria for this study were: (i) age: adults (18–64 years old); (ii) not having a physical or mental illness that prevents you from doing physical activity; (iii) intention to carry out all the tests that make up the study and; (iv) able to read and understand the informed consent as well as the object of the study. The exclusion criteria for this study were: (i) acute or terminal illness; (ii) myocardial infarction three months before starting the study; (iii) unstable cardiovascular disease; (iv) medical prescription that prevents the performance of the tests and; (v) injury or circumstance that makes it impossible to carry out the tests correctly.

All interested volunteers provided written informed consent to participate in the present study.

After providing written informed consent and being informed of the protocol to be carried out, they signed the “Physical Activity Readiness Questionnaire” (PAR-Q) questionnaire to detect possible contraindications to the practice of physical exercise, and a questionnaire to determine the physical activity level of the participants.

Participants were tested in 3 sessions during 3 weeks (one per week). In the first week, sociodemographic, anthropometric measurements, and a maximum treadmill test were carried out. In the second and third weeks, the 2-km walk test and the 20-m SRT were carried out, one per week (test–retest) in the same conditions as before.

Before field-based testing sessions, all participants completed a standardized 10-minute warm-up. All the participants received comprehensive instructions for the tests and were encouraged to do their best in each test. Participants were instructed to rest 24 h before evaluations and to maintain their eating and hydration habits.

Participants were initially classified as active/non-active when following/not following World Health Organization recommendations for adults (https://www.who.int/). The following self-reported question was asked: *how many days (in a typical week) do you practice physical activity/exercise or some sport, of at least moderate intensity, lasting at least 50 min per day*?

Height, weight, triceps and subscapular skinfolds, and hip and waist circumferences were measured using the protocol described by the International Society for the Advancement of Kinanthropometry (ISAK).[11] For the neck circumference, the protocol established by the Center for Disease Control and Prevention was followed.[12] Measurements were always performed by the same trained evaluator (to avoid intra-evaluator variability), of the same sex as the participant.

All measurements were collected with bare feet, in light sports clothing, and with a 3-h fast. Height was measured using a TANITA HR001 portable height rod (Tanita®, Illinois, USA; sensitivity, 1 mm). The margin of error that was established to make a third measurement was 1 cm. Weight was measured using an OMRON BF-400 electronic scale (Omron Healthcare Europe BV, Hoofddorp, The Netherlands; sensitivity, 100 g). The established margin of error by which a third measurement should be made was a difference of 1 kg. Body mass index (BMI) was calculated as weight (kg) divided by squared height (m^2).

Triceps and subscapular skinfolds were measured using the Harpenden Skinfold Caliper (Holtain, Dyfed, United Kingdom; range, 0–80 mm; sensitivity, 0.2 mm), and hip, waist and neck circumferences were assessed using the tape measure using SECA 201 (Seca Int, Hamburg, Germany; range, 0–205 cm; sensitivity, 0.1 cm). The margin of error for a third measurement was 1 mm for skinfolds and 1 cm for circumferences.

The percentage of body fat mass (%BF) and lean mass (kg) were determined by bioimpedance Tanita MC 780-P MA (Tanita Co., Guangzhou, China), according to the protocol described by the National Institute of Health (NIH).[13] For its correct evaluation, the participants were asked about their level of hydration.

Participants completed an incremental cardiopulmonary exercise test (CPET) on a treadmill (Lode Valiant, Groningen, Netherlands) for the determination of VO_{2max} through indirect calorimetry (Jaeger MasterScreen CPX®㈱, CareFusion, San Diego, USA). VO_{2max} was recorded as absolute values ($ml^* min^{-1}$) and relative per kilograms of body weight ($ml^* kg^{-1} * min^{-1}$). Heart rate and peak respiratory exchange ratio (RER) were also collected.

Three different protocols, performed until volitional exhaustion, were used given the large heterogeneity of our study sample (i.e., physical activity/fitness level and age). The protocol was selected for each participant to reach their limit of tolerance in approximately 10–12 min. In all protocols, the test involved a walking warm-up, an incremental, and a recovery phase.

For the selection of the CPET protocol for each participant, the following criteria was used: (i) participants with a low physical activity level or who could not run on a treadmill, performed the Balke protocol, which maintained a constant speed of 4.8 km/h and increased the slope by 2.5% every 2 min.[14] Those participants with no previous experience on a treadmill (both walking or running) underwent a brief familiarization period of 1–3 min, to feel confident enough with the test; (ii) participants with average physical activity/fitness warmed up at 4.8 km/h during 2 min, starting the test at 6 km/h and increasing by 1 km/h per minute; (iii) participants who presented a greater physical activity/fitness level, warmed up at 6 km/h during 2 min, starting the test at 8 km/h, increasing 1 km/h every minute. After exhaustion was attained, a 5 min recovery phase was performed in all the aforementioned protocols.

The main criteria used to determine the maximality of a test was the attainment of a VO_2 plateau. In our study we used the criteria of defining the plateau as a difference of ≤150 ml/min between consecutive stages. To this aim, we compared the 2-final time-averaged periods of 30 s of the test. Thus, when the difference in VO_2 between the final and the preceding 30-s period was ≤150 ml/min the plateau was established and the final VO_2 value was considered a real VO_{2max}.[15]

When the plateau criteria was not met, at least 3 of the following secondary criteria had to be met to establish that the test was maximal[16]: (i) volitional exhaustion or the incapacity to maintain the treadmill speed despite verbal encouragement, (ii) heart rate within 10 bpm of the maximal age predicted, (iii) RER ≥ 1.1, and (iv) rated perceived exertion (RPE) of ≥7 using the Borg scale from 0 to 10.

2-km walk test: The test consisted of walking 2-km at maximum speed in the shortest possible time.[17] The participants were instructed to walk at their maximum speed from the beginning to the end of the test in a 100-meter rectangular circuit. At the end of the test, the total time spent on the test (minutes and seconds) was recorded. Moreover, final heart rate was recorded using the activity bracelet, validated in adults, Xiaomi Mi Band 4 (Xiaomi Inc., Beijing, China).[18] VO_{2max} estimations were based on regression equations, established by Oja et al.[17]

20-m SRT: The test is an incremental intermittent running test between two separate lines at a distance of 20 m.[19] The initial speed, marked by acoustic signals, is 8.5 km/h, increasing 0.5 km/h every minute; so, the time that the participants have to cover the distance of 20 m decreases over time. The test ends when the participants reach physical exhaustion or are unable to follow the set pace, or when they cannot cover the distance in the set time during two consecutive acoustic signals. Results were registered as fully completed stages. VO_{2max} estimations were based on regression equations, established by Leger et al.[19]

Descriptive sample values and cardiorespiratory fitness tests are presented as mean ± SD. One-way analysis of variance (ANOVA) was performed to assess significant differences between age groups and t-test analysis for independent sample for sex and physical activity level groups.

Criterion-related validity: Bivariate correlations and simple linear regression were used to evaluate the agreement between cardiorespiratory fitness laboratory and field-based tests. When significant, the strength of the correlations was classified as follows: 0.00–0.25, very low; 0.26–0.49, low; 0.50–0.69, moderate; 0.70–0.89, high and; 0.90–1.00, very high.[20]

Subsequently, the mean difference and the 95% limits of agreement [95% LoA (mean difference ± 1.96 SD of the difference)] were calculated using the Bland–Altman method[21] to analyze the agreement between measured and estimated VO_{2max}, whose difference was calculated using an ANOVA test for repeated measures. Where appropriate, Cohen *d* was computed to quantify the magnitude of the effect size. Cohen *d* values of 0.8, 0.5, and 0.2 represented large, medium, and small effect sizes, respectively.[22]

Finally, in order to develop a more precise equation for the sample, stepwise linear regression model was used to estimated VO_{2max} [relative VO_{2max} ($ml^* \ kg^{-1} \ ^* \ min^{-1}$)]. To do this, the variables that presented a higher correlation were sequentially added, which were sex, age, physical activity level and %BF. Additional analyses were performed including BMI, or sum tricipital + subscapular skinfold, waist circumference, hip circumference and lean mass instead of %BF.

Reliability: To investigate the reliability of the 2-km walk test and the 20-m SRT, we compared test and retest (hereafter called T1 and T2), through t-test and intraclass correlation coefficient (ICC). ICC is commonly used to describe *relative reliability* (i.e., the consistency of measurements on individuals in a group relative to others). An ICC <0.8 were considered insufficient, values between 0.8 and 0.9 were considered moderate and values >0.9 were considered high.[23]

Since any reliability study should not be based on a single statistic method, we also examined the differences between T1 and T2 using different error measures. Generally, the lower the error value, the lower the dispersion between T1 and T2 measurements.

The sum of squared errors (SSE) was calculated as follows:

$$\mathrm{SSE} = \sum_{i=1}^{N} (yi - \hat{y})^2$$

where n is the cases to evaluate the error measurements, $\hat{y}$ is the T2, and y is the T1.

The mean sum of squared errors (MSE):

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N} (yi - \hat{y})^2$$

The root mean sum of squared errors (RMSE) was calculated by converting MSE into domain units by taking the root square:

$$\mathrm{RMSE} = \sqrt{MSE}$$

The percentage error was calculated as follows:

$$\%\mathrm{Error} = \frac{RMSE}{y_{max} - y_{min}} \times 100$$

The *absolute reliability* (consistency of repeated measurements for individuals) was analyzed by calculating the standard error of measurement (SEM) as percentage of the mean value of the measurements. The SEM quantifies the precision of individual scores in a test, and it is not influenced by variability among individuals (i.e., is considered a fixed characteristic of any measure, regardless of the sample of participants under investigation). A value ≤15% is considered acceptable.[24]

%SEM = mean of the difference scores between 2 trials × 100/mean of the first trial.

The coefficient of variation (CV) as follow:

$$\%\mathrm{CV} = \frac{\delta}{\overline{X}} \times 100$$

The CV method provides useful information in the presence of heteroscedasticity (assumes that greatest T1 and T2 variation occurs in individuals scoring the highest values in the test). A CV ≤ 10% was considered as acceptable reliability.[25]

The standard error of estimate was calculated as follows:

$$\mathrm{SEE} = SD\hat{y}\sqrt{\left(1 - R^2 y\hat{y}\right)}$$

Finally, Bland–Altman plots were used to evaluate the reproducibility[21] of the field-based cardiorespiratory fitness tests, whose difference was calculated using an ANOVA test for repeated measures. Where appropriate, Cohen *d* was computed to quantify the magnitude of the difference between T1 and T2.

We also examined the difference and the magnitude of the measurement (i.e., heteroscedasticity) by conducting regression analysis.

We conducted the analyses for the whole sample, as well as separately by sex, age groups, and physical activity level, for all criterion-related and reliability analysis.

All the analyses were performed using the Statistical Package for Social Sciences (IBM SPSS Statistics for Windows, version 26.0; Armonk, NY) and the level of significance was set at $p < 0.05$.

3. Results

The final sample size was composed of 410 adults aged 18–64 years (49.5% females). The descriptive characteristics of the participants, distributed by sex, age and, physical activity level are shown in Table 1. The mean age of the sample was 42 (±13.06) years old. Overall, significant differences were found according to sex, age, and physical activity level. Regarding sex, males presented higher anthropometric values than females (all, $p < 0.001$), except for %BF and tricipital skinfold, being higher in females (both, $p < 0.001$). Males performed faster the 2-km walk test and completed more stages in the 20-m SRT than females ($p < 0.001$); measured and estimated VO_{2max} was also higher in males than females ($p < 0.001$). Regarding age, the 50–64 years old group presented higher BMI, %BF, waist and neck circumference and subscapular skinfold than the 18–34 years old group (all, $p < 0.01$), and higher waist circumference than the 35–49 years old group ($p < 0.01$); the 35–49 years old group presented higher BMI, %BF and waist circumference than the 18–34 years old group (all, $p < 0.01$); the 50–64 years old group performed slower the 2-km walk test, completed less stages in the 20-m SRT and had measured and estimated VO_{2max} than their younger counterparts (all, $p < 0.001$). Regarding physical activity level, active participants shower lower anthropometric values (all, $p < 0.01$), expect for neck circumference ($p > 0.05$), and better performance in cardiorespiratory variables than non-active ones (all, $p < 0.001$).

Criterion-related validity: Bivariate correlation analysis between measured VO_{2max} ($ml^* kg^{-1} * min^{-1}$) with estimated VO_{2max} by 2-km walk test and 20-m SRT, and anthropometric variables, distributed by whole sample, sex, age groups, and physical activity level are displayed in Supplementary Table 1. In the whole sample, measured VO_{2max} was associated with estimated VO_{2max} by 2-km walk test ($r = 0.784$, $p < 0.01$), and 20-m SRT ($r = 0.875$, $p < 0.01$). Sex, age, physical activity level, and all anthropometric variables were associated with measured and estimated VO_{2max} ($r = -0.148$ to -0.751, all $p < 0.001$). %BF, tricipital skinfold and sum triceps + subscapular skinfold had the strongest association with measured VO_{2max} ($r = -0.750$, $r = -0.648$, $r = -0.656$, respectively; all, $p < 0.001$), with estimated VO_{2max} by 2-km walk test ($r = -0.730$, $r = -0.699$, $r = -0.723$, respectively; all, $p < 0.001$), and with estimated VO_{2max} by 20-m SRT ($r = -0.751$, $r = -0.650$, $r = -0.636$, respectively; all, $p < 0.001$). Similar results were found when the sample was distributed by sex, age, and physical activity level.

Fig. 1 shows the scatterplot of relationship between measured with estimated VO_{2max} by Oja's equation and time in the 2-km walk test, and relationship between measured with estimated VO_{2max} by Leger's equation, and final stage in the 20-m SRT. The measured VO_{2max} was associated with the estimated VO_{2max} by Oja's equation ($R^2 = 0.614$, $p < 0.01$; SEE = 5.027) and time in the 2-km walk test ($R^2 = 0.429$, $p < 0.01$; SEE = 6.117), and with estimated VO_{2max} by Leger's equation ($R^2 = 0.766$, $p < 0.01$; SEE = 3.917) and final stage in the 20-m SRT ($R^2 = 0.774$, $p < 0.01$; SEE = 3.865).

The association between measured VO_{2max} with estimated VO_{2max} by Oja's equation remained the same when the sample was distributed by sex, age, and physical activity level. The association between measured VO_{2max} with time in the 2-km walk test was slightly higher in females ($R^2 = 0.448$, $p < 0.01$) than in males ($R^2 = 0.300$, $p < 0.01$); in the 50–64 years group ($R^2 = 0.517$, all, $p < 0.01$) than in the younger ones ($R^2 = 0.346$ and $R^2 = 0.397$, respectively; both, $p < 0.01$); and in non-active ($R^2 = 0.449$, $p < 0.01$) than in active participants ($R^2 = 0.286$, $p < 0.01$) (Supplementary Fig. 1).

The association between measured VO_{2max} with estimated VO_{2max} by Leger's equation and final stage in the 20-m SRT remained the

Table 1
Descriptive characteristics of the participants, stratified by sex, age and, physical activity level.

	All (n = 410)	Sex		Age groups			Physical activity levels	
		Females (n = 203)	Males (n = 207)	18–34 yr (n = 136)	35–49 yr (n = 131)	50–64 yr (n = 143)	Non-active (n = 195)	Active (n = 215)
Age (years)	41.86 (13.06)	41.48 (12.84)	42.24 (13.28)	26.22 (0.36)	42.49 (0.36)	56.17 (0.35)	43.18 (13.10)*	40.62 (12.95)
Basal heart rate (bpm)	69.51 (12.09)	72.25 (12.20)***	66.82 (11.39)	71.83 (13.46)$^{a\wedge\wedge}$	67.39 (11.11)	69.24 (11.25)	73.51 (12.23)***	65.88 (10.80)
Maximal heart rate (bpm)	173.29 (13.77)	171.8 (13.0)*	174.66 (14.37)	182.47 (12.20)$^{a\wedge\wedge\wedge}$	173.68 (9.17)$^{b\wedge\wedge\wedge}$	164.41 (12.91)$^{c\wedge\wedge\wedge}$	172.79 (15.25)	173.73 (12.42)
Weight (kg)	71.16 (15.49)	62.06 (9.36)***	80.13 (15.12)	68.80 (16.24)	72.19 (16.29)	72.48 (13.77)	71.99 (16.82)	70.32 (14.14)
Height (cm)	168.25 (9.19)	161.38 (5.98)***	175.00 (6.36)	168.46 (0.79)	168.42 (0.80)	167.89 (0.77)	167.33 (9.17)	169.02 (9.12)
Body mass index	24.99 (4.10)	23.85 (3.45)***	26.12 (4.38)	24.07 (4.27)$^{a\wedge}$	25.32 (4.48)	25.56 (3.38)$^{c\wedge\wedge}$	25.55 (4.62)**	24.47 (3.50)
Body fat (%)	24.85 (7.47)	28.64 (6.30)***	21.11 (6.62)	22.77 (7.69)$^{a\wedge}$	24.99 (7.08)	26.70 (7.15)$^{c\wedge\wedge\wedge}$	27.38 (7.31)***	22.57 (6.89)
Lean mass (kg)	50.37 (10.58)	41.56 (3.92)***	59.06 (7.41)	49.85 (10.13)	51.09 (10.81)	50.22 (10.83)	49.10 (10.72)*	51.45 (10.31)
Waist circumference (cm)	83.14 (12.66)	76.90 (9.25)***	89.26 (12.58)	78.11 (11.68)$^{a\wedge\wedge}$	83.28 (13.11)$^{b\wedge\wedge}$	87.80 (11.34)$^{c\wedge\wedge\wedge}$	85.54 (14.21)***	80.91 (10.64)
Hip circumference (cm)	99.38 (8.69)	98.80 (7.27)	99.95 (9.87)	98.69 (9.75)	100.30 (9.19)	99.19 (6.96)	101.13 (9.10)***	97.77 (8.00)
Neck circumference (cm)	35.30 (3.83)	32.30 (2.00)***	38.24 (2.79)	34.40 (3.65)	35.44 (3.91)	36.01 (3.80)$^{c\wedge\wedge\wedge}$	35.31 (4.12)	35.27 (3.57)
Tricipital skinfold (mm)	16.92 (7.48)	19.69 (6.42)***	14.23 (7.46)	16.13 (0.64)	17.03 (0.65)	17.58 (0.63)	19.39 (7.67)***	14.72 (6.56)
Subscapular skinfold (mm)	19.50 (10.27)	20.33 (9.19)	18.68 (11.19)	17.37 (11.72)	19.46 (9.83)	21.57 (8.72)$^{c\wedge\wedge}$	22.74 (10.99)***	16.57 (8.62)
2-km walk test								
Total time (min:seg)	16.63 (1.80)	17.26 (1.66)***	16.00 (1.72)	16.50 (1.66)	16.14 (1.71)$^{b\wedge\wedge\wedge}$	17.16 (1.88)$^{c\wedge\wedge}$	17.32 (1.83)***	16.01 (1.54)
Final heart rate (bpm)	149.79 (21.36)	151.55 (20.05)***	148.05 (22.48)	151.89 (22.99)	151.38 (21.81)	146.46 (18.99)	152.54 (21.59)*	147.36 (20.90)
VO_{2max} ($ml^* kg^{-1} * min^{-1}$)	36.58 (10.14)	33.09 (6.43)***	40.06 (11.84)	40.38 (10.67)	37.97 (9.24)$^{b\wedge\wedge\wedge}$	31.89 (8.47)$^{c\wedge\wedge\wedge}$	32.65 (9.79)***	40.14 (9.12)
20-meter shuttle run test								
Final stage	5.10 (2.82)	3.71 (2.17)***	6.48 (2.72)	6.29 (2.71)	5.65 (2.69)$^{b\wedge\wedge\wedge}$	3.51 (2.26)$^{c\wedge\wedge\wedge}$	3.87 (2.31)***	6.20 (2.79)
VO_{2max} ($ml^* kg^{-1} * min^{-1}$)	35.42 (8.51)	31.30 (6.57)***	39.50 (8.25)	39.04 (8.26)	37.11 (8.04)$^{b\wedge\wedge\wedge}$	30.60 (6.76)$^{c\wedge\wedge\wedge}$	31.74 (6.98)***	38.70 (8.42)
Maximal treadmill test								
VO_{2max} absolute ($ml^* min^{-1}$)	2557.15 (720.28)	1999.87 (367.17)***	3097.71 (545.33)	2703.81 (744.90)	2655.09 (698.22)$^{b\wedge\wedge\wedge}$	2329.57 (720.28)$^{c\wedge\wedge\wedge}$	2299.89 (633.97)***	2784.21 (716.77)
VO_{2max} relative ($ml^* kg^{-1} * min^{-1}$)	36.13 (8.14)	32.78 (6.59)***	39.39 (8.20)	39.48 (7.86)$^{a\wedge}$	36.94 (7.69)$^{b\wedge\wedge\wedge}$	32.23 (7.17)$^{c\wedge\wedge\wedge}$	32.17 (6.74)***	39.67 (7.67)
RER final	1.20 (0.08)	1.18 (0.08)***	1.21 (0.08)	1.22 (0.07)	1.20 (0.07)$^{b\wedge\wedge\wedge}$	1.17 (0.08)$^{c\wedge\wedge\wedge}$	1.20 (0.08)	1.20 (0.08)

Differences between sex, and between physical activity level: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; differences between age groups: ^{a} = 18–34 years and 35–49 years, ^{b} = 35–49 years and 50–64 years, ^{c} = 18–35 years and 50–64 years ($^{\wedge}p < 0.05$; $^{\wedge\wedge}p < 0.01$, $p^{\wedge\wedge\wedge} < 0.001$).
Results are expressed as mean ± SD.
VO_{2max} estimated in the 2-km walk test by Oja's equation; VO_{2max} estimated in the 20-meter shuttle run test by Leger's equation.
Difference between sex, and physical activity levels measured with an independent t-test, difference between age groups measured with one-way repeated measures analysis of variance (ANOVA).

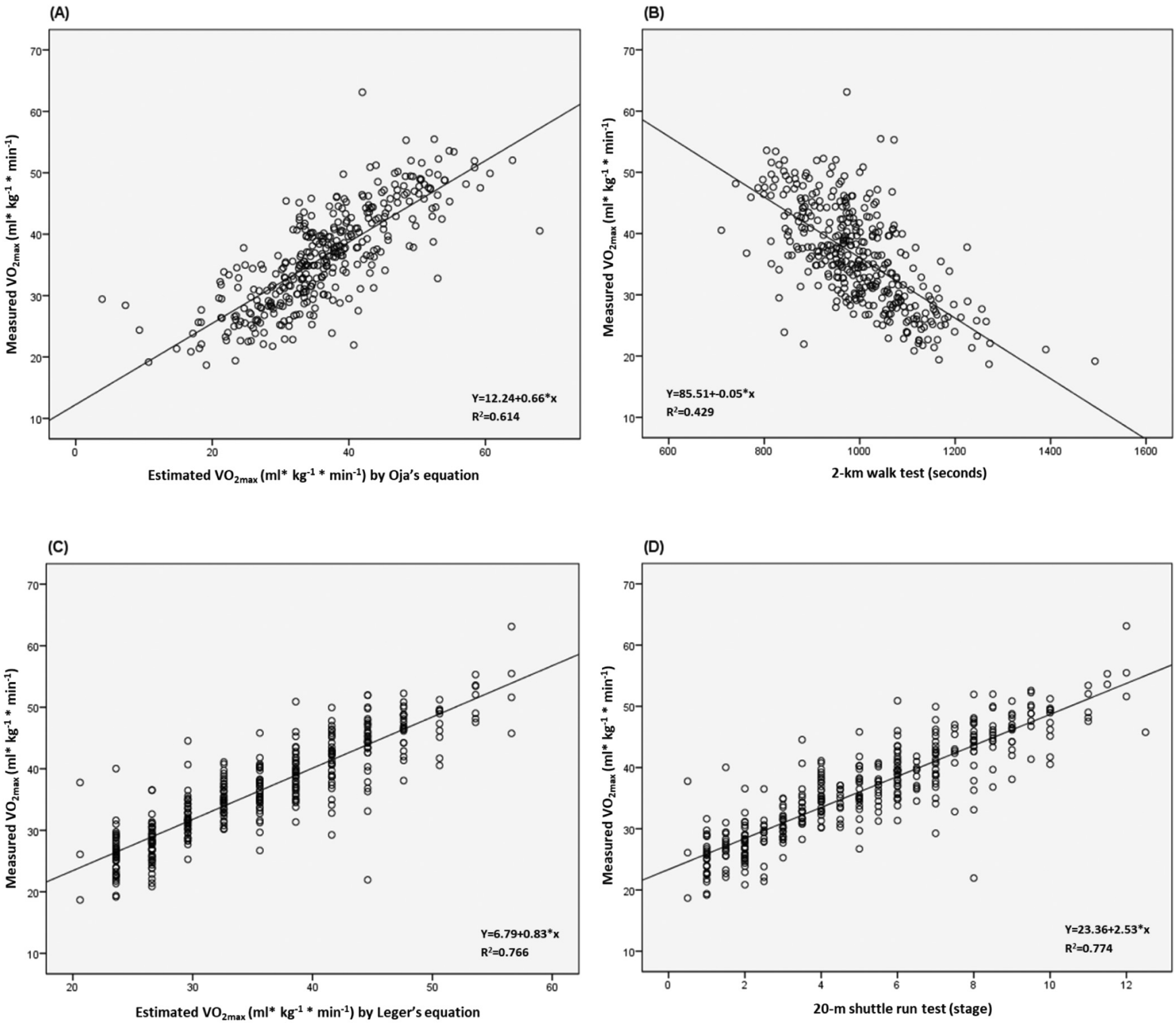

Fig. 1. Scatterplot of relationship between measured with estimated VO_{2max} (ml* kg^{-1} * min^{-1}) by Oja's equation (estimated from the 2-km walk test) and Leger's equation (estimated from the 20-m shuttle run test); and relationship between measured with total time in the 2-km walk test and final stage in the 20-m shuttle run test.
(A), Relationship between measured and estimated VO_{2max} (ml* kg^{-1} * min^{-1}) by Oja's equation; (B), relationship between measured VO_{2max} (ml* kg^{-1} * min^{-1}) and total time (sec) in the 2-km walk test; (C), relationship between measured and estimated VO_{2max} (ml* kg^{-1} * min^{-1}) by Leger's equation; (D), relationship between measured VO_{2max} (ml* kg^{-1} * min^{-1}) and final stage in the 20-m shuttle run test.
R^2 indicates lineal coefficient of determination.

same when the sample was distributed by sex, age, and physical activity level (Supplementary Fig. 2).

Fig. 2 shows the Bland–Altman difference plot between the measured and estimated VO_{2max}, by Oja's and Leger's equations. It can be observed that the difference was nearly 0 for both equations. The difference between measured and estimated VO_{2max} by Oja's equation was −0.30 ml* kg^{-1} * min^{-1} (95% LoA = −12.86 to 12.26, $p < 0.001$), and between measured and estimated VO_{2max} by Leger's equation was 0.86 ml* kg^{-1} * min^{-1} (95% LoA = −7.30 to 9.02, $p = 0.051$). Heteroscedasticity was observed between the measured and estimated VO_{2max} difference with the measured and estimated VO_{2max} mean by Oja's equation ($R^2 = 0.117$). The effect size (Cohen *d*) of the mean differences between measured and estimated VO_{2max} by Oja's equation was −0.141.

The differences between the measured and estimated VO_{2max}, by Oja's equation remained the same when the sample was distributed by sex and age groups. When the sample was distributed by physical activity level, the difference was lower in the non-active [−0.19 ml* kg^{-1} * min^{-1} (95% LoA = −13.45 to 13.06, $p < 0.001$; $R^2 = 0.249$)] than in the active participants [−3.36 ml* kg^{-1} * min^{-1} (95% LoA = −15.31 to 8.59, $p < 0.001$; $R^2 = 0.064$)] (Supplementary Fig. 3).

The difference between the measured and estimated VO_{2max}, by Leger's equation was lower in males [0.08 ml* kg^{-1} * min^{-1} (95% LoA = −8.92 to 9.08, $p = 0.525$; $R^2 = 0.002$)] than in females [1.65 ml* kg^{-1} * min^{-1} (95% LoA = −5.24 to 8.54, $p = 0.617$; $R^2 = 0.001$)]. It was also lower in the 18–34 years group [0.47 ml* kg^{-1} * min^{-1} (95% LoA = −7.49 to 8.43, $p = 0.116$; $R^2 = 0.020$)] and the 35–49 years group [0.16 ml* kg^{-1} * min^{-1} (95% LoA = −8.88 to 9.21, $p = 0.375$; $R^2 = 0.007$)] than in the 50–64 years group [1.79 ml* kg^{-1} * min^{-1} (95% LoA = −5.44 to 9.03, $p = 0.137$; $R^2 = 0.017$)] (Supplementary Fig. 4).

Table 2 shows the stepwise lineal regression analysis predicting the VO_{2max}. The 2-km walk test showed that total time represented the 42% of explained variance for measured VO_{2max} (SEE = 6.136 ml* kg^{-1} * min^{-1}, $p < 0.001$). When heart rate was added, the explained

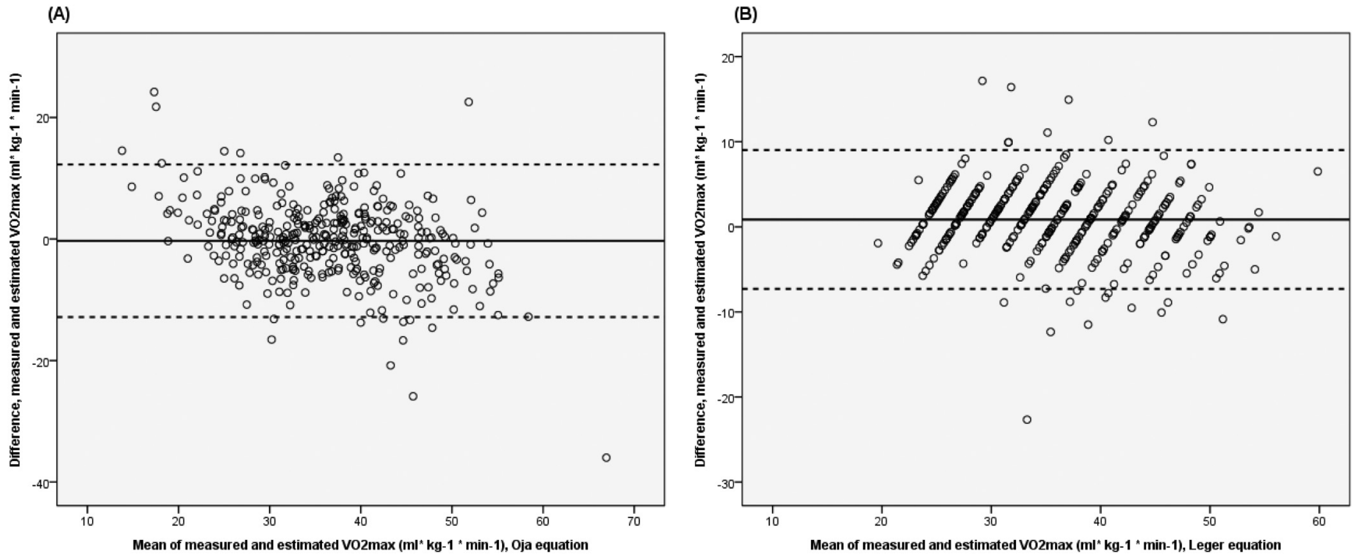

Fig. 2. Bland–Altman plot of agreement between measured and estimated VO_{2max} (ml* kg^{-1} * min^{-1}) by Oja's equation (estimated from the 2-km walk test), and Leger's equation (estimated from the 20-m shuttle run test).
(A), Agreement between measured and estimated VO_{2max} (ml* kg^{-1} * min^{-1}) by Oja's equation; (B), agreement between measured and estimated VO_{2max} (ml* kg^{-1} * min^{-1}) by Leger's equation. The central line represents the mean differences between measured and estimated VO_{2max}. The upper and lower dotted lines represent the upper and lower 95% limits of agreement (mean differences ± 1.96 SD of the differences), respectively.

variance increased until 47%. Finally, when age, physical activity level and %BF were included, the explained variance increased until 70% (SEE = 4.408 ml* kg^{-1} * min^{-1}, $p < 0.001$).

The same results were observed when replacing %BF by BMI, or sum tricipital + subscapular skinfold, waist circumference, hip circumference and lean mass in the prediction equations (*data not shown*).

The 20-m SRT showed that the final stage represented the 78% of explained variance for measured VO_{2max} (SEE = 3.841 ml* kg^{-1} * min^{-1}, $p < 0.001$). When physical activity and %BF mass was included, the explained variance only increased an additional 2% (SEE = 3.603 ml* kg^{-1} * min^{-1}, $p < 0.001$).

The same results were observed when replacing %BF by BMI, or sum tricipital + subscapular skinfold, waist circumference, hip circumference and lean mass in the prediction equations (*data not shown*).

Reliability: Test–retest reliability of 2-km walk test and 20-m SRT is shown in Table 3. Significant mean differences between T1 and T2 were found in the time to complete the 2-km walk test (−1.48 ± 0.51 s, p = 0.004) and in the final stage reached in the 20-m SRT (0.04 ± 0.01, p = 0.002). The effect size (Cohen *d*) of the mean differences was −0.014 and 0.002, respectively. Non-significant differences were found between T1 and T2 in the estimated VO_{2max} by Oja's (−0.29 ± 0.20 ml* kg^{-1} * min^{-1}, $p > 0.05$) and Leger's eqs. (0.03 ± 0.04 ml* kg^{-1} * min^{-1}, $p > 0.05$). The ICCs reported a high reproducibility, ranging from 0.95 to 0.99 (all, $p < 0.001$) in both tests. All the analyzed error measurements showed low values (RMSE = 0.38–10.09; %CV = 0.71–7.73; SEE = 0.25–9.88). Supplementary Table 2 shows test–retest reliability of 2-km walk test and 20-m SRT, distributed by sex, age, and physical activity levels. Overall, similar results to those of the whole sample were found.

Fig. 3 shows the Bland–Altman difference plot between the 2-km walk test and 20-m SRT. The systematic error was nearly 0 for all the cases: estimated VO_{2max} by Oja's equation [−0.29 ml* kg^{-1} * min^{-1} (95% LoA = −8.13 to 7.54, p = 0.034)], time to complete the 2-km walk test [−1.48 s (95% LoA = −21.07 to 18.10, p = 0.041)], estimated VO_{2max} by Leger's eq. [0.03 ml* kg^{-1} * min^{-1} (95% LoA = −1.51 to 1.58, p = 0.086)], and final stage reached in the 20-m SRT [0.04 (95% LoA = −0.45 to 0.53, p = 0.172)]. Heteroscedasticity (T1-T2 variability) was observed in the estimated VO_{2max} by Oja's equation ($R^2 = 0.012$), and time to complete the 2-km walk test ($R^2 = 0.011$).

Supplementary Fig. 4 displayed the results analyzed by sex, age, and physical activity level. Overall, similar results to those of the whole sample were found.

4. Discussion

The aim of the present study was to analyze the criterion-related validity and the reliability of the 2-km walk test and the 20-m SRT for evaluating cardiorespiratory fitness in adult population, according to sex, age, and physical activity level. The results showed that both tests are valid and reliable.

Our study included a homogeneously distributed sample by sex, age, and physical activity level, to analyze whether the criterion-related validity and reliability of the 2-km walk test and the 20-m SRT are dependent of these variables. Overall, we found no significant differences. Hence, the criterion-related validity and reliability of both tests were not determined by sex, age, or physical activity level.

The main results showed a high association between laboratory and field-based cardiorespiratory fitness tests, and between measured and estimated VO_{2max} by Oja's[17] ($R^2 = 0.61$, $p < 0.01$) and Leger's[19] ($R^2 = 0.77$, $p < 0.01$) equations. In fact, Bland–Altman analysis also indicated that the 2-km walk and the 20-m SRT test are valid to estimate cardiorespiratory fitness by Oja's and Leger's equations, showing a mean of differences close to 0 and narrow LoA, regardless of sex, age, and physical activity level.

The 2-km walk test is considered a user-friendly submaximal cardiorespiratory fitness test, since it allows the assessment of people with low physical fitness level or who are unable to run. In fact, the Oja's equation[17] is valid to assess VO_{2max}, in adults aged 20–65 years (R^2 = 0.73 to 0.75, $p < 0.05$),[17] with overweight/obesity (R^2 = 0.56 to 0.59, $p < 0.05$),[26] and with low or moderate fitness levels (R^2 = 0.30 to 0.64, $p < 0.05$), but not in adults with a high fitness level ($R^2 = 0.27$, $p < 0.05$).[27]

Oja et al.[17] found that the total time performed in the 2-km walk test was highly correlated with measured VO_{2max} in females (r = 0.74, p < 0.001), but moderately in males (r = 0.58, p < 0.001). These results are very similar to those found in our sample (r = 0.67 and r = 0.51 for females and males, respectively; both, p < 0.001). Moreover, in our

Table 2
Stepwise linear regression model predicting VO_{2max} by 2-km walk test and, by 20-meter shuttle run test.

Model	Independent variables	β	p value	r	R^2	R^2 change	SEE
2-km walk test							
1	2-km total time	−0.652	<**0.001**	0.652	0.423	0.425	6.136
2	2-km total time	−0.700	<**0.001**	0.685	0.466	0.044	5.903
	2-km heart rate	−0.216	<**0.001**				
3	2-km total time	−0.625	<**0.001**	0.703	0.489	0.025	5.772
	2-km heart rate	−0.176	<**0.001**				
	Sex	0.175	<**0.001**				
4	2-km total time	−0.559	<**0.001**	0.777	0.600	0.110	5.111
	2-km heart rate	−0.224	<**0.001**				
	Sex	0.200	<**0.001**				
	Age	−0.344	<**0.001**				
5	2-km total time	−0.486	<**0.001**	0.795	0.626	0.027	4.940
	2-km heart rate	−0.175	<**0.001**				
	Sex	0.205	<**0.001**				
	Age	−0.331	<**0.001**				
	Physical activity level	0.182	<**0.001**				
6	2-km total time	−0.313	<**0.001**	0.841	0.702	0.076	4.408
	2-km heart rate	−0.113	**0.001**				
	Sex	0.065	0.065				
	Age	−0.239	<**0.001**				
	Physical activity level	0.158	<**0.001**				
	% Body fat	−0.403	<**0.001**				
20-m shuttle run test							
1	Final stage	0.881	<**0.001**	0.881	0.776	0.777	3.841
2	Final stage	0.893	<**0.001**	0.881	0.776	0.000	3.843
	Sex	−0.023	0.423				
3	Final stage	0.911	<**0.001**	0.882	0.776	0.001	3.843
	Sex	−0.032	0.285				
	Age	−0.028	0.333				
4	Final stage	0.848	<**0.001**	0.888	0.786	0.0011	3.753
	Sex	−0.017	0.570				
	Age	−0.009	0.746				
	Physical activity level	0.116	<**0.001**				
5	Final stage	0.709	<**0.001**	0.897	0.803	0.017	3.603
	Sex	−0.053	0.069				
	Age	0.001	0.985				
	Physical activity level	0.110	<**0.001**				
	% Body fat	−0.206	<**0.001**				

β, Standardized regression coefficient; r, correlation coefficients; R^2, adjusted coefficients of determination; SEE, standard error of estimate. Bold values denote statistical significance at the $p < 0.05$ level.

study, that association was also higher in adults aged 50–64 years ($r = 0.71$, $p < 0.001$) and non-active participants ($r = 0.66$, $p < 0.001$).

In the prediction model, Oja et al. found that the total time, heart rate, age and weight predicted 66–76% (SEE = 6.2 to 3.0) of the variance in VO_{2max}, and 73–75% (SEE = 3.3 to 5.1) when replacing weight by BMI.[17] In this sense, we tried to develop a more accurate equation to estimate VO_{2max} considering total time, heart rate, sex, age, physical activity level, and anthropometric variables. The total time represented the 42% of explained variance, reaching 47% when heart rate was added. Finally, the explained variance increased until 70% when total time, heart rate, age, physical activity level and %BF were included. We also replaced %BF by BMI, sum tricipital + subscapular skinfold, waist circumference, hip circumference and lean mass, obtaining similar results. Based on this, our results did not improve the prediction equation proposed by Oja et al.[17] Accordingly, the Oja's equation seems more feasible to estimate VO_{2max}, since it includes a lower number of prediction variables.

Finally, Bland–Altman plots support the criterion-related validity of the 2-km walk test, with differences nearly 0 and narrow LoA, especially for non-active adults [−0.19 ml* kg^{-1} * min^{-1} (95% LoA = −13.45 to 13.06, $p < 0.001$), $d = -0.016$]. One explanation could be that the cardiovascular system of active people is less able to be stressed enough by regular walking to produce an accurate VO_{2max} prediction (i.e., underprediction).[28] Moreover, the heteroscedasticity analyses indicate that the higher VO_{2max} (i.e., the fitter) the worse the degree of agreement in the VO_{2max} prediction.

On the other hand, the 20-m SRT is an incremental maximal cardiorespiratory test, which has been found valid to estimate VO_{2max} by Leger's equation[19] in adults aged 18–64 years ($R^2 = 0.81$, $p < 0.05$). In this sense, we obtained similar results ($R^2 = 0.78$, $p < 0.01$) when testing the Leger's equation in our study sample. These results remained the same after analyzing them by sex, age and physical activity levels. Consequently, these variables seem not to affect the criterion-related validity of the Leger's equation.

There were different attempts to improve the Leger's equation.[19] In fact, we also tried to develop a more accurate equation to estimate VO_{2max} considering final stage, sex, age, physical activity, and anthropometric variables. We found that the final stage represented the 78% of explained variance for measured VO_{2max} (SEE = 3.841 ml* kg^{-1} * min^{-1}). When physical activity, was included, the explained variance only increased 1%. Our final model included final stage, physical activity level and %BF, explaining the 80% of explained variance for measured VO_{2max}, being still lower than those reported by Leger's equation (i.e., 81% of explained variance).[19] Furthermore, replacing %BF by BMI, or sum tricipital + subscapular skinfold, waist circumference, hip circumference and lean mass, yielded similar results. Hence, the Leger's equation seemed to be the most precise and feasible, since only final stage is needed.

Table 3
Test–retest reliability of 2-km walk test and 20-m shuttle run test in the whole sample.

	T1§	T2§	Intertrial difference (T2-T1)	p value	Cohen's *d*	ICC (95% CI)**	SSE	MSE	RMSE	% Error	% SEM	% CV	SEE
All													
2-km walk test (sec)	997.70 ± 108.22	994.81 ± 104.80	−1.48 ± 0.51	**0.004**	−0.014	0.99 (0.99–0.99)	38,479.00	101.80	10.09	1.29	0.15	0.71	9.88
2-km walk test (ml* kg^{-1} * min^{-1})	36.58 ± 10.14	36.43 ± 9.62	−0.29 ± 0.20	0.206		0.95 (0.94–0.96)	6042.61	16.02	4.00	5.02	0.79	7.73	3.80
20-m SRT (stage)	5.10 ± 2.82	5.15 ± 2.85	0.04 ± 0.01	**0.002**	0.015	0.99 (0.99–0.99)	54.50	0.14	0.38	3.16	0.98	3.45	0.25
20-m SRT (ml* kg^{-1} * min^{-1})	35. 42 ± 8.51	35.42 ± 8.49	0.03 ± 0.04	0.433		0.99 (0.99–0.99)	234.00	0.62	0.78	2.18	0.09	1.57	0.78

2-km walk test is expressed as total time to complete the test (sec), and as estimated VO_{2max} (ml* kg^{-1} * min^{-1}) by Oja's equation; 20-m shuttle run test is expressed as final stage reached, and as estimated VO_{2max} (ml* kg^{-1} * min^{-1}) by Leger's equation.
T2-T1 refers to retest (trial 2) minus test (trial 1). Values are displayed as mean ± SD.
ICC, intraclass correlation coefficients; CI, confident interval; SSE, sum of squared errors; MSE, mean sum of squared errors; RMSE, root mean sum of squared errors; %Error, percentage error; %SEM, standard error of measurement; %CV, percentage coefficient of variation; SEE, standard error of estimate.
§ T1 refers to test (trial 1) and T2 to retest (trial 2).
** All the ICCs were significant at $p<0.001$.

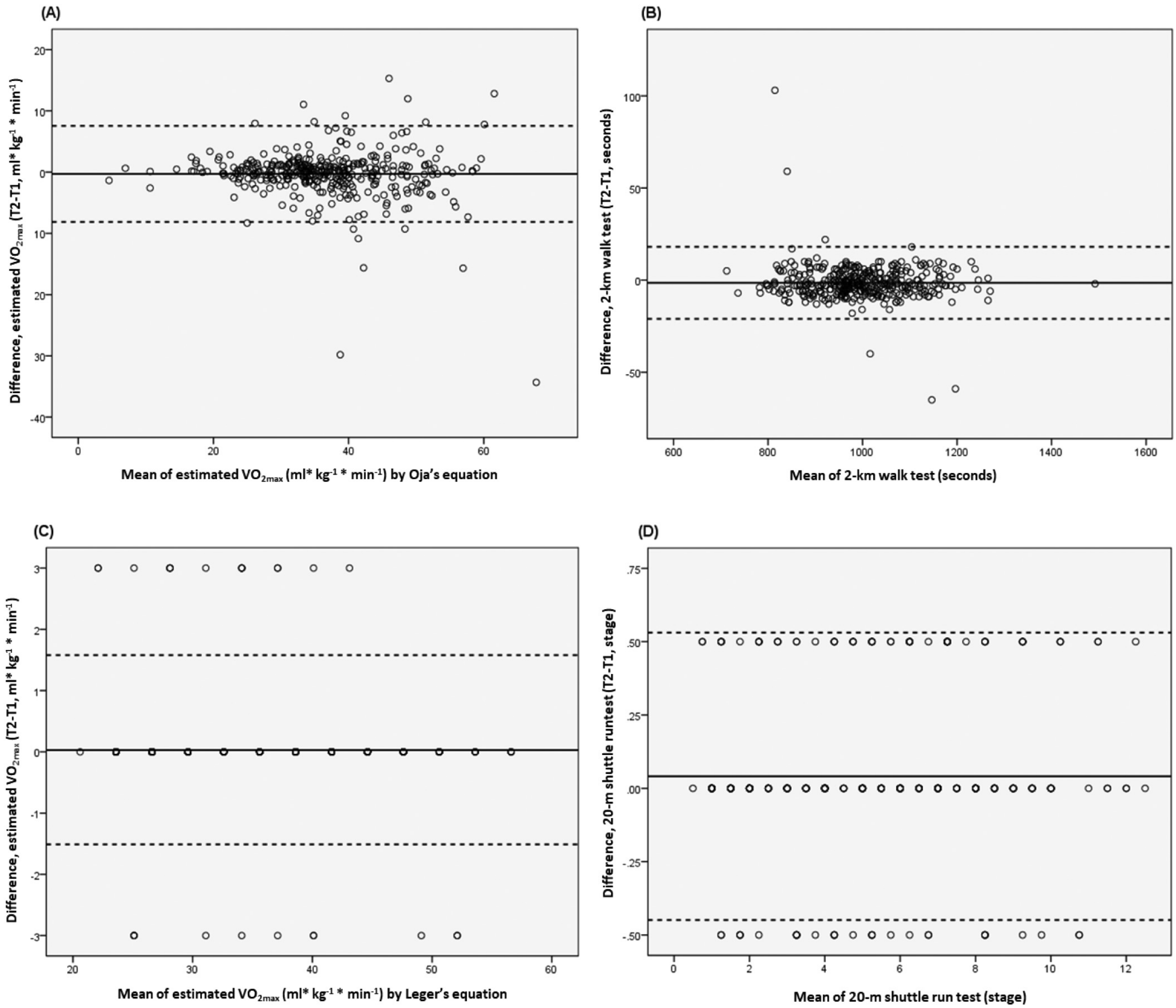

Fig. 3. Bland–Altman plot of the estimated VO_{2max} (ml* kg^{-1} * min^{-1}) by Oja's equation, and Leger's equation, the 2-km walk test (seconds), and the 20-m shuttle run test (stage). (A), Agreement between mean and difference in estimated VO_{2max} (ml* kg^{-1} * min^{-1}) by Oja's equation; (B), agreement between mean and difference in total time (sec) in the 2-km walk test; (C), agreement between mean and difference in estimated VO_{2max} (ml* kg^{-1} * min^{-1}) by Leger's equation; (D), agreement between mean and difference in final stage in the 20-m shuttle run test.
The central line represents the mean differences between retest (T2) and test (T1). The upper and lower dotted lines represent the upper and lower 95% limits of agreement (mean differences ± 1.96 SD of the differences), respectively.

Finally, Bland–Altman plots nearly 0 and narrow LoA, confirm the criterion-related validity of the 20-m SRT in the whole sample, as well as in the different groups of sex, age and physical activity level.

Mayorga et al.[29] reported that the protocol used by Leger et al.,[19] only including the final stage reached, was the protocol that presented a greater criterion-related validity with measured VO_{2max} ($r = 0.84$; $p < 0.05$), than the EUROFIT protocol ($r = 0.\ 73$, $p < 0.05$), Queen's University Belfast ($r = 0.71$, $p < 0.05$), and Dong-Ho ($r = 0.66$, $p < 0.05$). Moreover, sex and maximum oxygen uptake level did not seem to affect the criterion-related validity which is in line with our results.

A meta-analysis[30] highlights that the 20-m SRT has greater criterion-related validity than the distance tests in adults (such as the 2-km walk test) which concurs with our results. Due to the easiness in regulating the pace through an acoustic signal (where participants cannot choose their own pace) and the relatively short duration of this maximal test, the 20-m SRT seems likely to reduce the influence of psychological factors (e.g., self-motivation and monotonous) that may affect the performance, and thus the validity/reliability, when comparing with distance tests. Therefore, scientists and practitioners could use the 20-m SRT over the 2-km walk test, when no known physical impairments are present. Otherwise, the 2-km walk test is also a useful alternative to estimate cardiorespiratory fitness (i.e., participants with low physical fitness level or who is unable to run).

The main results showed a good reproducibility of the 2-km walk test and 20-m SRT, as well as the estimated VO_{2max} by Oja's and Leger's equations.

We found no significant difference between testing sessions in estimated VO_{2max} in both tests ($p > 0.05$). However, we found significant differences between 2-km walk test and 20-m SRT in terms of total time and final stage, respectively; although we cannot translate it as a *real* statistical difference, since these results are based on minimal performance changes. For instance, the mean time to complete the 2-km walk test was of 973 s, while the T1 and T2 difference was of −1.48 s. Likewise, the mean final stage reach in the 20-m SRT was 5, while the T1 and T2 difference was of 0.04 stages. Moreover, the effect size of the mean differences was small (all, Cohen $d < 0.016$).[22]

The reliability of the 2-km walk test and 20-m SRT were considered high based on coefficients of correlation (ICCs >0.90) in both tests for the whole sample. Overall, the reliability of both tests did not change according to sex, age or physical activity level. Although all of them reported a high reproducibility (ICC = 0.93 to 0.99), the 20-m SRT result and its estimated VO_{2max} by Leger's equation showed even higher reliability (all, ICC = 0.99) than the estimated VO_{2max} by Oja's eq. (ICC = 0.93 to 0.96). These results are also supported by low error values (i.e., MSE, RMSE, %Error, %SEM, %CV and SEE), indicating good data accuracy, especially for the 20-m SRT.

We have recently conducted a reliability systematic review of field-based fitness tests in adults,[10] and we found that the reliability of the 2-km walk test has been only previously analyzed in a sample of female and male adults aged 30–55 years.[27] They found a T1 and T2 difference of -0.9 ± 4.4 (95% LoA = −2.7 to 4.5 ml* kg^{-1} * min^{-1}, $p < 0.05$) for females, and -2.2 ± 3.5 (95% LoA = −2.8 to 6.6 ml* kg^{-1} * min^{-1}, $p < 0.05$) for males, with high correlation coefficients (ICCs = 0.88 to 0.91). In our study, we found no differences when comparing by sex, and taking similar age groups (without sex differences), we observed similar results to those reported by Laukkanen et al.,[27] Although with higher ICCs.

Finally, Bland–Altman plots support the reliability of the 2-km walk test, with differences nearly 0 and narrow LoA. Moreover, results from the heteroscedasticity analysis in the total time of the 2-km walk test, as well as in the estimated VO_{2max} by Oja's equation, indicate that the variability of these measurements could be greater when the participants performed better (i.e., the fitter).

Regarding the 20-m SRT, in the aforementioned systematic review, we found that it is a reliable test for young-adults (ICCs = 0.93–0.96, SEMs <15%).[10] In the present study, similar results were found for the whole sample (ICC = 0.99, SEM < 1%), as well as in adults aged 18–34 years (ICC = 0.99, SEM < 2%). Moreover, we found a T1 and T2 difference of 0.03 ± 0.04 ml* kg^{-1} * min^{-1} (95% LoA = −1.51 to 1.58, p = 0.086) for the whole sample, as well as 0.00 ± 0.07 ml* kg^{-1} * min^{-1} (95% LoA = −1.67 to 1.67, p = 0.064) in adults aged 18–34 years, being even lower than those reported in that systematic review.[10] Furthermore, results from Bland–Altman plots support the reliability of the 20-m SRT test, with differences nearly 0 and narrow LoA.

Nevertheless, some studies included in this systematic review analyzed reliability based only on the Pearson correlation coefficient which, despite being a common method to examine reliability, its use without other statistical support seems to be inappropriate.[25] In addition, the sample of these studies mainly included participants aged <45 years, being difficult to extrapolate that results to adults over this age.

In general terms, the Leger's equation (ICCs = 0.99; RMSE = 0.62–0.84; %CV = 1.57; SEE = 0.61–0.85) seems to be slightly more reliable than Oja's equation (ICCs = 0.93–0.96; RMSE = 2.37–5.13; %CV = 7.73; SEE = 2.27–4.84), to estimate VO_{2max}, regardless of sex, age or physical activity level. This fact may be explained by the difficulty in developing an appropriate pace during the 2-km walk test, starting too fast, so that the participants are unable to maintain their speed throughout the test; or too slow, increasing their speed at the end of the test (which may also mean an unexpected incremented heart rate at the end of the test). On the other hand, regarding the 20-m SRT, it is possible that following an acoustic signal could be easier for self-pace regulation.

Overall, both tests can be considered reliable. It was not possible to compare the results of the present study in terms of measurement errors, as none were available in the current literature. Nevertheless, high ICC values and low CV and SEM values suggest high levels of reliability and reproducibility,[23,24] regardless of the characteristics of individuals.

The main limitation of the present study was the lack of control of factors that can affect performance, such as genetics or experience and running economy. Moreover, although we maintained a high level of motivation throughout the participants' test performance, psychological issues as the discomfort of strenuous effort, self-motivation, and interest span for monotonous tasks may have had some uncontrolled effect on our results.

The major strengths of the study were the relatively large sample, as well as the homogeny distribution of the sample by sex, age, and physical activity level. The analysis of several anthropometric variables, including height, weight, %BF, lean mass, triceps and subscapular skinfolds, and hip and waist circumferences, also constitutes a strength, to detect which may result in a more explanatory variable for the validity equations. Finally, although ICC, Bland–Altman, SEM and CV are the most common statistic used to report reliability in sports medicine,[25] we have also included different measurement errors for a more complete interpretation of reliability.

5. Conclusions

This study was designed to analyze the criterion-related validity and the reliability of the 2-km walk test and the 20-m SRT for evaluating cardiorespiratory fitness in adult population, according to sex, age, and physical activity level. The results of this study indicate that the 2-km walk and the 20-m SRT, as well as their corresponding Oja's and Leger's equations, are valid and reliable for estimating cardiorespiratory fitness in adults aged 18–64 years. However, the 20-m SRT obtained slightly greater criterion-related validity and reliability, regardless of sex, age, and physical activity level.

Funding information

This project was supported by the Ministry of Economy, Industry and Competitiveness in the 2017 call for R&D Projects of the State Program for Research, Development and Innovation Targeting the Challenges of the Company; National Plan for Scientific and Technical Research and Innovation 2013–2016 (DEP2017-88043-R); National Plan for Scientific and Technical Research and Innovation 2017-2020 (PN/EPIF-FPU-CT/CP/2021-056); the Spanish Ministry of Education, Culture and Sport (FPU19/02961), and the Regional Government of Andalusia and University of Cadiz: Research and Knowledge Transfer Fund (PPIT-FPI19-GJ4F-10).

Confirmation of ethical compliance

The study was approved by the Review Committee for Research of Cadiz, Spain. The Declaration of Helsinki was strictly followed throughout the study.

CRediT authorship contribution statement

Nuria Marín-Jiménez: Data curation, Writing - original draft. **Sandra Sánchez-Parente:** Supervision. **Pablo Expósito-Carrillo:** Supervision. **José Jiménez-Iglesias:** Supervision. **Inmaculada C. Álvarez-Gallardo:** Supervision. **Magdalena Cuenca-García:** Conceptualization, Methodology, Supervision, Writing – review & editing. **José Castro-Piñero:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of interest statement

The authors declare that they have no competing interests.

Acknowledgements

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsams.2023.03.005.

References

1. Barry VW, Baruth M, Beets MW et al. Fitness vs. fatness on all-cause mortality: a meta-analysis. *Prog Cardiovasc Dis* 2014;56(4):382-390.
2. Lin X, Zhang X, Guo J et al. Effects of exercise training on cardiorespiratory fitness and biomarkers of cardiometabolic health: a systematic review and meta-analysis of randomized controlled trials. *J Am Heart Assoc* 2015;4(7):e002014.
3. Farrell SW, Fitzgerald SJ, McAuley PA et al. Cardiorespiratory fitness, adiposity, and all-cause mortality in women. *Med Sci Sports Exerc* 2010;42(11):2006-2012.
4. Blair SN, Kohl HW, Paffenbarger RS et al. Physical fitness and all-cause mortality: a prospective study of healthy men and women. *Jama* 1989;262(17):2395-2401.
5. Willis BL, Gao A, Leonard D et al. Midlife fitness and the development of chronic conditions in later life. *Arch Intern Med* 2012;172(17):1333-1340.
6. Sui X, Li H, Zhang J et al. Percentage of deaths attributable to poor cardiovascular health lifestyle factors: findings from the aerobics center longitudinal study. *Epidemiol Res Int* 2013;2013.
7. Castro-Piñero J, Marin-Jimenez N, Fernandez-Santos JR et al. Criterion-related validity of field-based fitness tests in adults: a systematic review. *J Clin Med* 2021;10(16):3743.
8. Castro-Piñero J, Artero EG, España-Romero V et al. Criterion-related validity of field-based fitness tests in youth: a systematic review. *Br J Sports Med* 2010;44(13):934-943.
9. Currell K, Jeukendrup AE. Validity, reliability and sensitivity of measures of sporting performance. *Sports Med* 2008;38(4):297-316.
10. Cuenca-Garcia M, Marin-Jimenez N, Perez-Bey A et al. Reliability of field-based fitness tests in adults: a systematic review. *Sports Med* 2022:1-20.
11. Marfell-Jones M, Olds T, Stewart A et al. *ISAK Accreditation Handbook*, 2006.
12. Control CfD, Prevention. *National Health and Nutrition Examination Survey: Anthropometry Procedures Manual*, Atlanta, GA, USA, CDC, 2007.
13. Research NIoHOoMAo. *Bioelectrical Impedance Analysis in Body Composition Measurement: National Institutes of Health Technology Assessment Conference Statement, December 12–14, 1994*, NIH Office of Medical Applications of Research, 1994.
14. Balke B, Ware RW. An experimental study of physical fitness of Air Force personnel. *US Armed Forces Med J* 1959;10(6):675-688.
15. Martin-Rincon M, González-Henríquez JJ, Losa-Reyna J et al. Impact of data averaging strategies on VO2max assessment: mathematical modeling and reliability. *Scand J Med Sci Sports* 2019;29(10):1473-1488.
16. Midgley AW, McNaughton LR, Polman R et al. Criteria for determination of maximal oxygen uptake. *Sports Med* 2007;37(12):1019-1028.
17. Oja P, Laukkanen R, Pasanen M et al. A 2-km walking test for assessing the cardiorespiratory fitness of healthy adults. *Int J Sports Med* 1991;12(04):356-362.
18. El-Amrawy F, Nounou MI. Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? *Healthc Inform Res* 2015;21(4):315-320.
19. Leger LA, Mercier D, Gadoury C et al. The multistage 20 metre shuttle run test for aerobic fitness. *J Sports Sci* 1988;6(2):93-101.
20. Silva P, Franco J, GUSMãO A et al. Trunk strength is associated with sit-to-stand performance in both stroke and healthy subjects. *Eur J Phys Rehabil Med* 2015;51(6):717-724.
21. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* 1986;327(8476):307-310.
22. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, Routledge, 2013.
23. Vincent-Smith B, Gibbons P. Inter-examiner and intra-examiner reliability of the standing flexion test. *Man Ther* 1999;4(2):87-93.
24. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 2005;19:231-240. doi:10.1519/15184.1.
25. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med* 1998;26(4):217-238.
26. Laukkanen R, Oja P, Pasanen M et al. Validity of a two kilometre walking test for estimating maximal aerobic power in overweight adults. *Int J Obes Relat Metab Disord* 1992;16(4):263-268.
27. Laukkanen R, Kukkonen-Harjula T, Oja P et al. Prediction of change in maximal aerobic power by the 2-km walk test after walking training in middle-aged adults. *Int J Sports Med* 2000;21(02):113-116.
28. Laukkanen R, Oja R, Pasanen M et al. Criterion validity of a two-kilometer walking test for predicting the maximal oxygen uptake of moderately to highly active middle-aged adults. *Scand J Med Sci Sports* 1993;3(4):267-272.
29. Mayorga-Vega D, Aguilar-Soto P, Viciana J. Criterion-related validity of the 20-m shuttle run test for estimating cardiorespiratory fitness: a meta-analysis. *J Sports Sci Med* 2015;14(3):536.
30. Mayorga-Vega D, Bocanegra-Parrilla R, Ornelas M et al. Criterion-related validity of the distance-and time-based walk/run field tests for estimating cardiorespiratory fitness: a systematic review and meta-analysis. *PloS One* 2016;11(3):e0151671.