



Article

Does the Multistage 20-m Shuttle Run Test Accurately Predict VO_{2max} in NCAA Division I Women Collegiate Field Hockey Athletes?

Meghan K. Magee^{1,2}, Jason B. White^{2,3}, Justin J. Merrigan^{2,4}  and Margaret T. Jones^{2,5,*} ¹ Kinesiology, George Mason University, Manassas, VA 20110, USA; mmagee2@gmu.edu² Frank Pettrone Center for Sports Performance, George Mason University, Fairfax, VA 22030, USA; whitej4@ohio.edu (J.B.W.); justin.merrigan@hsc.wvu.edu (J.J.M.)³ Exercise Physiology, Ohio University, Athens, OH 45701, USA⁴ Human Performance Innovation Center, Rockefeller Neuroscience Institute, West Virginia University, Morgantown, WV 26506, USA⁵ Sport, Recreation, and Tourism Management, George Mason University, Fairfax, VA 22030, USA

* Correspondence: mjones15@gmu.edu

Abstract: Laboratory assessments of maximal oxygen uptake (VO_{2max}) are considered the “gold standard” for ascertaining cardiovascular fitness, but they are not always practical for use in team sport settings. Therefore, the purpose of the current study was to compare the criterion assessment of VO_{2max} on a treadmill to the progressive, multistage 20-m shuttle run test (i.e., Beep test), and to determine the predictability of 6 previously established Beep test predictive equations (i.e., Chatterjee, Flouris, Leger, Leger and Gadoury, Ramsbottom, St. Clair-Gibson). Collegiate women field hockey athletes ($n = 65$, mean \pm SD: age 19.6 ± 1.2 years; weight 64.7 ± 6.1 kg) completed criterion VO_{2max} (mean \pm SD: 46.4 ± 4.6 mL \cdot kg $^{-1}\cdot$ min $^{-1}$) and Beep tests to volitional fatigue. According to Bland–Altman and Ordinary Least Products Regressions, the Ramsbottom (46.5 ± 4.2 mL \cdot kg $^{-1}\cdot$ min $^{-1}$) and Flouris (46.3 ± 3.8 mL \cdot kg $^{-1}\cdot$ min $^{-1}$) equations were considered valid predictions of criterion measured VO_{2max} (46.4 ± 4.6). The Chatterjee, Leger, Leger and Gadoury, and St. Clair-Gibson equations overestimated VO_{2max} , and are not recommended for use with women collegiate field hockey athletes. The Ramsbottom and Flouris estimates of VO_{2max} from 20-m shuttle performances may be used in this population. For accurate estimates of VO_{2max} , the clientele’s age, fitness level, and training history should be considered when selecting equations.

Keywords: aerobic power; women athletes; Beep test; cardiovascular fitness; oxygen consumption



Citation: Magee, M.K.; White, J.B.; Merrigan, J.J.; Jones, M.T. Does the Multistage 20-m Shuttle Run Test Accurately Predict VO_{2max} in NCAA Division I Women Collegiate Field Hockey Athletes? *Sports* **2021**, *9*, 75. <https://doi.org/10.3390/sports9060075>

Academic Editors: John J. McMahon and Paul Comfort

Received: 23 April 2021

Accepted: 21 May 2021

Published: 26 May 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Field hockey is a field-based, intermittent, team sport comprised of short bouts of high intensity sprints and longer bouts of walking and jogging [1]. The aerobic energy system is predominant during collegiate women’s field hockey, as games consist of two 35-min halves with high-intensity movements comprising approximately 20% of that duration [2]. Further, according to time motion analysis, 95% of men’s field hockey competitions consist of low intensity tasks (e.g., walking, jogging, standing), while 5% of the competition is comprised of high intensity tasks (e.g., sprinting) [3]. Due to the evidence that field hockey is predominately an aerobic sport [3–5], testing for aerobic capacity is imperative in understanding the athletes’ preparedness for competition and for modification of training programs.

Aerobic capacity can be validly and reliably assessed through measurement of inspired and expired gas exchange during a maximal graded exercise test [6]. The maximal oxygen consumption (VO_{2max}) refers to the highest physiological value attainable, as indicated by a plateau in maximal oxygen consumption during exercise. However, VO_{2peak} , the highest value obtained during exercise, should not be used interchangeably with VO_{2max} .

VO_{2peak} is more indicative of exercise tolerance than the maximal ability to transport and utilize oxygen, since individuals may select to discontinue exercise prior to achieving VO_{2max} [7–10]. Nonetheless, the aforementioned assessment of aerobic capacity requires expensive laboratory housed equipment and highly trained personnel, which are not always accessible or feasible for testing multiple athletes in team settings. Therefore, alternative, field based, methods for predicting VO_{2max} have been developed. A common and reliable test used to predict aerobic capacity is the progressive, multistage 20-m shuttle-run test or “Beep” test, in which the athlete runs 20-m laps at increasingly faster speeds until volitional fatigue [11–13]. Although VO_{2max} of men and women within the general population has proven to be correlated with the final shuttle speed during the Beep test [11,14,15], the ability of 20-m shuttle run tests to accurately predict VO_{2max} outcomes in collegiate athletes remains limited.

In order to estimate VO_{2max} from Beep test performance, several equations have been developed that use the speed of the final stage achieved during the Beep test [15–20]. Equations have been validated within the general population [16,17,19], recreationally active men and women [15], children ages 8–19 years [20], and men squash and distance athletes [18]. However, there are differences in the aerobic capacity of athletes depending upon the sport in question, which may affect the validity of VO_{2max} equations from the Beep test in certain populations. Further, despite previously reported similarities in aerobic capacity between elite men and women field hockey athletes and endurance athletes [4], prediction of VO_{2max} can be dependent upon sex [21]. Thus, further investigation of the accuracy of several VO_{2max} prediction equations is warranted in various athletic settings, such as women field hockey athletes, to provide practitioners with appropriate equations for their population. This is critical for reliably providing VO_{2max} predictions, to reduce the chance of error in training prescriptions from overestimations or underestimations of VO_{2max} . To assess true agreement between the predicted and criterion measured values, Bland–Altman and Ordinary Least Products Regressions are commonly conducted statistical procedures [22,23]. However, commonly reported statistics include correlational and linear regression analyses or t-tests, which assess the relationship or mean group differences between the assessments, not agreement. Therefore, this brings into question prior reporting on the accuracy of several 20-m shuttle VO_{2max} prediction equations. Thus, the aim of the current study was to determine the criterion validity of Beep test algorithms for predicting VO_{2max} in a sample of National Collegiate Athletic Association Division I women field hockey athletes.

2. Materials and Methods

2.1. Experimental Design

In a randomized order, field hockey athletes performed a laboratory-based maximal graded treadmill assessment and a field test of aerobic capacity at the same time of day on two days separated by 48 h. The laboratory assessment consisted of a continuous, incremental running protocol on a motorized treadmill. The field test was the progressive, multistage 20-m shuttle run test, also known as the Beep test. All athletes performed both assessments to volitional fatigue. Testing took place during the off-season. All athletes were familiar with treadmill running and the Beep test.

2.2. Subjects

Sixty-five National Collegiate Athletic Association Division I women field hockey athletes (age, 20 ± 1 years; body mass, 64.7 ± 6.1 kg; body fat, $24.5 \pm 5.5\%$; final Beep stage velocity, 12.7 ± 0.6 m·s⁻¹) participated in the current study. All athletes were under the direction of a certified strength and conditioning coach (NSCA-CSCS) and were following a similar training regimen. Athletes were instructed to refrain from exercise, alcohol, and supplementation 24 h prior to testing. Additionally, athletes were instructed to refrain from food and drink two hours prior to testing. All athletes completed a medical history form and were cleared for intercollegiate athletic participation.

2.3. Criterion Measure of Maximum Oxygen Uptake

Laboratory assessment of $\text{VO}_{2\text{max}}$ was conducted via a maximal graded treadmill (Marquette 1900, Milwaukee, WI, USA) protocol with 1-min stages. The protocol began at 3.0 mph and increased speed to 5.0 mph after one min. At the third min, speed was increased by 1.0 mph per min until speed was 8.0 mph. At min 6 and beyond, speed was held constant at 8.5 mph and only grade was increased by 1% per min. Heart rate was monitored continuously via heart rate monitor and watch (Polar Electro, Kempele, Finland), and the highest heart rate reached during each min was recorded. Oxygen consumption (VO_2) was attained via expired gas analysis (VO2000 Metabolic System, MedGraphics, St. Paul, MN, USA), which was calibrated before each test according to manufacturer guidelines. The $\text{VO}_{2\text{max}}$ corresponded to the highest VO_2 reached prior to volitional fatigue. Achievement of $\text{VO}_{2\text{max}}$ was dependent upon participants meeting three of the following criteria: RPE greater than or equal to 18; respiratory exchange ratio of 1.1 or greater; a plateau in VO_2 ($<150 \text{ mL}\cdot\text{min}^{-1}$) despite an increase in workload; a maximal attained heart rate within 10 beats per min of age predicted heart rate maximum ($206.9 - 0.67 \times \text{age}$); and a venous blood lactate $> 8 \text{ mM}$.

2.4. Progressive, Multistage 20-m Shuttle Run Test

The protocol followed for the progressive, multistage 20-m shuttle run (Beep) test was a commonly used modification [13,15,24,25] of the original developed by Leger and Lambert [11]. The Beep test was performed on an indoor basketball court with wooden floors and required athletes to run back and forth (“shuttle”) between two cones separated by 20 m. The initial speed was $2.22 \text{ m}\cdot\text{s}^{-1}$ for one min. At the end of the first min, the speed increased to $2.5 \text{ m}\cdot\text{s}^{-1}$ and progressively increased by $0.14 \text{ m}\cdot\text{s}^{-1}$ each min thereafter (Table 1). The speed was dictated by audible beeps from prerecorded audio, which was checked for accuracy prior to testing. Each min stage consisted of multiple “shuttles”, the number of which was dependent upon the stage speed. Athletes were instructed to keep running at the pace of the beeps for as long as possible, and they were grouped to provide a competitive environment. Once an athlete could no longer keep pace with the beeps (i.e., failed to complete two consecutive shuttles in time), the test was terminated. Estimated $\text{VO}_{2\text{max}}$ was predicted from the maximal speed attained during the test via six previously developed prediction equations available in Table 2 [13,15–18,26].

Table 1. Protocol for progressive, multistage 20-m shuttle run (Beep) test.

	Number of Shuttles	Speed ($\text{m}\cdot\text{s}^{-1}$)	Speed ($\text{km}\cdot\text{h}^{-1}$)
Stage 1	7	2.22	8
Stage 2	8	2.5	9
Stage 3	8	2.64	9.5
Stage 4	9	2.78	10
Stage 5	9	2.92	10.5
Stage 6	10	3.06	11
Stage 7	10	3.2	11.5
Stage 8	11	3.34	12
Stage 9	11	3.48	12.5
Stage 10	11	3.62	13
Stage 11	12	3.76	13.5
Stage 12	12	3.9	14
Stage 13	13	4.04	14.5
Stage 14	13	4.18	15
Stage 15	13	4.32	15.5
Stage 16	14	4.46	16
Stage 17	14	4.6	16.5
Stage 18	15	4.74	17
Stage 19	15	4.88	17.5
Stage 20	16	5.02	18
Stage 21	16	5.16	18.5

Table 2. The 20-m shuttle run prediction equation.

Reference	Equation
Ramsbottom et al. [15]	Table III
Leger et al. [20]	$-24.4 + 6.0 \times \text{MAS}$
Leger and Gadoury [16]	$(\text{MAS} \times 6.592) - 32.678$
Chatterjee et al. [17]	$-14.956 + 4.78 \times \text{MAS}$
St. Clair-Gibson et al. [18]	$6 \times \text{MAS} - 24$
Flouris et al. [19]	$(\text{MAS} \times 6.65 - 35.8) \times 0.95 + 0.182$

MAS, maximal aerobic speed.

2.5. Statistical Analysis

Data from criterion and Ramsbottom et al. [15] predicted $\text{VO}_{2\text{max}}$ were normally distributed according to histogram plots and Shapiro–Wilks tests. However, due to the nature of the $\text{VO}_{2\text{max}}$ predictions formulated from the final speed reached in the incremental treadmill test, these data were not normally distributed. Thus, these data do not meet the assumptions of standard parametric statistical assessments, such as correlations and t-tests. Nonparametric paired samples Wilcoxon Signed Rank tests and respective effect sizes were calculated to determine mean comparisons between predicted $\text{VO}_{2\text{max}}$ from the criterion and Beep test equations.

Bland–Altman and Ordinary Least Products Regressions were run to determine the systematic and proportionate agreement between each predictive equation and the measured $\text{VO}_{2\text{max}}$. The differences between the assessments were normally distributed and thus, assumptions of the Bland–Altman plots were met. Visualizations of the comparisons were created using Bland–Altman plots including 95% limits of agreement (mean bias $\pm [1.96 \times \text{SD of Differences}]$), their 95% confidence intervals, and a trendline of proportional bias and standard error. Bland–Altman plots and the respective statistics were created and assessed using BlandR package [27]. Since the interpretation of these plots has potential for error [23,28], ordinary least square products (Model II linear regression) analyses were performed to assess agreement between devices [29,30]. Systematic and proportional bias were noted when 95% confidence interval of the intercept did not include “0” and “1.0”, respectively. If the predicted $\text{VO}_{2\text{max}}$ of an assessment resulted in systematic or proportionate bias, then this method should not be used interchangeably with criterion measured $\text{VO}_{2\text{max}}$. All statistical procedures were conducted using R, version 3.6.2 (R Core Team, Vienna, Austria; <https://www.R-project.org>). The intercept, slope, and 95% confidence intervals of the Ordinary Least Products Regression analyses were calculated using the “sma” function within the “smatr” package [31]. For all statistical tests, alpha level was $p < 0.05$.

3. Results

Leger et al. [20], Leger and Gadoury [16], and St. Clair-Gibson et al. [18] predicted $\text{VO}_{2\text{max}}$ values were significantly different, with large overestimations, from the criterion $\text{VO}_{2\text{max}}$ (Table 3).

Table 3. Mean comparisons of maximal oxygen uptake ($\text{VO}_{2\text{max}}$) measurements.

Measurement	$\text{VO}_{2\text{max}}$ ($\text{mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$)	p-Value	Effect Size	Effect Magnitude
Criterion	46.4 ± 4.6	–	–	–
Ramsbottom et al. [15]	46.5 ± 4.2	0.563	0.07	small
Leger et al. [20]	50.2 ± 3.5	<0.001	0.71	large
Leger and Gadoury [16]	50.9 ± 4.0	<0.001	0.77	large
Chatterjee et al. [17]	45.7 ± 2.9	0.185	0.16	small
St. Clair Gibson et al. [18]	52.1 ± 3.6	<0.001	0.82	large
Flouris et al. [19]	46.3 ± 3.8	0.932	0.01	small

Values are represented as mean \pm standard deviation. p-values from Wilcoxon Signed Ranked tests and respective Rank Biserial effect sizes and magnitudes are displayed for comparison of Beep test calculations to criterion measures of $\text{VO}_{2\text{max}}$.

Results of the Bland–Altman and Least Products Regression analyses indicated true bias was not equal to 0 for Leger et al. [20], Leger and Gadoury [16], and St. Clair-Gibson et al. [18] (Table 4). The Ramsbottom et al. [15] (Figure 1) and Flouris et al. [19] (Figure 2) predicted VO_{2max} did not note bias according to Ordinary Least Products Regression or Bland–Altman procedures (Table 4). Systematic bias was indicated with wide limits of agreement for St. Clair-Gibson et al. [18] (Figure 3), Leger and Gadoury [16] (Figure 4), Leger et al. [20] (Figure 5). Least Products Regression analyses (Table 3) displayed systematic and proportional bias for Leger et al. [20] (Figure 5), Chatterjee et al. [17] (Figure 6), and St. Clair Gibson et al. (Figure 3) [18].

Table 4. Least Products Regression and BlandR results for agreement between VO_{2max} predictions (df = 64).

VO_{2max} Equation	R^2	Intercept ($CI_{95\%}$)	Slope ($CI_{95\%}$)	BlandR Stats
Ramsbottom et al. [15]	0.367 *	−3.925 (−14.051, 6.201)	1.081 (0.886, 1.319)	$t = -0.303; p = 0.763$
Leger et al. [20]	0.308 *	−18.816 (−32.513, −5.120)	1.300 (1.056, 1.600)	$t = 7.811; p < 0.001$
Leger and Gadoury [16]	0.308 *	−11.861 (−24.106, 0.384)	1.143 (0.929, 1.408)	$t = 9.059; p < 0.001$
Chatterjee et al. [17]	0.308 *	−25.643 (−40.765, −10.520)	1.577 (1.281, 1.941)	$t = -1.508; p = 0.136$
St. Clair-Gibson et al. [18]	0.308 *	−19.076 (−32.827, −5.326)	1.256 (1.020, 1.547)	$t = 11.714; p < 0.001$
Flouris et al. [19]	0.308 *	−8.865 (−20.485, 2.754)	1.193 (0.969, 1.469)	$t = -0.162; p = 0.872$

*, indicates significant R^2 . Bolded values indicate systematic bias (intercept $CI_{95\%}$ does not cross 0) and proportional bias (slope $CI_{95\%}$ does not cross 1).

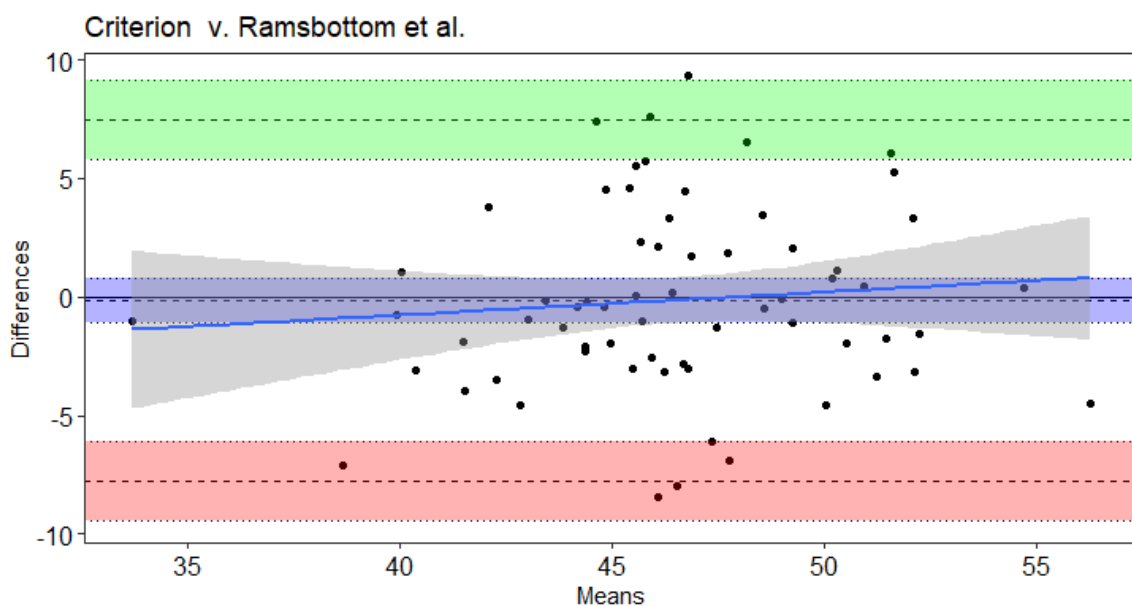


Figure 1. Bland–Altman plot of differences and means for criterion measured VO_{2max} from a maximal incremental treadmill test and the Ramsbottom et al. Beep test algorithm. The 3 dotted lines with blue shading represent mean bias, green shading represents upper 95% threshold, and red shading represents lower 95% threshold. The center line is either the mean difference or [$1.96 \times SD$ of Differences]), while the dotted lines above and below the center are the respective 95% confidence intervals. The solid blue line and the gray shading represent the trendline of proportional bias and the proportional standard error of the slope estimates, respectively.

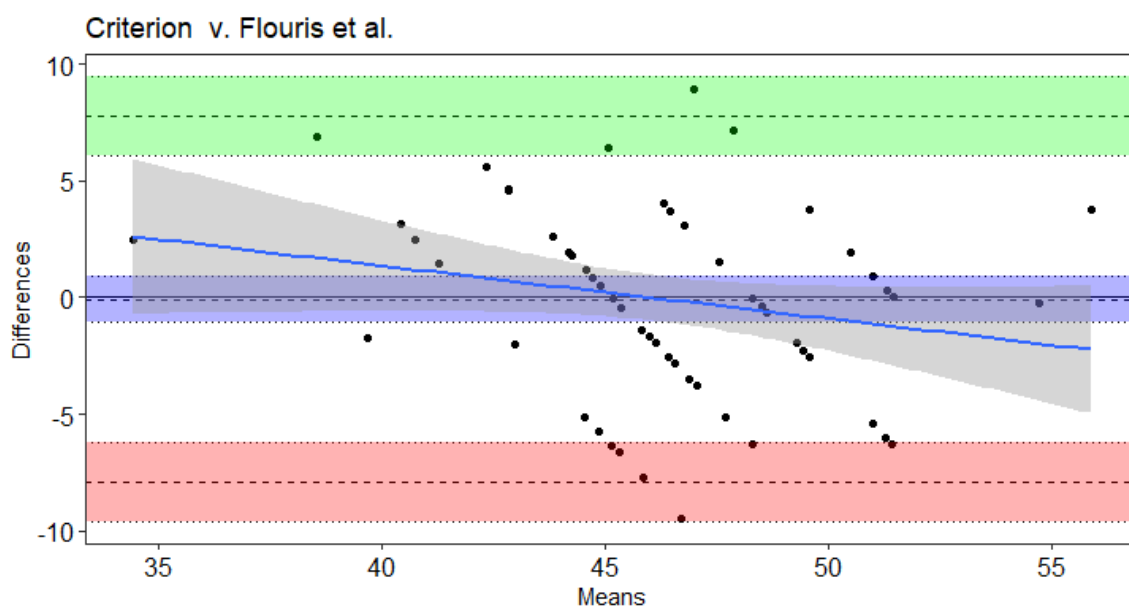


Figure 2. Bland–Altman plot of differences and means for criterion measured VO_{2max} from a maximal incremental treadmill test and the Flouris et al. Beep test algorithm. The 3 dotted lines with blue shading represent mean bias, green shading represents upper 95% threshold, and red shading represents lower 95% threshold. The center line is either the mean difference or $[1.96 \times SD \text{ of Differences}]$, while the dotted lines above and below the center are the respective 95% confidence intervals. The solid blue line and the gray shading represent the trendline of proportional bias and the proportional standard error of the slope estimates, respectively.

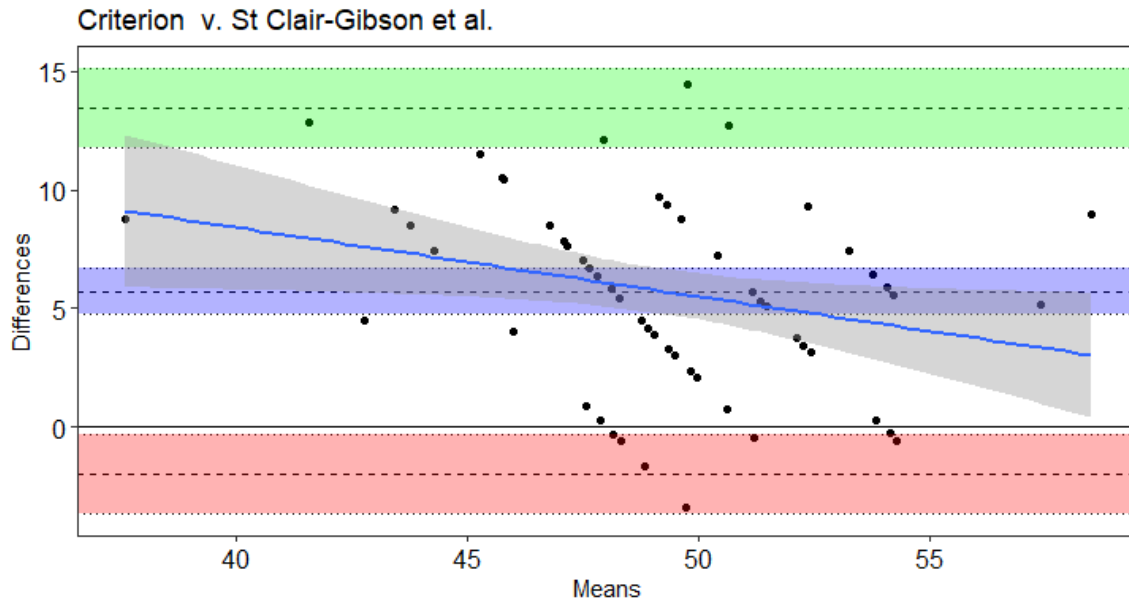


Figure 3. Bland–Altman plot of differences and means for criterion measured VO_{2max} from a maximal incremental treadmill test and the St. Clair Gibson et al. Beep test algorithm. The 3 dotted lines with blue shading represent mean bias, green shading represents upper 95% threshold, and red shading represents lower 95% threshold. The center line is either the mean difference or $[1.96 \times SD \text{ of Differences}]$, while the dotted lines above and below the center are the respective 95% confidence intervals. The solid blue line and the gray shading represent the trendline of proportional bias and the proportional standard error of the slope estimates, respectively.

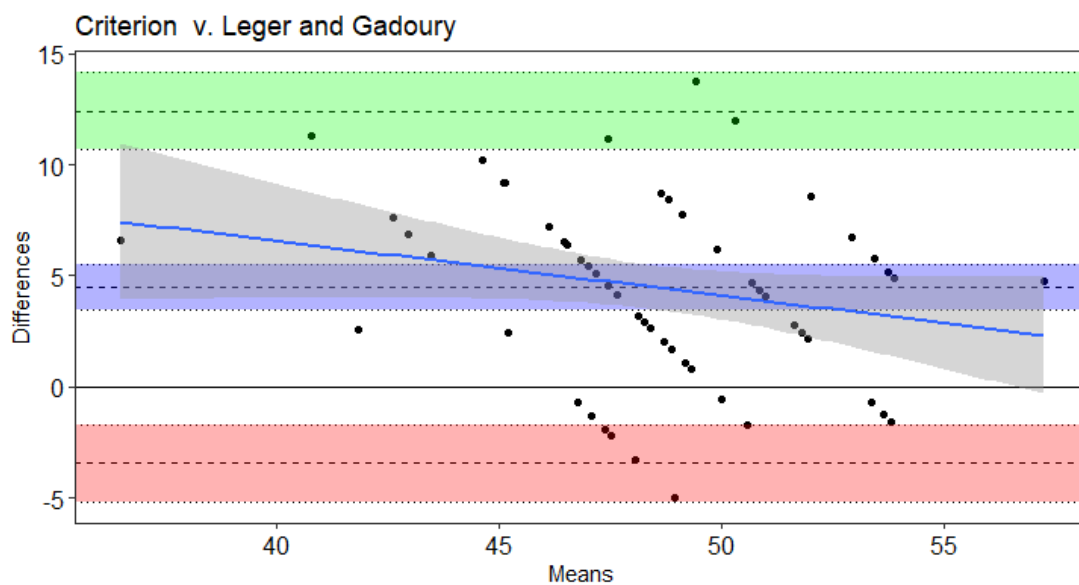


Figure 4. Bland–Altman plot of differences and means for criterion measured VO_{2max} from a maximal incremental treadmill test and the Leger and Gadoury Beep test algorithm. The 3 dotted lines with blue shading represent mean bias, green shading represents upper 95% threshold, and red shading represents lower 95% threshold. The center line is either the mean difference or $[1.96 \times SD \text{ of Differences}]$, while the dotted lines above and below the center are the respective 95% confidence intervals. The solid blue line and the gray shading represent the trendline of proportional bias and the proportional standard error of the slope estimates, respectively.

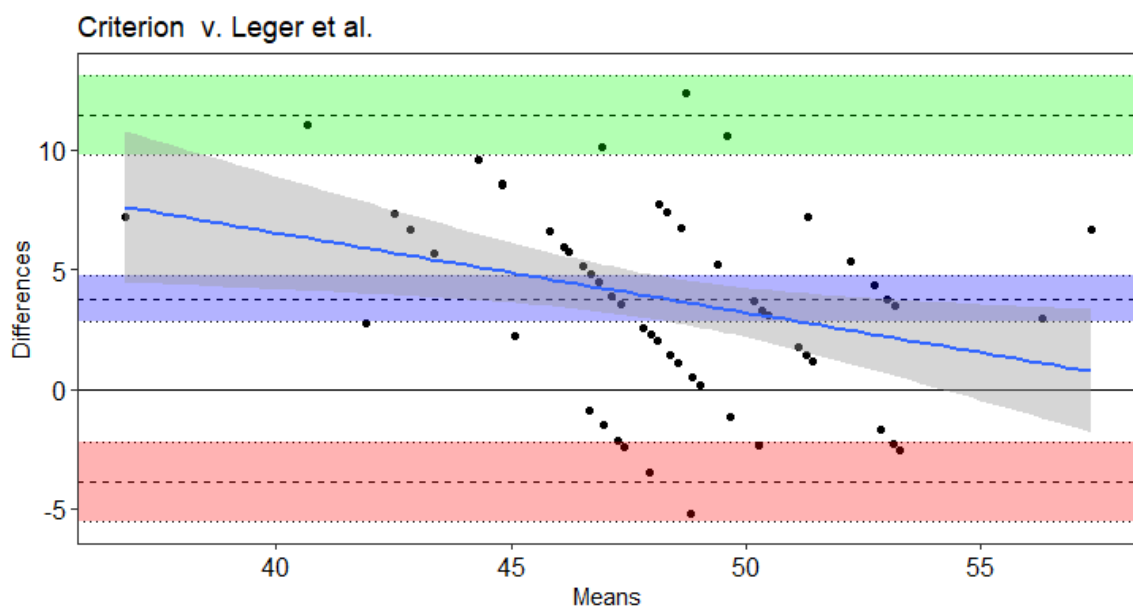


Figure 5. Bland–Altman plot of differences and means for criterion measured VO_{2max} from a maximal incremental treadmill test and the Leger et al. Beep test algorithm. The 3 dotted lines with blue shading represent mean bias, green shading represents upper 95% threshold, and red shading represents lower 95% threshold. The center line is either the mean difference or $[1.96 \times SD \text{ of Differences}]$, while the dotted lines above and below the center are the respective 95% confidence intervals. The solid blue line and the gray shading represent the trendline of proportional bias and the proportional standard error of the slope estimates, respectively.

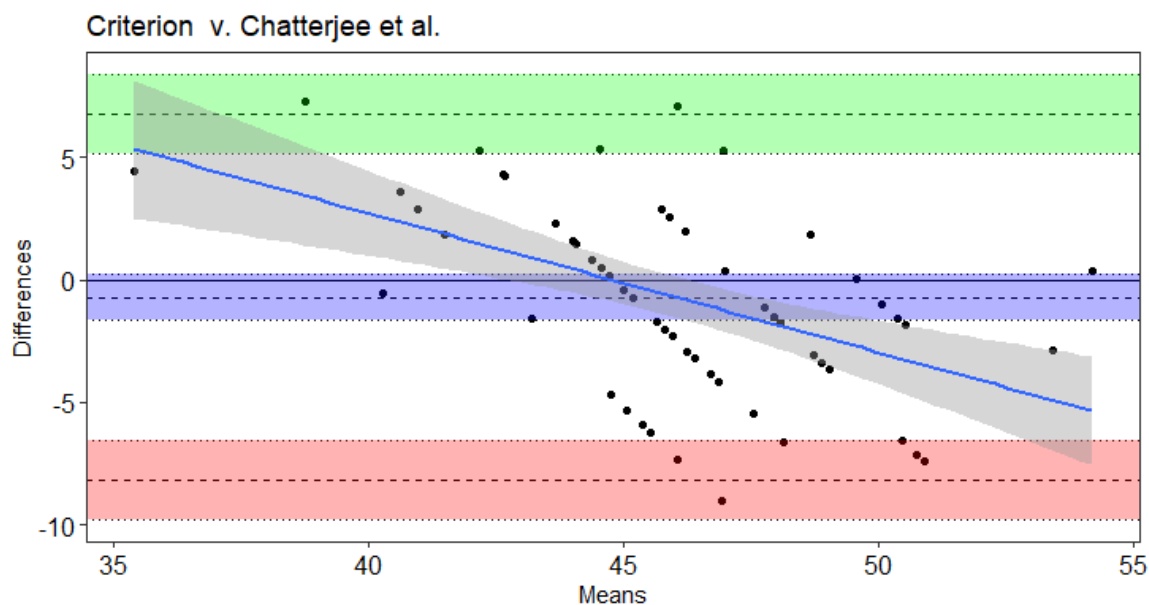


Figure 6. Bland–Altman plot of differences and means for criterion measured $\text{VO}_{2\text{max}}$ from a maximal incremental treadmill test and the Chatterjee et al. Beep test algorithm. The 3 dotted lines with blue shading represent mean bias, green shading represents upper 95% threshold, and red shading represents lower 95% threshold. The center line is either the mean difference or $[1.96 \times \text{SD of Differences}]$, while the dotted lines above and below the center are the respective 95% confidence intervals. The solid blue line and the gray shading represent the trendline of proportional bias and the proportional standard error of the slope estimates, respectively.

4. Discussion

The aim of the current study was to determine the criterion validity of Beep test algorithms for predicting $\text{VO}_{2\text{max}}$ of National Collegiate Athletic Association Division I women field hockey athletes. The Leger, Leger and Gadoury, and St. Clair-Gibson equations largely overestimated $\text{VO}_{2\text{max}}$. Yet, according to Least Products Regression and Bland–Altman analyses, the Leger, Leger and Gadoury, Chatterjee, and St. Clair-Gibson equations displayed systematic and proportionate bias when predicting $\text{VO}_{2\text{max}}$. However, the Ramsbottom and Flouris equations were considered valid predictors of $\text{VO}_{2\text{max}}$ in the current sample and thus, are considered appropriate equations for women field hockey athletes when compared to the other equations investigated in this study. Although the Ramsbottom and Flouris equations were statistically in agreement, the wide limits of agreement suggest each equation may have error in predictability up to $10 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$. Thus, these equations should still be used with caution as comparisons at the individual level have displayed unacceptable error rates that would likely influence practical implications from the testing results.

The Ramsbottom et al. equation validly predicted $\text{VO}_{2\text{max}}$ in the current study, but previously has underestimated $\text{VO}_{2\text{max}}$ by $5.18\text{--}7.90 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ in male adolescent basketball players [32]. The underestimations in the aforementioned study may be a result of the higher average $\text{VO}_{2\text{max}}$ of the male basketball players ($55.45 \pm 4.98 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$). According to the current study, the Flouris et al. equation of the Beep test was also a valid predictor of $\text{VO}_{2\text{max}}$. However, previous findings indicated the Flouris et al. equation to significantly underestimate laboratory $\text{VO}_{2\text{max}}$ in women soccer athletes by $2.4 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ [13]. In each of the aforementioned studies, the protocol of the laboratory $\text{VO}_{2\text{max}}$ tests were different from the current study, which may yield different results. Though the Flouris equation was validated in men, it is important to note that the $\text{VO}_{2\text{max}}$ achieved by their sample was $46.9 \pm 5.7 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$, which is similar to that of the present study [19].

The Leger and Gadoury equation, developed in recreationally active adult men and women [16], overestimated $\text{VO}_{2\text{max}}$ by $\sim 4.5 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ in the current sample of

women field hockey athletes. Previous literature has reported this equation to be valid in collegiate women soccer athletes [13]. Despite both of these athletic populations including women athletes in high aerobic demanding sports with similar $\text{VO}_{2\text{max}}$ performances, differences between the criterion treadmill test of the Green et al. [13] study (using Leger and Gadoury equation) and current study may have resulted in discrepancies in the findings. Yet, the $\sim 5.7 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ overestimation of the predicted $\text{VO}_{2\text{max}}$ from the St. Clair-Gibson et al. equation is likely due to its development in men squash and endurance running athletes. The men squash players ($63.4 \pm 6.1 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$) and endurance runners ($69.6 \pm 4.2 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$) attained greater $\text{VO}_{2\text{max}}$ values than women field hockey players in the current study ($46.4 \pm 4.6 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$). However, the overestimations of the St. Clair-Gibson equation seemed to be greater at lower $\text{VO}_{2\text{max}}$ levels, according to the proportional bias of Bland–Altman analyses.

St. Clair-Gibson et al. investigated the relationship between estimation of $\text{VO}_{2\text{max}}$ and laboratory $\text{VO}_{2\text{max}}$ in men athletes participating in squash ($n = 10$) and endurance running ($n = 10$), and reported the Ramsbottom equation to significantly underpredict $\text{VO}_{2\text{max}}$ in endurance runners [18]. The cyclical nature of endurance running may render a 20-m shuttle assessment unsuitable for this population as endurance running does not involve a high volume of intermittent bursts of anaerobic effort nor frequent changes of direction. Additionally, previous literature has reported that predicted $\text{VO}_{2\text{max}}$ determined from the Beep test requires significant contribution from anaerobic metabolism because of the need to slow down and accelerate every 20 m [13,33,34]. There was no difference between measured and predicted $\text{VO}_{2\text{max}}$ values with squash athletes, which was attributed, in part, to the squash athletes' skill in making calculated movements during intermittent exercise [35].

Contrary to current findings, the Leger et al. and Leger and Gandoury equations have resulted in accurate estimations of $\text{VO}_{2\text{max}}$ in men Air Force cadets with a $\text{VO}_{2\text{max}}$ of $56.8 \pm 4.3 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ [12]. Of note, the Leger and Gandoury equation was developed in 53 males and 24 females aged 19–47 years with varying levels of aerobic fitness [16] while the Leger et al. equation was developed in a sample of children ages 8–19 years. Similar to the present study, Green et al. found the Leger and Gandoury equation to overestimate $\text{VO}_{2\text{max}}$ by $5.4 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ in women soccer athletes [13]. Fitness level is another important characteristic to consider when predicting $\text{VO}_{2\text{max}}$ from selected equations. The Chatterjee estimation equation was developed with college-aged women attending university in India [17]. Although the current study involved collegiate women subjects of similar age, they were trained, competitive athletes and of a higher fitness level than the untrained women in the Chatterjee study who obtained a $\text{VO}_{2\text{max}}$ of $32.8 \pm 2.9 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$. This is evident within the results of the current study as the Chatterjee equation becomes increasingly inaccurate at greater $\text{VO}_{2\text{max}}$ values. Findings from the current study in combination with previous literature highlight the importance of accounting for sex, sport, fitness level, and age when assessing aerobic fitness.

There are other necessary considerations, such as the individual's age, fitness level, training history, and criterion measured $\text{VO}_{2\text{max}}$ when analyzing the validity of estimation equations and comparing to prior literature. For example, Bland–Altman plots displayed the influence of an individual's $\text{VO}_{2\text{max}}$ on the accuracy of 20-m shuttle equations to estimate $\text{VO}_{2\text{max}}$. Thus, discrepancies between the current and prior findings are likely attributed to differences in fitness level ($\text{VO}_{2\text{max}}$). Discrepancies in findings may also be attributed to differences in statistical analyses used to test the validity of the equations. Although common in prior literature, use of correlation analyses or group mean comparisons to determine agreement are often inappropriate as they fail to provide valuable information for reliable comparisons [23]. For example, the Chatterjee equation was not significantly different from laboratory $\text{VO}_{2\text{max}}$ according to group mean differences (i.e., Wilcoxon tests); however, according to Ordinary Least Products Regressions, the Chatterjee equation demonstrated systematic bias (underestimations) and proportional bias (changes in accuracy across $\text{VO}_{2\text{max}}$ values). Thus, comparisons of average values between the

criterion and equation estimated $\text{VO}_{2\text{max}}$ are likely not accurate assessments of mean agreement. Instead, analyses such as Ordinary Least Products Regression and Bland–Altman analyses, that consider the mean and spread of agreement at the individual level should be conducted.

5. Conclusions

The Ramsbottom and Flouris equations applied to 20-m shuttle run tests demonstrated valid estimations of $\text{VO}_{2\text{max}}$ compared to criterion metabolic cart measures during maximal graded treadmill assessments. Thus, these equations may be a valid assessment of maximal oxygen consumption abilities of women field hockey athletes, when laboratory testing of $\text{VO}_{2\text{max}}$ is unavailable. However, the equations should still be used with caution as the wide limits of agreement between the criterion and equations suggest high error rates at the individual level. The findings from the current study stress the importance of using caution when estimating $\text{VO}_{2\text{max}}$ from prediction equations as results were inconsistent. In order to select the appropriate prediction equation, it is recommended that coaches and practitioners make themselves aware of the population involved in the equation's development. Overestimations and underestimations of aerobic capacity may be detrimental for program design. This may lead to overtraining or deconditioning in athletes if the right estimation is not completed. For example, overestimations and underestimations of $\text{VO}_{2\text{max}}$ would result in prescribed training intensities at $\text{VO}_{2\text{max}}$ percentages higher or lower than an athlete should be training. Thus, intended adaptations may not persist due to prolonged training above or below necessary intensities. Practitioners should select an estimation equation that has been validated in a population as similar to their athletes and sport as possible. In order to achieve this, further research directed at the investigation of such equations in men and women athletes from different sports and training backgrounds is recommended.

Author Contributions: Conceptualization, J.B.W.; methodology, J.B.W.; formal analysis, J.J.M.; investigation, M.T.J., J.J.M., J.B.W.; resources, J.B.W. and M.T.J.; data curation, J.J.M., J.B.W.; writing—original draft preparation, M.K.M. and J.B.W.; writing—review and editing, J.B.W., J.J.M., M.K.M., M.T.J.; visualization, J.J.M.; supervision, J.B.W. and M.T.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The Ohio University Institutional Review Board for Human Subjects approved all procedures (IRB #12 × 059). The study was conducted according to the Declaration of Helsinki guidelines.

Informed Consent Statement: Risks and benefits were explained to athletes and an institutionally approved consent form was signed prior to participation.

Data Availability Statement: Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lemmink, K.A.P.M.; Visscher, S.H. Role of Energy Systems in Two Intermittent Field Tests in Women Field Hockey Players. *J. Strength Cond. Res.* **2006**, *20*, 682–688. [[PubMed](#)]
2. Lothian, F.; Farrally, M.R. A Comparison of Methods for Estimating Oxygen Uptake during Intermittent Exercise. *J. Sports Sci.* **1995**, *13*, 491–497. [[CrossRef](#)]
3. Spencer, M.; Lawrence, S.; Rechichi, C.; Bishop, D.; Dawson, B.; Goodman, C. Time-Motion Analysis of Elite Field Hockey, with Special Reference to Repeated-Sprint Activity. *J. Sports Sci.* **2004**, *22*, 843–850. [[CrossRef](#)]
4. Hinrichs, T.; Franke, J.; Voss, S.; Bloch, W.; Schänzer, W.; Platen, P. Total Hemoglobin Mass, Iron Status, and Endurance Capacity in Elite Field Hockey Players. *J. Strength Cond. Res.* **2010**, *24*, 629–638. [[CrossRef](#)]
5. Reilly, T.; Borrie, A. Physiology Applied to Field Hockey. *Sports Med.* **1992**, *14*, 10–26. [[CrossRef](#)] [[PubMed](#)]
6. Thoden, J. Testing aerobic power. In *Physiological Testing of the High-Performance Athlete*; Human Kinetics: Champaign, IL, USA, 1990; pp. 107–174.

7. Poole, D.C.; Jones, A.M. Measurement of the Maximum Oxygen Uptake $\dot{V}O_{2\max}$: $\dot{V}O_{2\text{peak}}$ Is No Longer Acceptable. *J. Appl. Physiol.* **2017**, *122*, 997–1002. [CrossRef]
8. Poole, D.C.; Wilkerson, D.P.; Jones, A.M. Validity of Criteria for Establishing Maximal O_2 Uptake during Ramp Exercise Tests. *Eur. J. Appl. Physiol.* **2008**, *102*, 403–410. [CrossRef]
9. Day, J.R.; Rossiter, H.B.; Coats, E.M.; Skasick, A.; Whipp, B.J. The Maximally Attainable $\dot{V}O_2$ during Exercise in Humans: The Peak vs. Maximum Issue. *J. Appl. Physiol.* **2003**, *95*, 1901–1907. [CrossRef]
10. Green, S.; Askew, C. $\dot{V}O_{2\text{peak}}$ Is an Acceptable Estimate of Cardiorespiratory Fitness but Not $\dot{V}O_{2\max}$. *J. Appl. Physiol.* **2018**, *125*, 229–232. [CrossRef] [PubMed]
11. Léger, L.A.; Lambert, J. A Maximal Multistage 20-m Shuttle Run Test to Predict VO_2 Max. *Eur. J. Appl. Physiol.* **1982**, *49*, 1–12. [CrossRef]
12. Aandstad, A.; Holme, I.; Berntsen, S.; Anderssen, S.A. Validity and Reliability of the 20 Meter Shuttle Run Test in Military Personnel. *Mil. Med.* **2011**, *176*, 513–518. [CrossRef]
13. Green, M.S.; Esco, M.R.; Martin, T.D.; Pritchett, R.C.; McHugh, A.N.; Williford, H.N. Crossvalidation of Two 20-M Shuttle-Run Tests for Predicting $\dot{V}O_{2\max}$ in Female Collegiate Soccer Players. *J. Strength Cond. Res.* **2013**, *27*, 1520–1528. [CrossRef]
14. Paliczka, V.J.; Nichols, A.K.; Boreham, C.A. A Multi-Stage Shuttle Run as a Predictor of Running Performance and Maximal Oxygen Uptake in Adults. *Br. J. Sports Med.* **1987**, *21*, 163–165. [CrossRef] [PubMed]
15. Ramsbottom, R.; Brewer, J.; Williams, C. A Progressive Shuttle Run Test to Estimate Maximal Oxygen Uptake. *Br. J. Sports Med.* **1988**, *22*, 141–144. [CrossRef]
16. Leger, L.; Gadoury, C. Validity of the 20 m Shuttle Run Test with 1 Min Stages to Predict $VO_{2\max}$ in Adults. *Can. J. Sport Sci.* **1989**, *14*, 21–26.
17. Chatterjee, P.; Banerjee, A.K.; Das, P.; Debnath, P.; Chatterjee, P. Validity of 20 Meter Multi Stage Shuttle Run Test for Prediction of Maximum Oxygen Uptake in Indian Female University Students. *Kathmandu Univ. Med. J. KUMJ* **2008**, *6*, 176–180. [PubMed]
18. St Clair Gibson, A.; Broomhead, S.; Lambert, M.I.; Hawley, J.A. Prediction of Maximal Oxygen Uptake from a 20-m Shuttle Run as Measured Directly in Runners and Squash Players. *J. Sports Sci.* **1998**, *16*, 331–335. [CrossRef] [PubMed]
19. Flouris, A.D. Enhancing the Efficacy of the 20 m Multistage Shuttle Run Test. *Br. J. Sports Med.* **2005**, *39*, 166–170. [CrossRef] [PubMed]
20. Léger, L.A.; Mercier, D.; Gadoury, C.; Lambert, J. The Multistage 20 Metre Shuttle Run Test for Aerobic Fitness. *J. Sports Sci.* **1988**, *6*, 93–101. [CrossRef] [PubMed]
21. Grant, J.; JOSEPH, A.; Campagna, P. The Prediction of $Vo_{2\max}$: A Comparison of 7 Indirect Tests of Aerobic Power. *J. Strength Cond. Res.* **1999**, *13*. [CrossRef]
22. Penry, J.T. Validity and Reliability Analysis of Cooper's 12-Minute Run and the Multistage Shuttle Run in Healthy Adults. Ph.D. Thesis, Oregon State University, Corvallis, OR, USA, 2008.
23. Merrigan, J.J.; Martin, J.R. Is the OUTPUT Sports Unit Reliable and Valid When Estimating Back Squat and Bench Press Concentric Velocity? *J. Strength Cond. Res.* **2021**. Publish Ahead of Print. [CrossRef]
24. Aziz, A.R.; Tan, F.H.Y.; Teh, K.C. A Pilot Study Comparing Two Field Tests with the Treadmill Run Test in Soccer Players. *J. Sports Sci. Med.* **2005**, *4*, 105–112. [PubMed]
25. Nassis, G.P.; Geladas, N.D.; Soldatos, Y.; Sotiropoulos, A.; Bekris, V.; Souglis, A. Relationship between the 20-m Multistage Shuttle Run Test and 2 Soccer-Specific Field Tests for the Assessment of Aerobic Fitness in Adult Semi-Professional Soccer Players. *J. Strength Cond. Res.* **2010**, *24*, 2693–2697. [CrossRef] [PubMed]
26. Flouris, A.D.; Koutedakis, Y.; Nevill, A.; Metsios, G.S.; Tsiotra, G.; Parasiris, Y. Enhancing Specificity in Proxy-Design for the Assessment of Bioenergetics. *J. Sci. Med. Sport* **2004**, *7*, 197–204. [CrossRef]
27. Datta, D. Deepankardatta/Blandr. Available online: <https://cran.r-project.org/web/packages/blandr/blandr.pdf> (accessed on 24 May 2021).
28. Ludbrook, J. Confidence in Altman–Bland Plots: A Critical Review of the Method of Differences. *Clin. Exp. Pharmacol. Physiol.* **2010**, *37*, 143–149. [CrossRef] [PubMed]
29. Ludbrook, J. Linear Regression Analysis for Comparing Two Measurers or Methods of Measurement: But Which Regression? *Clin. Exp. Pharmacol. Physiol.* **2010**, *37*, 692–699. [CrossRef] [PubMed]
30. Ludbrook, J. A Primer for Biomedical Scientists on How to Execute Model II Linear Regression Analysis. *Clin. Exp. Pharmacol. Physiol.* **2012**, *39*, 329–335. [CrossRef]
31. Warton, D.I.; Duursma, R.A.; Falster, D.S.; Taskinen, S. Smatr 3— an R Package for Estimation and Inference about Allometric Lines. *Methods Ecol. Evol.* **2012**, *3*, 257–259. [CrossRef]
32. Stojanovic, M.D.; Calleja-Gonzalez, J.; Mikic, M.; Madic, D.M.; Drid, P.; Vuckovic, I.; Ostojic, S.M. Accuracy and Criterion-Related Validity of the 20-M Shuttle Run Test in Well-Trained Young Basketball Players. *Montenegrin J. Sports Sci. Med.* **2016**, *6*, 5–10.
33. McGregor, S.J.; Nicholas, C.W.; Lakomy, H.K.; Williams, C. The Influence of Intermittent High-Intensity Shuttle Running and Fluid Ingestion on the Performance of a Soccer Skill. *J. Sports Sci.* **1999**, *17*, 895–903. [CrossRef]
34. Thompson, D.; Nicholas, C.W.; Williams, C. Muscular Soreness Following Prolonged Intermittent High-Intensity Shuttle Running. *J. Sports Sci.* **1999**, *17*, 387–395. [CrossRef] [PubMed]
35. Steinger, K.; Wodick, R.E. Sports-Specific Fitness Testing in Squash. *Br. J. Sports Med.* **1987**, *21*, 23–26. [CrossRef] [PubMed]