ISSN 2379-6391

SPORTS AND EXERCISE MEDICINE

OpenJournal

Openventio PUBLISHERS

Original Research

Agreement and Validity of Existing Anthropometric Calibration Models to Estimate Whole Body Density in Male Professional Football Players

Claire Mills, PhD[1*]; Mark De Ste Croix, PhD[2]; Stephen-Mark Cooper, PhD[1]; David James, PhD[1]

[1]*School of Sport and Exercise, University of Gloucestershire, Oxstalls Campus, Gloucester, GL2 9HW, UK*
[2]*Cardiff School of Sport and Health Sciences, Cardiff Metropolitan University, Cyncoed Campus, Cardiff, UK*

[*]***Corresponding author***
Claire Mills, PhD
Senior Lecturer, School of Sport and Exercise, University of Gloucestershire, Oxstalls Campus, Gloucester, GL2 9HW, UK; E-mail: clairem@glos.ac.uk

Article information
Received: January 12th, 2022; **Revised:** February 24th, 2022; **Accepted:** March 2nd, 2022; **Published:** March 5th, 2022

Cite this article
Mills C, De Ste Croix M, Cooper S-M, James D. Agreement and validity of existing anthropometric calibration models to estimate whole body density in male professional football players. *Sport Exerc Med Open J.* 2022; 8(1): 1-9. doi: 10.17140/SEMOJ-8-184

ABSTRACT

Introduction

To date no calibration models exist to estimate whole body density (Db) of professional footballers, as such the development of practical anthropometric calibration models to make sound body composition judgements is imperative. The aim of this study was to investigate the agreement and validity of estimating Db from 15 existing calibration models through comparison to a criterion method of hydrostatic weighing (HW) in male professional footballers.

Materials and Methods

Data were gathered from a total of n=206 professional football participants (x±s; age=24.1±5.4-years, body mass=78.8±8.4 kg, stature=180.1±7.0 cm, Db 1.075±0.01 g.ml^{-1}). Bland and Altman limits of agreement (LoA) approaches were used to determine bias and random variation derived from the calibration models.

Results

Bias and random errors for the published calibration models ranged from -0.005 to +0.015 g.ml^{-1} and 1.012 to 1.090 g.ml^{-1} respectively. A priori criterion (±3.8% p=<0.05 (g.ml^{-1})) was set as acceptable limits for the LoA method of which 13 calibration models found that (on average) estimated Db derived from HW was greater than Db derived from the models.

Discussion

A rank order of LoA identified the best model to use, however, LoA were not narrow enough for measurements to be of practical use and in most instances, selected models are not appropriate for estimating Db in male professional footballers.

Keywords

Whole body density; Calibration models; Anthropometry; Professional football players; Agreement; Validity.

INTRODUCTION

Given the accessibility of subcutaneous fat around the body, may be a reason which has led to the proliferation of formulae and equations to estimate whole body density (g.ml^{-1}) from various components of body composition.[1,2] The formulae are normally subdivided into regression equations generally developed on anthropometric based formulae that predicts the dependent variable (whole body density (g.ml^{-1})) from a series of independent variables such as body mass, stretched stature, skinfolds, girths, breadths, depths and widths.[2-4] Within scientific literature these regression equations are correctly termed calibration models, and the development of generalised calibration models provides a wealth of body composition information relating to different ages, sex and ethnicity.[4,5] From a sport science research perspective, questions remain about measurement validity issues and in retrospective can affect the confidence of what sport scientists have to say about their meaningfulness of their measurement data. Therefore, if these calibration models for the estimation of whole-body density are to be useful in a football or indeed a sporting context, their validity must be established. When further scrutinising the methodological approaches to the design and development of some of

© Copyright 2022 by Mills C. This is an open-access article distributed under Creative Commons Attribution 4.0 International License (CC BY 4.0), which allows to copy, redistribute, remix, transform, and reproduce in any medium or format, even commercially, provided the original work is properly cited.

the most popular calibration models, various limitations quickly become apparent. These limitations are no means exhaustive and could be replicated in other limitations, never-the-less they account for four main areas of concern:

Limitation 1: The Number of Variables Used as Individual Components within a Calibration Model

Questions have risen relating to the restrictive range of anthropometric measures used within a calibration model and in particular, the anthropometric site location.[4,6] What is evident is that of the most commonly used approach of four main skinfold sites of the biceps, triceps, suprailiac and subscapular which have been all or partly used within calibration models for young adult men.[7-10] The reason why these specific measures were used in their models is likely to be because of their impact on the estimation of whole body density (g.ml^{-1}). Another possible reason could be the recommendations from previous research indicating that the site location provides an accessibility advantage, although this is unclear.

Limitation 2: The Emphasis of How Variables are Used Interchangeably within a Calibration Model

In some instances, the manner in which the body composition variables are used interchangeably within the calibration model regression equation can provide an outcome with a different bias, which can result in significant errors in whole body density (g.ml^{-1}).[11-13] Examination of various calibration models found that some variables were provided as stand-alone outcomes, some as a combination of summed variables, some squared or even logged. What is often not discussed in calibration studies, is whether there have been any collinearity issues. As collinearity can potentially have an impact on the validity of the predictors in the models, what is concerning that given the wide range of anthropometric variables that are used within the development of these calibration models, that this has not been flagged as a concern or appear to be addressed by the authors in their findings.[14,15]

Limitation 3: The Sample Size Employed when Developing a Calibration Model

Too frequently the restrictive nature of the sample sizes raises concern over its practical use with a given population due to its predictive nature of the calibration model.[16,17] It is important to consider recommendations by Atkinson[17] and Sun et al[18] where the larger the sample size, the more statistical power, with recommendations to achieve 9 participants per variable.[17-19] Yet, literature investigated indicated that this this is not the case, thereby questioning the confidence and usefulness of such calibration models on a population of professional footballers.[18]

Limitation 4: The Use of Inappropriate Analytical Methods and the Lack of Cross-Validation Approaches When Designing a Calibration Model

If these calibration models are to be useful in a football context, their predictive accuracy for the estimation of whole body density must be established through careful examination using the most appropriate analysis methods.[6,20,21] The majority of the calibration studies investigated used stepwise linear regression analyses as a method of choice, and none of them employed the Bland et al[26] 95% limits of agreement analyses. Given that this approach was first introduced in the biostatistics literature by Altman and Bland[22] and recommended to the sport science community by Nevill et al[23] previous calibration models did not have access to these analyses. On another note, the calibration model(s) should be cross-validated by comparing values in a different sample of participants drawn from the population of interest, than those originally used to develop the calibration model, in order to test the accuracy of the prediction results.[4,5] Too frequently however, the sample sizes for cross-validation, and the range of variables considered, have been too restrictive to be effective indicators of the predictive nature of the existing calibration models.[16,17] Of the fourteen most commonly used calibration models within the literature, only the Jackson et al[24] model was found to have cross validated their model, where they used a sample of n=308 male participants to develop and n=95 to cross validate their model. Yet the remaining 13 calibration studies did not cross-validate their models and one plausible reason was due to their low sample sizes and cross-validation was not an option. What is disappointing is none of these studies discussed why cross-validation was not conducted, which in itself is a fundamental error.

With these limitations in mind and the number of calibration models available, it is still questionable to whether the majority of these calibration studies are robust when estimating whole body density (g. ml^{-1}) in professional footballers. Therefore, careful examination of the validity of the whole body densities predicted by these models is warranted. Hence, the aim of this study was to determine the agreement and subsequent validity from a range of previously published calibration models used to predict whole body density (g.ml^{-1}) determined from hydrostatic weighing as the criterion measurement method. The data entered into the models were gained from careful measurements with known reliability from Mills et al[25] study, the sample size was large (n=206 participants) and the criterion validity was determined with reference to Bland et al[26] 95% limits of agreement approach.

MATERIALS AND METHODS

Participants and Recruitment

Participants included two hundred and six Fédération Internationale de Football Association (FIFA) registered contracted professional football players (x±s; age=24.1±5.4-years, body mass=78.8±8.4 kg, stretched stature=180.1±7.0 cm and whole body density=1.075±0.010 (g.ml^{-1}) who were all over 18-years of age and free from disease. They were recruited from eight professional football clubs that represented Barclays Premiership, npower Championship, npower League One, npower League Two and Blue Square Premier Leagues during the 2007-2008, 2008-2009 and 2009-2010 playing seasons. All participants completed a health screening questionnaire and agreed to take part in the study by giving their informed consent. Ethical approval was granted from the University of Gloucestershire (UK) Research Ethics Committee (MIL-UOGREC/10).

Sport Exerc Med Open J. 2022; 8(1): 1-9. doi: 10.17140/SEMOJ-8-184

Procedure

This study needed to interrogate the degree to which models could potentially be applied to the present population of male professional football players. All potential calibration models were considered, and a selection criterion was put in place for accepting or rejecting them for further investigation (Table 1). Of the 18 models identified, 4 models were rejected on the basis that they required the chest skinfold to be measured, which is considered

Table 1. *Calibration Models for Consideration*

Author(s)	Sample No	Sample Characteristics	Country	Ethnicity	Skinfold Caliper	Age	Body Mass	Stretched Stature	Accept or Reject Model
Brožek et al[45]	159	University students	USA	Undisclosed	Undisclosed	x 20.4	x 69.1	x 177.8	Rejected
Pascale[46]	88	Soldiers	USA	Caucasian	Medical Nutrition	17.0-25.0	49.7-109.8	94.0-193.0	Rejected
Durnin et al[7]	60	Volunteers	Scotland	Undisclosed	Harpenden	18.1-33.8	43.6-95.6	154.8-192.0	Accept
Sloan[31]	50	University students	South Africa	Undisclosed	Medical Nutrition	18.0-26.0	57.8-85.7	163.0-191.0	Accept
Wilmore et al[32]	133	University students	USA	Undisclosed	Lange	16.8-36.8	53.2-121.2	159.0-193.4	Accept
Sloan et al[35]	50	Volunteers	Scotland	Undisclosed	Medical Nutrition	18.0-26.0	Undisclosed	Undisclosed	Accept
Forsyth et al[36]	50	University sports students	USA	Undisclosed	Lange	19.0-22.0	68.5-85.9	178.4-179.6	Accept
Forsyth et al[30]	50	University sports students	USA	Undisclosed	Lange	19.0-22.0	68.5-85.9	178.4-179.6	Accept
Katch et al[33]	53	University sports students	USA	Caucasian	Lange	18.0-21.0	62.8-80.0	169.4-183.4	Accept
Behnke et al[37]	54	University students	USA	Undisclosed	Undisclosed	Undisclosed	Undisclosed	Undisclosed	Accept
Durnin et al[8]	209	Sports enthusiasts	Scotland	Undisclosed	Harpenden	17.0-72.0	49.8-121.4	150.0-193.0	Accept
Wickkiser et al[38]	65	American footballers	USA	Undisclosed	Lange	x 17.2	x 88.0	x 182.5	Accept
Pollock et al[24]	95	Volunteers	USA	Undisclosed	Lange	18.0-22.0	74.6-82.2	179.6-179.8	Rejected
Jackson et al[24]	403	Volunteers	USA	Undisclosed	Lange	18.0-61.0	54.0-123.0	163.0-201.0	Accept
White et al[38]	58	American footballers	USA	Undisclosed	Undisclosed	x 19.9	x 89.7	x 182.0	Accept
Lohman[9]	61	University students	USA	Undisclosed	Undisclosed	Undisclosed	Undisclosed	Undisclosed	Accept
Thorland et al[10]	141	Athletes of national calibre	USA	Undisclosed	Lange	16.5-18.4	56.2-78.8	167.9-185.1	Accept
Withers et al[47]	207	State representatives	USA	Undisclosed	Harpenden	15.4-39.1	53.3-117.3	154.1-215.1	Rejected

Table 2. *Calibration Models for the Prediction of Body Density ($g.ml^{-1}$)*

Durnin et al[7]	Db=1.1610-0.0632 X Where X=$\log_{10}$ of the ∑of biceps, triceps, subscapular and suprailiac skinfolds
Sloan[31]	Db=1.1043-0.001327 (X_1)-0.001310 (X_2) Where X_1=front thigh skinfold, X_2=subscapular skinfold
Wilmore et al[32]	Db=1.08543-0.000886 (X_1)-0.00040 (X_2) Where X_1=abdominal skinfold, X_2=front thigh skinfold
Sloan et al[35]	Db=1.1043-0.00133 (X_1)-0.00131 (X_2) Where X_1=anterior thigh skinfold, X_2=subscapular skinfold
Forsyth et al[36]	Db=1.10647-0.00162 (X_1)-0.00144 (X_2)-0.00077 (X_3)+0.00071 (X_4) Where X_1=subscapular skinfold, X_2=abdominal skinfold, X_3=triceps skinfold, X_4=mid-axilla skinfold
Forsyth et al[30]	Db=1.03523-0.00156 (X_1)+0.00207 (X_2)-0.00140 (X_3) Where X_1=subscapular skinfold, X_2=biiliocristal breadth, X_3=abdominal skinfold
Katch et al[33]	Db=1.09665-0.00103 (X_1)-0.00056 (X_2)-0.00054 (X_3) Where X_1=triceps skinfold, X_2=subscapular skinfold, X_3=abdominal skinfold
Behnke et al[37]	Db=1.08543-0.00086 (X_1)-0.00040 (X_2) Where X_1=abdominal skinfold, X_2=anterior thigh skinfold
Durnin et al[8]	Db=1.1765-0.0744 ($\log_{10}X_1$) Where X_1=∑4 skinfolds (triceps, biceps, subscapular and iliac crest)
Wickkiser et al[38]	Db=1.10148-0.00118 (X_1)-0.00114(X_2)+0.00044 (X_3) Where X_1=waist circumference, X_2=triceps skinfold, X_3=stretched stature
Jackson et al[24]	Db=1.0982-0.000815 (X) +0.0000084 $(X)^2$ Where X=∑3 skinfolds (triceps, abdomen and subscapular)
White et al[38]	Db=1.0958-0.00088 (X_1)-0.00060 (X_2) Where X_1=suprailiac skinfold, X_2=anterior thigh skinfold
Lohman[9]	Db=1.1091-0.00052 (X_1) +0.00000032 $(X_1)^2$ Where X_1=∑7 skinfolds (triceps, subscapular, mid-axilla, iliac crest, abdominal, front thigh and medial calf)
Thorland et al[10]	Db=1.0988-0.0004 (X_1) Where X_1=∑7 skinfolds (triceps, subscapular, biceps, supraspinale, abdominal, front thigh and medial calf)

outdated in relation to International Society for the Advancement of Kinanthropometry (ISAK) accreditations. Subsequently, 14 calibration models and the process *via* air displacement plethysmography were accepted to predict whole body density (g.ml^{-1}) in professional football players.

Table 2 provides the components of the calibration model regression equations for the accepted pre-published calibration models used in this study. It is important to note that none of the models were designed specifically for use within professional football players.

Statistical Analysis

In order to establish the requisite indices of validity for each calibration model for participants whole body density (g.ml^{-1}), there were some major elements to be determined: (i) the systematic bias in the whole body density (g.ml^{-1}) values between each calibration model, (ii) establish the degree of random variation, (iii) the degree of heteroscedasticity (the condition of un-equal residual variance) which would be investigated by the application of the Bland et al[26] 95% limits of agreement method, finally, (iv) to identify whether the quantification of agreement was narrow enough for the whole body density (g.ml^{-1}) to be providing practically valid values, thereby, would it have any detrimental effect of the practical use for this population of participants. Therefore a priori was established for the Bland and Altman 95% LoA method that presented acceptable tolerable limits within the context of this study and set at ±3.8% (p<=0.05).[25,27-29] For instance, whole body density of 1.075 g.ml^{-1} could be considered average within the context of this study, therefore ±3.8% acceptable limit, whole body density ranged from 1.034 to 1.116 g.ml^{-1} in the studied population. Reports from Mayhew et al,[20] Cooper[21] and Heyward[6] suggest that too high whole body density (g.ml^{-1}) could pose a possible impact on team selection, conversely too low could be deemed a definite danger to the health and wellbeing of the participant if their training prescription is not carefully considered. When illustrated in this manner, it is clear that the acceptable limits can be used interchangeably with the criterion measurement method to assess whole body density (g.ml^{-1}) in professional football players.

RESULTS

Table 3 provides an overview of bias, random variation and heteroscedasticity of 95% limits of agreement for all calibration models. Inspection indicated that in 13 of the calibration models (on average) whole body density (g.ml^{-1}) derived from hydrostatic weighing, was greater than whole body density (g.ml^{-1}) derived from the models, so there was a positive bias.

Results from the 95% limits of agreement analyses indicated bias (systematic errors) between criterion measured whole body density (g.ml^{-1}) and whole body densities predicted by calibration models ranged from 0.005 to 0.009 g.ml^{-1} and random errors ranged from 1.012 to 1.079 g.ml^{-1}. Thirteen (13) of the 15 calibration models were normally distributed and within acceptable limits thus suggesting underestimation of whole body densities (g.ml^{-1}) of professional football players. In contrast, 2 of the calibration models by Forsyth et al[30] and Durnin et al[8] found that (on average) whole body density (g.ml^{-1}) was lower than whole body density (g.ml^{-1}) derived from the models, so had a negative bias. Results from the 95% limits of agreement analyses indicated negative bias (systematic errors) between criterion and densities predicted by calibration models ranged from 0.009 to 0.015 g.ml^{-1} and random errors ranged from 1.027 to 1.090 g.ml^{-1} thereby indicating overestimation of whole body density (g.ml^{-1}) of professional football players.

Figure 1 exhibits the scatter plots of heteroscedasticity to demonstrate the relationship between the criterion method of hydrostatic weighing and each calibration model. The extent to which heteroscedasticity is present in these scores can be quantified by correlating absolute differences against mean scores for calibration model and criterion method values and can be illustrated on a scatter plot of these two variables. Heteroscedasticity for all calibration models demonstrate r values ranging from -0.064 to 0.374% and R^2 (%) coefficients ranging from 0.00000005 to 0.1398%. All models illustrated heteroscedastic data with deviations from the line of identity between the whole body density (g.ml^{-1}) values.

Overall, 7 models indicated statistical significance of p=0.01[7,24,31-34] compared to 8 models which indicated a statistical significance of p=0.05.[8-10,30,35-38] Wider deviation from the line of identity was particularly prevalent with the model developed by Sloan et al[35] and Jackson et al[24] and provided deviation from the line of identity and demonstrated heteroscedasticity. Furthermore, the air displacement plethysmography demonstrates an ordinal scale visual that discriminates intervals between the range, thereby not reflecting visuals like the remaining calibration models and as such can be seen as a deviation.

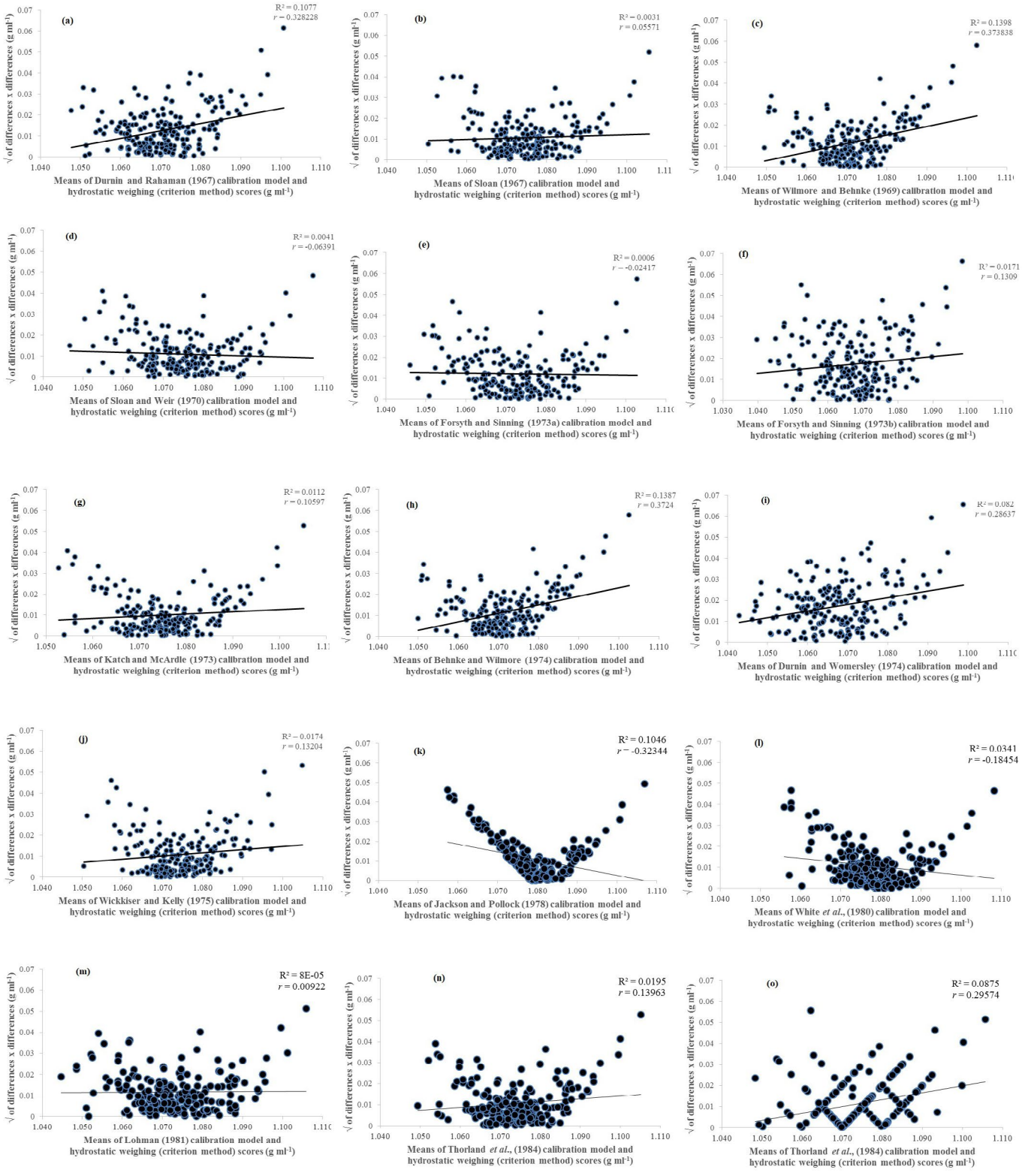
At this point of statistical analysis, it was worth considering the issue of heteroscedasticity and whether there is a need to find the cause and resolve it. Given that the Jackson et al[24] was the only calibration model which demonstrated heteroscedasticity from Figure 1, a judgement was made not to log transform and therefore keep the data in its present condition. As there can be obvious patterns of distribution within each case, it is therefore important to obtain some clarity over the most appropriate calibration model to use. To avoid repetition when reporting outcomes, Table 4 reports the hypothetical predictions of whole body density (g.ml^{-1}) of 1.045 g.ml^{-1} for each model and presented in rank order from lowest to highest in terms of differences of agreement indicators.

Examination of the calibration models indicated that Behnke et al[37] (lowest) and Forsyth et al[30] (highest) models found there is a bias of 0.007 g.ml^{-1} and 0.015 g.ml^{-1} and 95% limits of agreement of 0.019 g.ml^{-1} to 0.033 g.ml^{-1} and 0.015 g.ml^{-1} to 0.045 g.ml^{-1} respectively.

Based on these findings, the issue is whether an error of this magnitude would detrimentally affect anything this research has to say about the participants' whole body density (g.ml^{-1}) derived from the chosen calibration models. In other words, could this research replace hydrostatic weighing method with the cali-

Sport Exerc Med Open J. 2022; 8(1): 1-9. doi: 10.17140/SEMOJ-8-184

Figure 1. *Scatter Plots for the Heteroscedasticity of Hydrostatic Weighing (criterion method) Compared to (a) Durnin et al,[7] (b) Sloan,[31] (c) Wilmore et al[32] (d) Sloan et al[35] (e) Forsyth et al,[36] (f) Forsyth et al,[30] (g) Katch et al,[33] (h) Behnke et al,[37] (i) Durnin et al,[8] (j) Wickkiser et al,[38] (k) Jackson et al,[24] (l) White et al,[34] (m) Lohman,[9] (n) Thorland et al,[10] and (o) Air Displacement Plethysmography Calibration Model (means) for Whole Body Density (g.ml^{-1})*



Sport Exerc Med Open J. 2022; 8(1): 1-9. doi: 10.17140/SEMOJ-8-184

bration model. It is important to stress that it is unlikely that the calibration models will agree exactly for estimated whole body density (g.ml^{-1}) amongst themselves, but these findings suggest a need to provide an explanation as to why 13 calibration models (on average) under-reported and two calibration models (on average) over-reported body density for a group of professional footballers. The statistics provided here cannot answer this question. Indeed, the question that this research needed to ask is, are the 95% limits of agreement narrow enough for measurements to be of practical use.

Whilst reviewing the literature, various limitations with the design and development of their calibration models became apparent. Five main limitations were identified. Firstly, the most frequently used measures that were used as part of the components of the calibration models included the subscapular, triceps, supraspinale, iliac crest, abdominal and anterior thigh skinfolds, biacromial and biiliocristal breadths and chest and waist girths (Table 2). Yet further scrutiny found that half of the models did to incorporate of any limb measures which failed to support empirical studies and recommendations in the literature about the value that skinfolds from the lower limb account for a significantly greater proportion of variance in body fat.[39-41] Secondly, in some instances some variables were provided as stand-alone outcomes some as a combination of summed variables, some squared or even logged (Table 2), thereby providing the outcome with a different bias and if used indiscriminately, the strength of the outcome is lost.[11-13,42] Thirdly, too frequently the sample sizes have been too restrictive to be effective indicators of the predictive nature of the existing calibration model and therefore raises concern over its practical use.[16,17] Ten studies investigated had less than n=100 participants in their sample and in some instances as low as n=50 participants. Fourthly, evidence that cross-validation has either been ignored or used on very low restrictive sample numbers can raise doubt over the models specificity and validity.[4] Examination of the 14 published calibration models found that 10 failed to cross-validate their data with another sample from the population of interest, 3 cross-validated but with other published calibration models and only one model[24] cross-validated their data. The fifth limitation was arguably the inappropriate analytical methods to develop these models. Correlation coefficients and linear regression methods as indices of the equations' validity and decisions based upon outcomes generated from employing the 95% limits of agreement analyses are the preferred analysis of choice. Furthermore, the decision to log transform heteroscedastic data has also become an area of debate, particularly when considering whether there is a need to find the cause and resolve heteroscedasticity.[43]

DISCUSSION

The aim of the present study was to gain some insight into the validity of estimating whole body density (g.ml^{-1}) from 15 calibration models that already exist in the public domain by comparing them to those gathered from the criterion hydrostatic weighing method in a large sample of professional footballers. The Bland and Altman 95% limits of agreement approaches findings were too wide to state with authority that these calibration models can be used interchangeably with the criterion method to assess whole body density in professional football players. Thirteen (13) calibration models found that (on average) estimated whole body density (g.ml^{-1}) derived from hydrostatic weighing was greater than whole body density (g.ml^{-1}) derived from the models. Bias ranged from -0.005 to +0.009 g.ml^{-1} and random errors ranged from 1.012 to 1.079 g.ml^{-1}. Furthermore, Bland et al[26] 95% limits of agreement approaches were used to determine heteroscedasticity. Of the 15 calibration models used within this study only Jackson et al[24] model illustrated heteroscedasticity with r values of -0.323 and R^2 (%)

Table 3. *Overview of 95% Upper and Lower Limits of Agreement and Bias Indicators for (n=14) Calibration Models Plus Air Displacement Method*

Calibration Models	Bias (g.ml^{-1})	(95 LoA)	
		Lower Limit (g.ml^{-1})	Upper Limit (g.ml^{-1})
Durnin et al[7]	+0.009	-0.018	+0.036
Sloan[31]	+0.000	-0.027	+0.028
Wilmore et al[32]	+0.008	-0.019	+0.034
Sloan et al[35]	+0.001	-0.027	+0.029
Forsyth et al[36]	+0.006	-0.023	+0.034
Forsyth et al[30]	+0.015	-0.015	+0.045
Katch et al[33]	+0.001	-0.026	+0.027
Behnke et al[37]	+0.007	-0.019	+0.033
Durnin et al[8]	+0.015	-0.012	+0.043
Wickkiser et al[38]	+0.001	-0.028	+0.029
Jackson et al[24]	-0.006	-0.033	+0.021
White et al[38]	-0.005	-0.031	+0.021
Lohman[9]	+0.004	-0.024	+0.032
Thorland et al[10]	+0.003	-0.024	+0.030
Air displacement plethysmography	+0.004	-0.024	+0.033

Table 4. *Rank order of 95% Upper and Lower Limits of Agreement and Differences Indicators for (n=14) Calibration Models Plus Air Displacement Method for Whole Body Density of 1.045 (g.ml^{-1})*

Calibration Models	Lower Limit (g.ml^{-1})	Upper Limit (g.ml^{-1})	Difference (g.ml^{-1})
Behnke et al[37]	1.026	1.078	0.052
White et al[38]	1.014	1.066	0.052
Wilmore et al[32]	1.026	1.079	0.053
Katch et al[33]	1.019	1.072	0.053
Durnin et al[7]	1.027	1.081	0.054
Jackson et al[24]	1.012	1.066	0.054
Thorland et al[10]	1.021	1.075	0.054
Sloan[31]	1.018	1.073	0.055
Durnin et al[8]	1.033	1.088	0.055
Sloan et al[35]	1.018	1.074	0.056
Lohman[9]	1.021	1.077	0.056
Air displacement plethysmography	1.021	1.078	0.057
Wickkiser et al[38]	1.017	1.074	0.057
Forsyth et al[36]	1.022	1.079	0.057
Forsyth et al[30]	1.030	1.090	0.060

coefficients of 0.1046% and p=0.01.

Given that the Jackson et al[24] was the only calibration model which demonstrated heteroscedasticity from Figure 1, a decision was needed whether to log transform all data. As there was no measurement error or reliability issues across the variables or calibration models, a judgement was made not to log transform and therefore keep the data in its present condition. Due to the obvious patterns of distribution for whole body density ($g.ml^{-1}$) from 15 calibration models, a hypothetical whole body density value of 1.045 $g.ml^{-1}$ was applied for each individual calibration model *via* limits of agreement. A rank order of 95% upper and lower limits was determined to provide an overview that would best identify the best model to use for the current population of professional footballers. The model developed by Forsyth et al[30] was considered the best where bias ranged from -0.015 to +0.045 $g.ml^{-1}$ with whole body density values ranging from 1.025 to 1.081 $g.ml^{-1}$.

Whilst reviewing the literature, it became apparent that the researchers cited had various limitations with the design and development of their calibration models. Nonetheless, what is questionable is the reason why the literature does not have any calibration models to estimate whole body density ($g.ml^{-1}$) available for professional football players, especially given the popularity and income generation with the sport. Generally speaking, results indicated that most models that were used had high reliability values, but exploitation of whole body density values occurred with severe underestimation of whole body density ($g.ml^{-1}$) of professional footballers, thereby raising doubt over the validity of such calibration models for use within the current population. Therefore, sport scientists that are cognisant that calibration models have the potential to provide an insight into players' body composition, thus contributing towards the optimisation of performance potential.[44] Yet, in reality there are no hard-set rules on what calibration models to use for a specific population sample, but care must be taken by the sport scientist when selecting an appropriate calibration model for application in a professional football context by considering the most appropriate criteria about what constitutes practical significance. For instance, by considering (i) circumstances of initial validation and any collinearity issues (ii) a large enough sample size even after the division of gender and age groupings and (iii) at least 9 variables per participant.[6,20,21]

In essence, the 95% limits of agreement of this study's findings were not narrow enough for measurements to be of practical use. In most instances, the error (the disagreement) was too great, and as such it would be detrimental to what this study can conclude about professional players' whole body density ($g.ml^{-1}$). Therefore, there is a need to develop a sport specific calibration model for use with professional football players which is both warranted and desirable.[2,42] Moreover, new models should ideally be based on measures from large sample sizes and these should also include the entire playing spectrum of football players and above all, decisions about the calibration model's specificity should be established with reference to sound research principles such as cross-validation procedures.

DECLARATION OF INTEREST STATEMENT

N/A.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

1. Clark RR, Kuta JM, Sullivan JC. Prediction of percent body fat in adult males using dual energy x-ray absorptiometry, skinfolds, and hydrostatic weighing. *Med Sci Sports Exerc.* 1993; 25(4): 528-535. doi: 10.1249/00005768-199304000-00017

2. Provyn S, Scafoglieri A, Tresignie J, Lumé C, Clarys JP, Bautmans I. Critical appraisal of selected body composition data acquisition techniques in public health. In: Maddock J, ed. *Public Health-Social and Behavioural Health*. Rijeka, Croatia: IntechOpen; 2012: 133-152.

3. Russo GE, Gruppioni G, Gueresi P, Belcastro MG, Marchesini V. Skinfolds and body composition of sports participants. *J Sports Med Phys Fitness.* 1992; 32(3): 303-313.

4. Atkinson G, Nevill AM. Selected issues in the design and analysis of sport performance research. *J Sports Sci.* 2001; 19: 811-827. doi: 10.1080/026404101317015447

5. Vincent WJ. *Statistics in Kinesiology*. Champaign, IL, USA: Human Kinetics; 1995.

6. Heyward VH. Nutrition and athletic performance: A position of the American Dietetic Association, Dietitians of Canada, and the American College of Sports Medicine. *J Am Diet Assoc.* 2000; 100 (12): 1543-1556.

7. Durnin JVGA, Rahaman MM. The assessment of the amount of fat in the human body from skinfold thickness. *Br J Nutr.* 1967; 21(3): 681-689. doi: 10.1079/bjn19670070

8. Durnin JVGA, Womersley J. Body fat assessed from total body density and its estimation from skinfold thickness: Measurements on 481 men and women aged from 16 to 72 years. *Br J Nutr.* 1974; 32(1): 77-97. doi: 10.1079/bjn19740060

9. Lohman TG. Skinfolds and body density and their relation to body fatness: A review. *Hum Biol.* 1981; 53(2); 181-221.

10. Thorland WG, Johnson GO, Tharp GD, Fagot TG, Hammer RW. Validity of anthropometric equations for the estimation of body density in adolescent athletes. *Med Sci Sports Exerc.* 1984; 16(1): 77-81.

11. Ball SD, Altena TS, Swan PD. Comparison of anthropometry to DXA: A new prediction equation for men. *Eur J Clin Nutr.* 2004; 58(11): 1525-1531. doi: 10.1038/sj.ejcn.1602003

12. Ishiguro N, Kanehisa H, Miyani M, Masuo Y, Fukunaga T. A

comparison of three biological impedance analyses for predicting lean body mass in a population with a large difference in muscularity. *Eur J Appl Physiol.* 2005; 94(1-2): 25-35. doi: 10.1007/s00421-004-1259-2

13. Peeters MW, Goris M, Keustermans G, Pelgrim K, Claessens AL. Body composition in athletes: A comparison of densitometric methods and tracking of individual differences. *European Journal of Sport Sciences.* 2013; 13(1): 78-85. doi: 10.1080/17461391.2011.606836

14. Mason CH, Perreault WD. Collinearity, power, and interpretation of multiple regression analysis. *Journal of Marketing Research.* 1991; 28(3): 268-280. doi: 10.2307/3172863

15. Tormod N, Bjørn-Helge M. Understanding the collinearity problem in regression and discriminant analysis. *Journal of Chemometrics.* 2001; 15(4); 413-426. doi: 10.1002/cem.676

16. Hawes MR, Martin AD. Human body composition. In: Eston R, Reilly T, eds. *Kinanthropometry and Exercise Physiology Laboratory Manual.* 2nd ed. London, England: E. and F.N. Spon; 2001: 1-46.

17. Atkinson G. What is this thing called measurement error? In: Reilly T, Marfell-Jones M, eds. Paper presented at: Kinanthropometry VIII: Proceedings of the 8th International Conference of the International Society for the Advancement of Kinanthropometry; 2003; London, England: Taylor and Francis. 2003. 1-12.

18. Sun SS, Chumlea WC. Statistical methods. In: Heymsfield SB, Lohman TG, Wang Z, Going SB, eds. *Human Body Composition.* 2nd ed. Champaign, IL, USA: Human Kinetics; 2005.

19. Cohen J. *Statistical Power Analysis.* 2nd ed. New Jersey, NY, USA: Lawrence Erlbaum Associates; 1988.

20. Mayhew JL, Clarke BA, McKeown BC, Montaldi DH. Accuracy of anthropometric equations for estimating body composition in female athletes. *J Sports Med Phys Fitness.* 1985; 25(3): 120-126.

21. Cooper S-M. Consistency of anthropometric predictive equations for the assessment of body composition characteristics in adolescent Rugby Union players. *Journal of Sports Sciences.* 1995; 13(6): 499-500.

22. Altman DG, Bland MJ. Commentary on quantifying agreement between two methods of measurement. *Clinical Chemistry.* 2002; 48(5): 801-802. doi: 10.1093/clinchem/48.5.801

23. Nevill AM, Atkinson G. Assessing agreement between measurements recorded on a ratio scale in sports medicine and sports science. *Br J Sports Med.* 1997; 31(4): 314-318. doi: 10.1136/bjsm.31.4.314

24. Jackson AS, Pollock ML. Generalized equations fat predictions of body density of men. *Br J Nutr.* 1978; 40(3): 497-504. doi: 10.1079/BJN19780152

25. Mills CD, Cooper S-M, Croix MDS, James DVB. Measurement reliability and precision of anthropometric measures for the estimation of whole body density in professional football players. *Open Access Journal of Biomedical Science.* 2020; 3(1): 658-666. doi: 10.38125/OAJBS.000227

26. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986; 1: 307-310.

27. International Society for the Advancement of Kinanthropometry (ISAK). Paper presented at: International Standards for Anthropometric Assessment. 2001; Sydney, Australia.

28. Ludbrook J. Confidence in Altman-Bland plots: A critical review of the method of differences. *Clin Exp Pharmacol Physiol.* 2010; 37(2): 143-149. doi: 10.1111/j.1440-1681.2009.05288.x

29. Woodman RJ. Bland-Altman beyond the basics: Creating confidence with badly behaved data. *Clin Exp Pharmacol Physiol.* 2010; 37(2): 141-142. doi: 10.1111/j.1440-1681.2009.05320.x

30. Forsyth HL, Sinning WE. The anthropometric estimation of body density and lean body weight of male athletes. *Med Sci Sports.* 1973b; 5(3): 174-180.

31. Sloan AW. Estimation of body fat in young men. *J Appl Physiol.* 1967; 23(3): 311-315. doi: 10.1152/jappl.1967.23.3.311

32. Wilmore JH, Behnke AR. An anthropometric estimation of body density and lean body weight in young men. *J Appl Physiol.* 1969; 27(1): 25-31. doi: 10.1152/jappl.1969.27.1.25

33. Katch FI, McArdle WD. Prediction of body density from simple anthropometric measurements in college-age men and women. *Hum Biol.* 1973; 45(3): 445-451.

34. White J, Mayhew JL, Piper FC. Prediction of body composition in college football players. *J Sports Med Phys Fitness.* 1980; 20(3): 317-324.

35. Sloan AW, Weir JB. Nomograms for prediction of body density and total body fat from skinfold measurements. *J Appl Physiol.* 1970; 28(2): 221-222. doi: 10.1152/jappl.1970.28.2.221

36. Forsyth HL, Sinning WE. The anthropometric estimation of body density and lean body weight of male athletes. *Med Sci Sports.* 1973a; 5(3): 174-180.

37. Behnke AR, Wilmore JH. *Evaluation and Regulation of Body Build and Composition (International Research Monograph Series in Physical Education).* New Jersey, USA: Prentice-Hall Inc; 1974.

38. Wickkiser J, Kelly J. The body composition of a college football team. *Med Sci Sports.* 1975; 7(3): 199-202.

39. Clarys JP, Martin AD, Drinkwater DT, Marfell Jones MJ. The skinfold: Myth and reality. *J Sports Sci.* 1987; 5(1): 3-33. doi:

10.1080/02640418708729760

40. Eston RG. Prediction of body fat from skinfolds: The importance of including sites from the lower limb. *J Sports Sci.* 2003; 21: 369-370. doi: 10.1080/0264041031000071236

41. Bellisari A, Roche AF. Anthropometry and Ultrasound. In: Heymsfield SB, Lohman TG, Wang Z, Going SB, eds. *Human Body Composition.* 2nd ed. Champaign, IL, USA: Human Kinetics; 2005: 109-127.

42. Guo SS, Siervogel RM, Chumlea CWM. Epidemiological applications of body composition. The effects and adjustment of measurement errors. *Annals of the New York Academy of Sciences.* 2000; 904(1): 312-316.

43. Manning WG, Mullahy J. Estimating log models: To transform or not to transform? *J Health Econ.* 2001; 20: 461-494. doi: 10.1016/s0167-6296(01)00086-8

44. Hencken CL. Anthropometric measurement in elite football players. *J Sports Sci.* 2004; 22(3): 266-267.

45. Brozek J, Keys A. The evaluation of leanness-fatness in man; norms and interrelationships. *Br J Nutr.* 1951; 5(2): 194-206. doi: 10.1079/bjn19510025

46. De Pascale A. Values for protein-bound iodine in the blood of infants. *Lattante.* 1956; 27: 84-91.

47. Withers RT, Craig NP, Bourdon PC, Norton KI. Relative body fat and anthropometric prediction of body density of male athletes. *Eur J Appl Physiol Occup Physiol.* 1987; 56(2): 191-200. doi: 10.1007/BF00640643