UNIVERSITY OF
GLOUCESTERSHIRE

This is a peer-reviewed, post-print (final draft post-refereeing) version of the following published document, © 2023 Informa UK Limited, trading as Taylor & Francis Group. This is an Accepted Manuscript of an article published by Taylor & Francis in Science and Medicine in Football on 17/01/2023, available at: http://www.tandfonline.com/10.1080/24733938.2023.2166680 and is licensed under Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0 license:

Mills, Claire D ORCID logoORCID: https://orcid.org/0000-0003-4156-4593, De Ste Croix, Mark B ORCID logoORCID: https://orcid.org/0000-0001-9911-4355, James, David V ORCID logoORCID: https://orcid.org/0000-0002-0805-7453 and Cooper, Stephen-Mark (2024) Development of novel calibration model(s) to predict whole-body density in professional football players. Science and Medicine in Football, 8 (2). pp. 170-178. doi:10.1080/24733938.2023.2166680

Official URL: http://doi.org/10.1080/24733938.2023.2166680
DOI: http://dx.doi.org/10.1080/24733938.2023.2166680
EPrint URI: https://eprints.glos.ac.uk/id/eprint/12265

Disclaimer

The University of Gloucestershire has obtained warranties from all depositors as to their title in the material deposited and as to their right to deposit such material.

The University of Gloucestershire makes no representation or warranties of commercial utility, title, or fitness for a particular purpose or any other warranty, express or implied in respect of any material deposited.

The University of Gloucestershire makes no representation that the use of the materials will not infringe any patent, copyright, trademark or other property or proprietary rights.

The University of Gloucestershire accepts no liability for any infringement of intellectual property rights in any material deposited but will remove such material from public view pending investigation in the event of an allegation of any such infringement.

PLEASE SCROLL DOWN FOR TEXT.

Development of novel calibration model(s) to predict whole-body density in professional football players

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

[1] School of Sport and Exercise, University of Gloucestershire, UK and [2] Cardiff School of Sport and Health Sciences, Cardiff Metropolitan University, UK.

ABSTRACT

Introduction: Questions continue to be raised about the validity that are in existence to estimate D_b, in professional male footballer players. **Methods:** Phase1: n = 28 anthropometric variables were used on n = 206 footballers, using regression analyses to determine SEE and R^2. A cut-off correlation coefficient set at r = 0.950 and 90% R^2. Phase2: all variables (z-scores, $\bar{x}$ = 0.0, SD = ± 1.0) to help reduce heteroscedasticity, β, r, t, significance of t and P-values were calculated. Phase3: a forced stepwise – backwards regression analysis approach with 9 predictors which met the acceptance criteria (r = 0.950, R^2 = 90%, and β weights) was used to develop a '*best fit*' and a '*practical*' calibration model. Phase4: cross-validation of the 2 newly developed calibration method using LoA. **Results:** The 'best fit' model SEM (0.115 g ml^{-1}), the highest R^2 (6.6%) (P = < 0.005), whereas the 'practical' calibration model SEM (0.115 g ml^{-1}), R^2 (4.7%) (P = < 0.005) with r values = 0.271 and 0.596 and R^2 (%) coefficients = 0.3526 for the 'best fit' and 'practical' calibration models respectively (P = 0.01). **Conclusions:** The 2 calibration models supported an ecologically and statistically valid contribution and can provide sound judgements about professional footballers' body composition.

Key words: anthropometry · calibration models · whole body density · professional football players · validity · cross validation

INTRODUCTION

There is a plethora of calibration models that exist in the literature to estimate various components of body composition, although questions have been raised about the validity of such models[1]. Since publication, researchers have identified limitations that can have an impact on the estimation of whole-body density (g ml^{-1}) when applied to a specific professional football population[2,3,4,5]. Valid and reliable estimations of whole body density are the cornerstone of understanding the other more commonly referred to characteristics of body composition like %body fat, fat mass, fat-free mass, lean body mass, and minimal body mass[1,2,3]. The tables that coaches and sport scientists might use to convert total skinfold thicknesses into percentage body fat have their genesis in estimations of whole body density, and if players, coaches, and support staff know this, and how it is derived, they will better understand the characteristics of body composition[4,5]. Understanding that fractionalisation of total body mass into its fat (storage and essential depots) and fat-free (muscle, bone, and residual) and that essential fat plus the fat-free components are equivalent to the lean body mass which is the functional characteristic of body composition that will impact football performance – all fundamentally derived from valid estimations of whole body density[4,5].

Findings from Mills *et al*.,[4] study have indicated that published calibration models had significant differences of under estimation of whole body density in professional footballers[6,7,8,9,10,11,12,13,14,15] and over estimation of whole body density[8,16,17]. These inconsistencies may be one reason why different calibration models produce different body densities on the same participant[18]. Fortunately attempts have been made to cross validate previously published calibration models for the estimation of body composition parameters specifically on football populations[19,20,21,22]. However, results indicated that although these models have high measurement reliability, exploitation of whole-body density values with severe underestimation, will not provide for accurate monitoring of professional football players body composition changes during training[2,5]. It is no surprise therefore, that research has been on the increase to develop population specific calibration models for various populations, ages, sports and levels of activity[2,14]. These population specific approaches have helped to contribute to increasing understanding of body composition in relation to health, fitness, sport, exercise, growth and the ageing

process[14]. Consequently, a newly developed calibration model(s) should be cross-validated by establishing how well the predicted values agree with measured criterion values in a different sample of participants from that used to develop the calibration model[5]. Yet, evidence has found many pre-published calibration models are strictly speaking only calibration studies, where, controversially, original authors did not cross-validate values generated by their calibration models with those from a different sample of participants that were used to develop the model[5,23]. Indeed, some authors have even called for a halt to the development of new calibration models unless they are cross-validated in order to test the validity of the prediction results[23]. Given that there are no specific calibration models that exist in the literature to estimate whole body density in professional football players, evidence suggests that the development of a model(s) with cross-validation techniques can provide sport scientists with an essential mechanism for making sound body composition judgements for the football profession[5,24,25]. Thus, the main aims of this study were to develop 2 separate calibration models to estimate whole body density (g ml^{-1}) in professional football players and to cross-validate the models to determine validity.

METHODS

Two hundred and six Fédération Internationale de Football Association (FIFA) registered contracted professional football players (Goalkeepers n = 14; Defenders n = 67; Midfielders n = 70 and Strikers n = 55) ($\bar{x} \pm s$; 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}) were recruited from 8 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. G* Power, a post hoc power analysis on the correlation and regression coefficients were utilised to show adequate power and thereby adequate sample size(s) (95% confidence level with n = 206 margin of error 5% ideal sample size of n = 135). Sampling included players who were all over 18 years of age, free from disease or illness and who agreed to act as participants for the study by giving their written informed consent. Signs and symptoms of disease and diagnosed disease were determined

through health screening questionnaire and ethical approval was granted (MILLS_SREC09) from the University of Gloucestershire's Research Ethics Committee.

All participants were asked to arrive at the sports science laboratories at the University of Gloucestershire at least 1 hour before testing was to begin within the first half of the playing season. Assessments were conducted in the mornings so that the primary investigator could control for diurnal fluctuations and that participants could more easily adhere to the following strict standardised pre-testing procedures, which required all participants to: (a) refrain from consuming food or fluid for at least 4 hours before assessment (b) refrain from exercising for a 12 hour period before assessment (c) refrain from smoking for at least 4 hours before assessment (d) to empty their bowel and bladder before assessment (e) to wear light fitting shorts or underpants and (f) to remove all jewellery. In advance of any testing procedures, a health questionnaire and consent form were read, dated, and signed by the participant and counter–signed by the primary investigator. Before testing, a thorough verbal explanation of the study's aims, duration (~ 1.5 hours), visual demonstration of all procedures, consequences of the research and how the results were likely to be disseminated to each participant. Furthermore, they were asked to comment on whether they had an injury(s), bruising, swelling, scaring or muscle atrophy which might impede accurate measurement. If necessary, the injury(s) were documented, and it was also noted whether participants had excessive body hair and/or facial hair. All measures were recorded in sequential order as listed on Kinanthropometric data collection proforma (a) anthropometric measurements (b) forced vital capacity (c) air displacement measurements and (d) hydrostatic weighing. Hydrostatic weighing measurements were conducted last, to ensure that the participant's skin was dry and lotion free.

Phase1 statistical analyses, a total of n = 28 anthropometric variables from Mills *et al*.,[4] study (8 skinfolds (mm) (triceps, subscapular, biceps, iliac crest, supraspinale, abdominal, anterior thigh and medial calf), 11 girths (cm) (neck, arm (relaxed), arm (flexed), forearm, wrist, chest, waist, hip, thigh, calf and ankle), 2 breadths (cm) (biacromial and biiliocristal), 2 depths (cm) (transverse chest and anterior-posterior chest), 2 widths (cm) (humerus and femur), body mass (kg), stretched stature (cm)

and sitting height (cm) using ISAK protocols[26]) were used to establish a correlation matrix on n = 206 participants using SPSS (see Table 1). The correlation matrix provided Pearson's correlation coefficients (r) and P values between the dependent variable (Y = whole body density) and independent variable (Xs). Those variables that had a level of significance (P value) at 0.01 or below were considered potential candidates for the development of the calibration models. The remainder of the matrix was interrogated for collinearity - linear relationships between the independent variables. Regression analyses for whole body density (Y) and each potential predictor (X) was conducted to determine the standard error of estimate (SEE), coefficient of determination (R^2) and R^2 - adjusted values for each variable (see Table 1). A cut-off correlation coefficient was set at 0.950 (90% R^2) and those variables that were above 0.950 were rejected and those that were below were used in the next phase of analyses (see Table 1).

Phase2 statistical analyses, all remaining potential variables gathered from Phase1 were standardised (z-scores, $\bar{x}$ = 0.0, SD = ± 1.0) thereby converting them into one unit of measurement to help reduce heteroscedasticity. Beta weight (β) (or standardised regression coefficient), r, t, significance of t and P-values were calculated via SPSS on the CM group of n = 140 participants (see Table 2).

Phase3 statistical analyses, two groups were constructed: CM (calibration model group; n = 140) and CV (cross-validation group; n = 66). Due to low numbers and potentially contentious issues relating to the estimation of whole body density, non-Caucasians and goalkeepers were positively assigned into the two separate groups. The CM group had n = 13 non-Caucasians whereas the CV group had n = 12 non-Caucasians, whilst both groups had n = 7 goalkeepers. Remaining participants were randomly assigned into each group. The sample size for the CM and CV groups has been regarded as large enough (recommended maximum of 9 variables given the sample number = 15.5 participants per variable) which Atkinson (2005) agree would provide more stability for whom the calibration model was to be developed. Forced regression analysis using an ordinary least squares stepwise approach was conducted from the values obtained from Phase2 on the CM group on those that did not exceed r = 0.80 or a negative beta (β) weight. Pre-selection of the most applicable and worthy anthropometric

variables are exhibited in Figure 1. As part of the model development process, the stepwise analysis procedure involved the elimination of one variable at each stage and was determined by the *t* value and *P* value. At each stage analysis of variance (ANOVA) values such as *F* and *P* values were obtained to determine significance and testing for heteroscedastic (multiplicative) residual errors were calculated including *r* and *P* values. Finally, establishing the most practical and statistically sound calibration models were determined by having the lowest SEE and the highest R^2 values[14].

Phase4 statistical analyses, cross-validation on $n = 66$ of the sample was conducted to test the veracity of the two newly developed calibration models using Bland and Altman[27] 95% Limits of Agreement (LoA) method. Predicted whole body density (g ml^{-1}) was plotted on a Bland and Altman scatter plot to identify agreement between each calibration models and the criterion (see Figure 2). The extent to which heteroscedasticity is present can be illustrated in scatter plots (see Figure 3) which included R^2, *r* and *P*-values and the distribution line to allow a visual overview of the relationship between the calibration model and the criterion values. The final part of the treatment of validity was to identify error and to contextualise and interpret the quantification of agreement where it would be expected to lie for both models ('best fit' and 'practical') for the estimation of whole-body density (g ml^{-1}). Using research from pertinent literature and from the International Society for the Advancement of Kinanthropometry (ISAK), the primary investigator set *a priori* of acceptable tolerable limits at ± 3.8%, $P \leq 0.05$ (g ml^{-1})[26,27]. These limits were set to determine that the agreement had minimal impact (considered a danger to the health and wellbeing of a participant) on the determination of whole-body density in professional football players.

RESULTS

For Phase1 of analyses, the correlation matrix provided outcomes for calculating r, R^2 (%), R^2 - adjusted, SEE and *P*-values for all variables measured in study one ($n = 28$) and is illustrated in Table 1. Results found that of the 28 variables used, 17 variables were statistically significant ($P = < 0.01$) and considered potential candidates for use in the development of the calibration models, whereas 11

variables did not achieve an alpha level of 0.01 and were therefore rejected and subsequently not used for further statistical analyses (see Table 1).

Insert Table 1

Examination of the correlation matrix for collinearity – linear relationships between the independent variables resulted in the cut-off correlation coefficient being set at 0.950 which would give a coefficient of determination (R^2) of 90%, which included 7:8 of the skinfold thicknesses, 5:11 girths, with 2 from the upper limb, 2 from the core body and none from the lower limb, 1:2 breadth, 1:2 depth were accepted, thereby providing a wide range of upper limb, lower limb and core body variables (see Table 1). Of the 17 potential variables, none had a correlation coefficient with any other variable of 0.950 or R^2 of 90% or above. Therefore, all variables were subsequently accepted and used in the next phase of the analyses. Further examination of the correlation matrix led to the grouping of the 17 variables accepted to help improve the prediction. Three groups were considered (skinfold thicknesses, girths and other variables (body mass, stretched stature, sitting height, transverse chest depth and biiliocristal breadth) and re-entered into another correlation matrix. The predictions did improve, for instance the medial calf skinfold rose from an original value of r = -0.203 to r = -0.211. Results from these recalculations found that improvements in grouped predictions were so minimal that it was thought sufficient to continue with values from the original correlation matrix.

For Phase2 of analyses, seventeen variables from Phase1 were standardised into *z*-scores to help reduce heteroscedasticity. β weight, *r*, *t*, significance of *t* and *P*-values on the CM group of n = 140 participants are shown in Table 2.

Insert Table 2

When interrogating the *r*-values for relationships between independent variables and multicollinearity illustrated in Table 2, no measures exceeded the recommended r = 0.80. The highest *r*-values however

were recorded for the anterior thigh skinfold (0.144), hip girth (0148), body mass (0.171), stretched stature (0.188) and sitting height (0.103). Both the supraspinale skinfold and arm (relaxed) girth having the lowest *r*-values of 0.001. However, results summarised in Table 2 indicate that of the 17 potential predictor variables, the hip girth provided the highest β weight (0.210) and thereby the greatest impact on whole body density (g ml^{-1}). Sitting height had the lowest β weight (-0.027) with eight other variables having negative values (triceps skinfold, supraspinale skinfold, medial calf skinfold, arm (flexed) girth, waist girth, biiliocristal breadth and transverse chest depth), indicating that these variables had the smallest impact on whole body density (g ml^{-1}) and therefore not fulfilling the acceptance criteria as explained earlier.

The next phase of analyses (Phase3) was to construct two separate calibration models (a 'best fit' calibration model and a 'practical calibration model) using data from the CM group (n = 140). For the CM group of n = 140 participants ordinary least squares forced regression analysis employing a backward stepwise approach was conducted using the remaining 9 variables to establish the 'best fit' calibration model. The variables included were: subscapular skinfold, iliac crest skinfold, abdominal skinfold, anterior thigh skinfold, neck girth, arm (relaxed) girth, hip girth, body mass and stretched stature. Table 3 summarises the nine 'best fit' calibration models for the estimation of whole-body density developed using measurements from the CM group of n = 140 participants.

Insert Table 3

Examination of the regression analyses summarised in Table 3 revealed 4 potential variables for the most practical 'best fit' calibration model. Further examination found that the most statistically robust calibration model considering the 'best fit' criteria was that which used 6 independent variables: body mass, stretched stature, anterior thigh skinfold, neck girth, arm (relaxed) girth and hip girth. This model had the lowest SEE (0.115 g ml^{-1}) and highest R^2 (6.6%) of the nine 'best fit' potential calibration models. Furthermore, from the ANOVA analysis Fdf-value = 1.56 and P = 0.164, and with

a heteroscedastic coefficient (multiplicative) residual error at this stage of r = -0.213 and P = 0.011 and was statistically significant (P = < 0.005).

For Phase3, the 'practical' calibration model was primarily designed to be used within a football environment of field-testing monitoring and sport science support. However, given the practical nature of this intended design, there was an assumption that professional football club might not be the anthropometric equipment available or the technical support to collect the anthropometric values in these practical environments compared to the 'best fit' model intended for research and academia. 28 of the potential variables are shown in Figure 1, where the primary investigator's judgement on the most applicable and worthy anthropometric variables were selected.

Insert Figure 1

Of the 8 skinfold thicknesses available, at least 1 skinfold should be taken from the lower limb, and 1 from the core body given the physiological demands of the game and recommendations previously reported by Mills *et al*.,[4]. With 11 potential girths available, a judgement that the upper limb was not needed, but at least 1 variable was needed from the lower limb and 1 from the core body. The 2 breadths of the biacromial and biiliocristal were considered as important variable(s), from a practical point of view, as these variables are an important measure of body frame and size[28]. Whereas the 2 depths and 2 widths were considered non-essential in the practical model because they are generally associated with growth and maturation. Finally, at least 1 potential variable was needed from either body mass, stretched stature or sitting height. It could be argued that sitting height and stretched stature is unlikely to have an influence on the estimation of whole-body density (g ml^{-1}) and was rejected. Following the rigorous statistical approach, all judgements were supported by Phase1 and 2 analyses, due to their low impact on the estimation of whole-body density (g ml^{-1}). Five variables (subscapular skinfold, iliac crest skinfold, anterior thigh skinfold, hips girth and body mass) remained for the next phase of analyses in a forced ordinary least squares backward stepwise regression analysis approach to develop the 'practical' calibration model. These five variables consisted of a variation of upper, lower

and trunk locations which according to Hencken[25], provide an excellent subset of measuring total subcutaneous fat levels to estimate whole-body density (g ml^{-1}).

Table 4 illustrates the 'practical' calibration models generated for the estimation of whole-body density (g ml^{-1}) using various combinations of anthropometric measures on the CM group of n = 140 participants. The order in which elimination of variables occurred was as follows: 1) subscapular skinfold, 2) iliac crest skinfold, 3), hips girth and, 4) anterior thigh skinfold.

Insert Table 4

Examination of the calibration model summarised in Table 4 found that the most statistically robust model exhibited the lowest SEE (0.115 g ml^{-1}) and highest R^2 (4.7%) of the 5 potential calibration models. ANOVA components included *Fdf*-value = 1.68 and *P*-value of 0.159. Testing for residual errors heteroscedastic (multiplicative) found r = -0.176 and P = 0.038, with the overall practical calibration model indicating P = < 0.005.

Phase4 of these analyses required the consideration of the cross-validation of the newly developed calibration models on n = 66 of the sample. Table 5 summarises the nine 'best fit' and five 'practical' calibration models to predict whole body density (g ml^{-1}) on the CV group of n = 66 participants.

Insert Table 5

To test the veracity of the newly developed calibration models a cross-validation was conducted on the values from the n = 66 CV sample. This involved using the LoA approach to determine the bias, random variation and heteroscedasticity between whole body density (g ml^{-1}), values measured using the criterion method of hydrostatic weighing against both the for both 'best fit' and 'practical' calibration models (Figure 2).

Insert Figure 2

The LoA plot for the 'best fit' calibration model, shown in Figure 2 identifies a positive bias of +0.005 g ml^{-1} and 95% LoA of –0.026 to +0.036 g ml^{-1}. Whereas the plot for the 'practical' calibration model (Figure 2), identified a positive bias of +0.011 g ml^{-1} and 95% LoA of –0.019 to +0.041 g ml^{-1}. There is some evidence of systematic bias and random variation, with some data clusters around the bias line, and some outliers on both plots. However, the limits of agreement are within acceptable limits, which, indicates that there are minimal issues for sport scientists to consider with respect to the predictions from both calibration models to estimate whole-body density (g ml^{-1}) in professional football players. For illustrative purposes, Figure 3 shows a scatter plot of heteroscedasticity to demonstrate the relationship between the criterion method of whole-body density (g ml^{-1}) for 'best fit' and 'practical' calibration models, where the 'best fit' calibration model provided *r* values = 0.271 and R^2 (%) coefficients = 0.3526, whereas the 'practical' calibration model provided *r* values = 0.596 and R^2 (%) coefficients = 0.3526. There was some deviation from the line of identity, demonstrating some heteroscedasticity between the criterion method of hydrostatic weighing with normal distribution (P = 0.01).

Insert Figure 3

Based on these findings, the question that the primary investigator needs to ask is are the 95% LoA narrow enough for the measurements to be of use for (i) academia, research, and sports science and (ii) field testing monitoring and sports science. Results from the *a priori* criteria established that both calibration models are within the acceptable limits and would be of practical use to the population of professional footballers.

CONCLUSIONS

It is not uncommon for sport scientists to assume responsibility for monitoring and managing their players' body composition over the playing season[22,25,29]. Therefore, the knowledge and understanding of whole body density and how it influences the body could be useful to quantify the effectiveness of a prescribed training programme and/or to help players reach optimal performance potential[25,29]. Previous research from Reilly *et al*.,[22], Mills *et al*.,[4,29] and Gardasevic *et al*.,[30] have indicated the varied under and over estimation of whole body density and therefore, the need for a valid and reliable calibration model(s) specifically for football populations.

The aim of the present study was to develop two separate calibration models on a large sample of n = 140 participants using a forced ordinary least squares backward stepwise regression method approach and then to cross-validate the newly developed models on n = 66 participants to establish the validity of such models when estimating whole body density (g ml^{-1}) in professional footballer players. The purpose of the 'best fit' model could be used within an environment that includes research, academia and/or sports science, where there is an expectation of expertise and understanding within the area of body composition analysis[4,5]. Secondly a 'practical' model which could be used within a football environment including field testing monitoring and/or sports science but can be used for regular monitoring of a player(s) and/or squad(s) and providing informative insight into body composition and possible performance potential. Furthermore, to cross-validate the two calibration models on n = 66 participants to determine the validity and relevance by using Bland and Altman[26] 95% limits of agreement approach.

Following a rigorous statistical approach of 4 distinct phases, results from the regression analysis approach found that the 'best fit'[6] calibration model had the lowest SEE (0.115 g ml^{-1}) and highest R^2 (6.6%) of the 9 calibration models to predict whole body density (g ml^{-1}). Furthermore, ANOVA analysis Fdf-value = 1.56 and P-value = 0.164, and with testing for heteroscedastic (multiplicative) residual errors at this stage revealed r = -0.213 and P = 0.011. This model was statistically significant (P = <0.005). The 'best fit' predictive regression equation developed model was determined as:

Whole body density (g ml-1) = 1.01 + (0.000066 x body mass) + (0.000220 x stretched stature) + (0.000393 x anterior thigh skinfold) + (0.000336 x neck girth) – (0.000587 x arm (relaxed)) + (0.000154 x hip girth)

Furthermore, results from the regression analysis approach found that the 'practical'[4] calibration model had the lowest SEE (0.115 g ml^{-1}) and highest R^2 (4.7%) of the five potential calibration models to predict whole body density. Furthermore, ANOVA analysis *Fdf*-value = 1.68 and *P*-value = 0.159, and with testing for heteroscedastic (multiplicative) residual errors at this stage discovered *r* = -0.176 and *P* = 0.038. This model was statistically significant (*P* = <0.005). The 'practical' predictive regression equation developed model was determined as:

Whole body density (g ml^{-1}) = 1.03 + (0.000160 x body mass) – (0.000072 x iliac crest) + (0.000382 x anterior thigh skinfold) + (0.000173 x hip girth)

Scatter plots of heteroscedasticity provided *r* values = 0.271 and 0.596 and R^2 (%) coefficients = 0.3526 for the 'best fit' and 'practical' calibration models. There was some evidence of heteroscedasticity and deviations from the line of identity between the criterion method of hydrostatic weighing and the calibration models. Both plots provided normal distribution and statistical significance of *P* = 0.01. In summary, reliability findings from Mills *et al.*,[4] had a huge influence on the power of prediction for each calibration model, thereby, providing confidence by which sound judgements on whole body density (g ml^{-1}) could be made. . In essence, given the nature of this study, and the four phased statistical approach, the two calibration models can provide an ecologically and statistically valid contribution to applied sport science knowledge.

PRACTICAL IMPLICATIONS

- A 'best fit' model could be used within an environment that includes research, academia and/or sports science, where there is an expectation of expertise and understanding within the area of body composition analysis
- A 'practical' model which could be used within a football environment including field testing monitoring and/or sports science but can be used for regular monitoring of a

player(s) and/or squad(s) and providing informative insight into body composition and possible performance potential.

- The two calibration models can provide an ecologically and statistically valid contribution to applied sport science knowledge.

ACKNOWLEDGEMENT

N/A

DECLARATION OF INTEREST STATEMENT

The authors report there are no competing interests to declare

INFORMED CONSENT

All participants described in the paper have given written consent to the inclusion of material pertaining to themselves, that they acknowledge that they cannot be identified via the paper; and they have been fully anonymized.

REFERENCES

1. Atkinson, G., (2005). How important is sample size in research? *The Sport and Exercise Scientist*. **5** 22-23
2. Guo, S.S., Siervogel, R.M., and Chumlea, C.W.M., (2000). Epidemiological applications of body composition. The effects and adjustment of measurement errors. *Annals of the New York Academy of Sciences*. **904**(1) 312-316
3. Heyward, V.H., (2000). Nutrition and athletic performance: A position of the American Dietetic Association, Dietitians of Canada, and the American College of Sports Medicine. *Journal of American Dietetic Association.* **100**(12) 1543-1556
4. Mills, C.D., De Ste Croix, M., Cooper, S-M., and James, D.V.B., (2020). Measurement reliability and precision of anthropometric measures for the estimation of whole-body density in professional football players. *Open Access Journal of Biomedical Science*. **3**(1) 658-666
5. Mills, C.D., De Ste Croix, M., Cooper, S-M., and James, D.V.B., (2022). Agreement and validity of existing anthropometric calibration models to estimate whole body density in male professional football players. *Sports and Exercise Medicine – Open Journal*. 8(1) 1-9
6. Wilmore, J.H., and Behnke, A.R., (1969). An anthropometric estimation of body density and lean body weight in young men. *Journal of Applied Physiology*. **27**(1) 25-31
7. Sloan, A.W., and Weir, J.B., (1970). Nomograms for prediction of body density and total body fat from skinfold measurements. *Journal of Applied Physiology*. **28**(2) 221-222
8. Forsyth, H.L., and Sinning, W.E., (1973). The anthropometric estimation of body density and lean body weight of male athletes. *Medicine and Science in Sports and Exercise*. **5**(3) 174-180

9. Katch, F.I., and McArdle, W.D., (1973). Prediction of body density from simple anthropometric measurements in college-age men and women. *Human Biology*. **45**(3) 445-451
10. Behnke, A.R., and Wilmore, J.H., (1974). *Evaluation and Regulation of Body Build and Composition*. International Research Monograph Series in Physical Education. New Jersey: Prentice-Hall Inc.
11. Wickkiser, J., and Kelly, J., (1975). The body composition of a college football team. *Medicine and Science in Sports.* **7**(3) 199-202
12. Jackson, A.S., and Pollock, M.L., (1978). Generalized equations fat predictions of body density of men. *British Journal of Nutrition*. **40**(3) 497-504
13. White, J., Mayhew, J.L., and Piper, F.C., (1980). Prediction of body composition in college football players. *Journal of Sports Medicine and Physical Fitness*. **20**(3) 317-324
14. Lohman, T.G., (1981). Skinfolds and body density and their relation to body fatness: A review. *Human Biology*. **53**(2) 181-221
15. Thorland, W.G., Johnson, G.O., Tharp, G.D., Fagot, T.G., and Hammer, R.W., (1984). Validity of anthropometric equations for the estimation of body density in adolescent athletes. *Medicine and Science in Sports and Exercise*. **16**(1) 77-81
16. Durnin, J.V.G.A., and Rahaman, M.M., (1967). The assessment of the amount of fat in the human body from skinfold thickness. *British Journal of Nutrition*. **21**(3) 681-689
17. Durnin, J.V.G.A., and Womersley, J., (1974). 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. *British Journal of Nutrition*. **32**(1) 17-97
18. Becque, D., Katch, V., and Moffatt, R.J., (1986). Time course of skin-plus-fat compression in males and females. *Human Biology*. **58**(1) 33-42
19. Sinning, W.E., and Wilson, J.R., (1984). Validity of "generalized" equations for body composition analysis in women athletes. *Research Quarterly for Exercise and Sport*. **55**(2) 153-160
20. Ramadan, J., and Byrd, R., (1987). Physical characteristics of elite soccer players. *Journal of Sports medicine and Physical Fitness*. **27**(4) 424-428
21. Withers, R.T., Norton, K.I., Craig, N.P., Hartland, M.C., and Venables, W., (1987). The relative body fat and anthropometric prediction of body density of South Australian females aged 17-35 years. *European Journal of Applied Physiology and Occupational Physiology*. **56**(2) 181-190
22. Reilly, T., Bangsbo, J., and Franks, A., (2000). Anthropometric and physiological predispositions for elite soccer. *Journal of Sports Sciences*. **18**(9) 669-683
23. Atkinson, G., and Nevill, A.M., (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine.* **26**(4) 217-238
24. Casajūs, J.A., and Bosco, S.J., (2001). Seasonal variation in fitness variables in professional soccer players. *Journal of Sports Medicine and Physical Fitness*. **41**(4) 463-469
25. Hencken, C.L., (2004). Anthropometric measurement in elite football players. *Journal of Sports Sciences*. **22(**3) 266-267
26. ISAK (International Society for the Advancement of Kinanthropometry) (2001). *International Standards for Anthropometric Assessment.* Sydney: National Library of Australia.
27. Bland, J.M., and Altman, D.G., (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* **1** 307-310
28. Frisancho, A.R., (1984). New standards of weight and body composition by frame size and height assessment of nutritional status of adults and the elderly. *The American Journal of Clinical Nutrition*. **40**(4) 808-819
29. Mills, C.D., De Ste Croix, M., James, D.V.B., and Cooper, S-M., (2017). The importance of measuring body composition in football players. A Commentary. *Sport and Exercise Medicine – Open Journal.* 3(1) 35-40
30. Gardasevic., J., Bjelica, D., Vasiljevic, I., and Corluka, M., (2019). Differences in Body Composition of Two Top Football Clubs. *Journal of Anthropology of Sport and Physical Education.* **3** 15-19

Table 1 Overview of r, R^2 (%), R^2 - adjusted, SEE and P values for n = 28 variables

Variables	*r*	R^2 (%)	R^2 adjusted	SEE	*P*	Accept / reject
Skinfolds (mm)						
Triceps	-0.249	6.2	5.8	0.014	0.001	Accept
Subscapular	-0.302	9.1	8.7	0.014	0.001	Accept
Biceps	-0.129	1.7	1.2	0.014	0.066	Reject
Iliac crest	-0.378	14.3	13.9	0.013	0.001	Accept
Supraspinale	-0.337	11.3	10.9	0.014	0.001	Accept
Abdominal	-0.354	12.5	12.1	0.013	0.001	Accept
Anterior thigh	-0.271	7.3	6.9	0.014	0.001	Accept
Medial calf	-0.203	4.1	3.7	0.014	0.001	Accept
Girths (cm)						
Neck	-0.269	7.2	6.8	0.014	0.001	Accept
Arm (relaxed)	-0.233	5.4	5.0	0.014	0.001	Accept
Arm (flexed)	-0.191	3.7	3.2	0.014	0.006	Accept
Forearm	-0.079	0.6	0.1	0.014	0.260	Reject
Wrist	-0.022	0.0	0.0	0.014	0.756	Reject
Chest	-0.163	2.7	2.2	0.014	0.019	Reject
Waist	-0.235	5.5	5.1	0.014	0.001	Accept
Hip	-0.283	8.0	7.5	0.014	0.001	Accept
Thigh	-0.138	1.9	1.4	0.014	0.048	Reject
Calf	-0.173	3.0	2.5	0.014	0.013	Reject
Ankle	-0.117	1.4	0.9	0.014	0.094	Reject
Breadths (cm)						
Biacromial	-0.135	1.8	1.3	0.014	0.054	Reject
Biiliocristal	-0.240	5.8	5.3	0.014	0.001	Accept
Depths (cm)						
Transverse chest	-0.201	4.0	3.6	0.014	0.004	Accept
Anterior-posterior chest	-0.177	3.1	2.7	0.014	0.011	Reject
Widths (cm)						
Humerus	-0.100	1.0	0.5	0.014	0.155	Reject
Femur	0.004	0.0	0.0	0.014	0.956	Reject
Other variables						
Body mass (kg)	-0.439	19.2	18.8	0.013	0.001	Accept
Stretched stature (cm)	-0.271	7.3	6.9	0.014	0.001	Accept
Sitting height (cm)	-0.188	3.5	3.1	0.014	0.001	Accept

Table 2 Overview of *r*, beta, *t*, significance of *t* and *P*-values for *n* = 17 variables

Variables	*r*	β	*t*	Sig of *t*	*P*
Skinfolds (mm)					
Triceps	0.019	-0.111	-0.772	0.442	0.410
Subscapular	0.078	0.070	0.471	0.638	0.179
Iliac crest	0.067	0.112	0.730	0.467	0.215
Supraspinale	0.001	-0.107	-0.785	0.434	0.496
Abdominal	0.071	0.015	0.099	0.921	0.204
Anterior thigh	0.144	0.188	1.454	0.148	0.045
Medial calf	0.065	-0.056	-0.438	0.662	0.222
Girths (cm)					
Neck	0.079	0.104	0.769	0.443	0.176
Arm (relaxed)	0.001	0.130	0.694	0.489	0.495
Arm (flexed)	-0.048	-0.188	-1.072	0.286	0.288
Waist	-0.078	-0.201	-1.082	0.072	0.180
Hip	0.148	0.210	1.772	0.079	0.040
Breadths (cm)					
Biiliocristal	-0.043	-0.143	-1.246	0.215	0.305
Depths (cm)					
Transverse chest	-0.091	-0.171	-1.539	0.126	0.142
Other variables					
Body mass (kg)	0.171	0.107	0.667	0.506	0.022
Stretched stature (cm)	0.188	0.198	1.314	0.191	0.013
Sitting height (cm)	0.103	-0.027	-0.206	0.837	0.114

Table 3 'Best fit' calibration models for the estimation of whole-body density using anthropometric measures as predictors in professional football players (n = 140)

Variable included	Calibration model (D_b = g ml^{-1})	SEE	R^2
BM, StS, SS, IC, Ab, AT, Nek, Armr, Hip	D_b = 1.01 + (0.000070 x BM) + (0.000214 x StS) – (0.000054 x SS) – (0.000008 x IC) + (0.000007 x Ab) + (0.000405 x AT) + (0.000358 x Nek) – (0.000599 x Armr) + (0.000159 x Hip)	0.012	6.6
BM, StS, SS, Ab, AT, Nek, Armr, Hip	D_b = 1.01 + (0.000070 x BM) + (0.000214 x StS) – (0.000059 x SS) + (0.000003 x Ab) + (0.000404 x AT) + (0.000365 x Nek) – (0.000601 x Armr) + (0.000158 x Hip)	0.012	6.6
BM, StS, SS, AT, Nek, Armr, Hip	D_b = 1.01 + (0.000071 x BM) + (0.000214 x StS) – (0.000056 x SS) + (0.000405 x AT) + (0.000364 x Nek) – (0.000602 x Armr) + (0.000158 x Hip)	0.012	6.6
BM, StS, AT, Nek, Armr, Hip	D_b = 1.01 + (0.000066 x BM) + (0.000220 x SS) + (0.000393 x AT) + (0.000336 x Nek) – (0.000587 x Armr) + (0.000154 x Hip)	0.012	6.6
StS, AT, Nek, Armr, Hip	D_b = 0.997 + (0.000258 x StS) + (0.000409 x AT) + (0.000429 x Nek) – (0.000551 x Armr) + (0.000180 x Hip)	0.011	6.5
StS, AT, Armr, Hip	D_b = 1.00 + (0.000270 x StS) + (0.000384 x AT) – (0.000409 x Armr) + (0.000224 x Hip)	0.011	6.3
StS, AT, Hip	D_b = 0.997 + (0.000263 x StS) + (0.000375 x AT) + (0.000160 x Hip)	0.011	5.8
StS, AT	D_b = 1.00 + (0.000309 x StS) + (0.000394 x AT)	0.011	5.5
StS	D_b = 1.01 + (0.000314 x StS)	0.012	3.5

KEY:
(skinfolds): SS = subscapular; IC = iliac crest; Ab = abdominal; AT = anterior thigh. (girths): Nec = neck; Armr = arm (relaxed); Hip = hip. (other variables) BM = body mass. StS = stretched stature. D_b = estimate of whole-body density (g ml^{-1}); SEE = standard error of the estimate; R^2 = coefficient of determination (%)

Table 4 'Practical' calibration models for the estimation of whole-body density (g ml^{-1}) from anthropometric measures in professional football players (n = 140)

Variable included	Calibration model (D_b = g ml^{-1})	SEE	R^2
BM, SS, IC, AT, Hip	D_b = 1.03 + (0.000161 x BM) – (0.000037 x SS) – (0.000063 x IC) + (0.000384 x AT) + (0.000175 x Hip)	0.012	4.7
BM, IC, AT, Hip	D_b = 1.03 + (0.000160 x BM) – (0.000072 x IC) + (0.000382 x AT) + (0.000173 x Hip)	0.012	4.7
BM, AT, Hip	D_b = 1.03 + (0.000160 x BM) – (0.000072 x IC) + (0.000382 x AT) + (0.000173 x Hip)	0.012	4.7
BM, AT	D_b = 1.04 + (0.000210 x BM) + (0.000343 x AT)	0.011	4.4
BM	D_b = 1.05 + (0.000234 x BM)	0.012	2.9

KEY:
(skinfolds): SS = subscapular; IC = iliac crest; AT = anterior thigh. (girths): Hip = hip. (other variables) BM = body mass. SS = stretched stature. D_b = estimate of whole-body density (g ml^{-1}); SEE = standard error of the estimate; R^2 = Coefficient of Determination (%)

Table 5 General summary ($\bar{x} \pm s$) of characteristics for the 'best fit' and 'practical' calibration models on the cross validation (CV) group of n = 66 participants to predict whole-body density (g ml^{-1})

Calibration models	$\bar{x} \pm s$	**Range**
'best fit'^{9}	1.068 ± 0.003	1.061 – 1.075
'best fit'^{8}	1.068 ± 0.003	1.061 – 1.075
'best fit'^{7}	1.068 ± 0.003	1.061 – 1.075
'best fit'^{6}	1.069 ± 0.003	1.062 – 1.075
'best fit'^{5}	1.065 ± 0.003	1.056 – 1.072
'best fit'^{4}	1.062 ± 0.003	1.054 – 1.069
'best fit'^{3}	1.064 ± 0.003	1.057 – 1.071
'best fit'^{2}	1.061 ± 0.003	1.053 – 1.067
'best fit'^{1}	1.067 ± 0.002	1.061 – 1.071
'practical'^{5}	1.063 ± 0.003	1.059 – 1.069
'practical'^{4}	1.063 ± 0.003	1.059 – 1.069
'practical'^{3}	1.062 ± 0.003	1.058 – 1.068
'practical'^{2}	1.061 ± 0.003	1.056 – 1.068
'practical'^{1}	1.069 ± 0.002	1.064 – 1.072
Hydrostatic weighing	**1.075 ± 0.015**	**1.034 – 1.132**

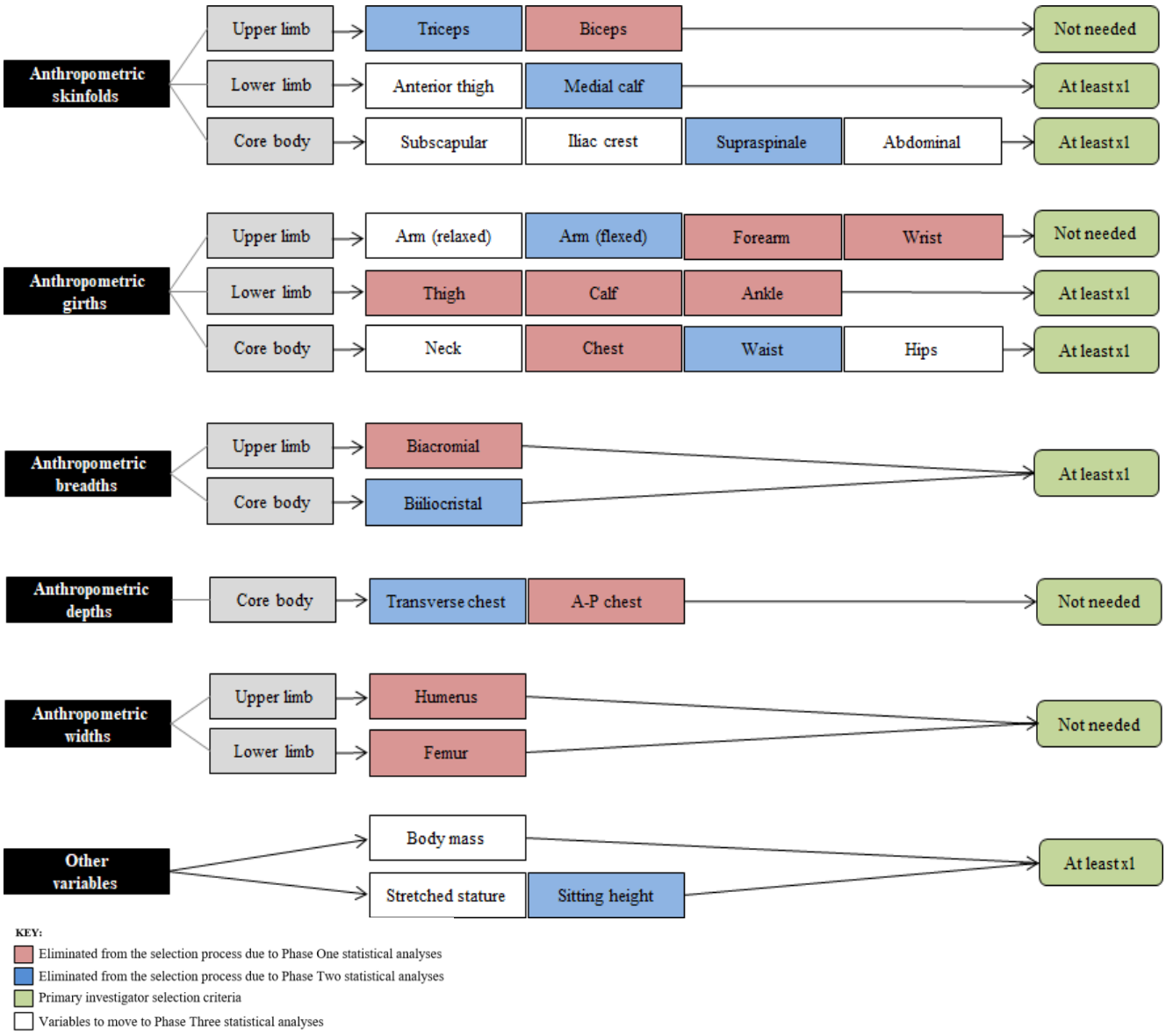

Figure 1 Flow chart to illustrate the variables (n = 9) available for selection for the 'practical' calibration model

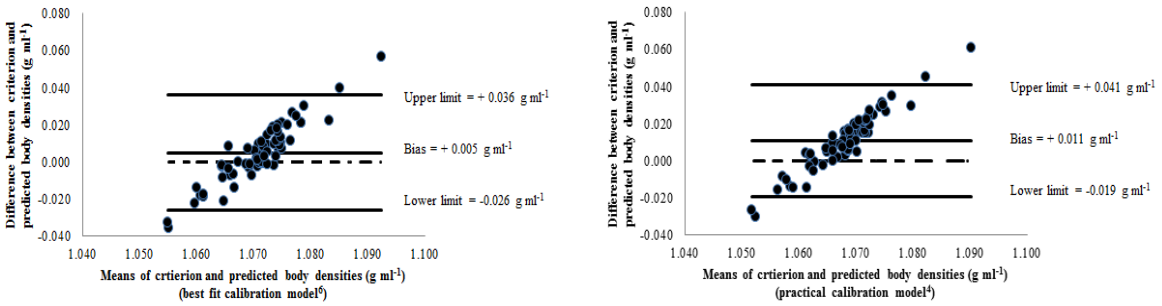

Figure 2 Bland and Altman plot summarising the 95% limits of agreement for comparisons between criterion body densities and those predicted from the 'best fit' and 'practical' calibration models. Note: Direction of bias [hydrostatic weighing – calibration model]

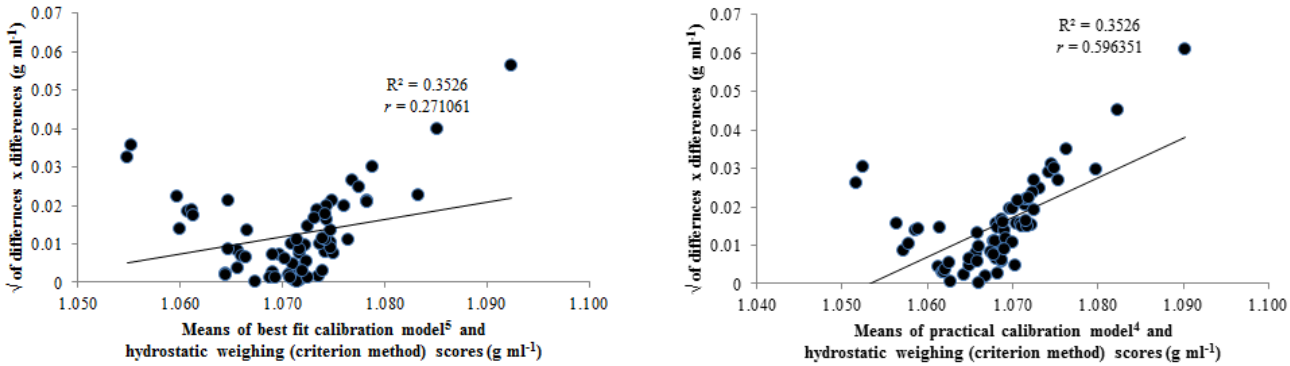

Figure 3 Scatter plots for heteroscedasticity of hydrostatic weighing (criterion method) compared to 'best fit' and 'practical' calibration models (means) for whole body density (g ml^{-1})