Measurement agreement in percent body fat estimates among laboratory and field assessments in college students: Use of equivalence testing

Ryan D. Burns[1]*, **You Fu**[2], **Nora Constantino**[2]

1 Department of Health, Kinesiology, and Recreation, University of Utah, Salt Lake City, Utah, United States of America, **2** School of Community Health Sciences, University of Nevada Reno, Reno, Nevada, United States of America

* ryan.d.burns@utah.edu

Abstract

The purpose of this study was to examine the agreement in percent body fat estimates among 7 laboratory and field assessments against dual-emission x-ray absorptiometry using equivalence testing. Participants were 437 college students (mean age = 19.2±0.6 years). Dual-emission x-ray absorptiometry was used as the criterion with hydrostatic weighing, skinfold thickness, air displacement plethysmography, near infrared reactance, and three methods of bioelectrical impedance analysis examined as surrogate assessments. Relative agreement was examined using intraclass correlation coefficients. Group level agreement was examined using equivalence testing. Individual-level agreement was assessed using Mean Absolute Percent Error and Bland-Altman Plots. Single measure intraclass correlation coefficient scores ranged from 0.71–0.80. Hydrostatic weighing, skinfold thickness, air displacement plethysmography, and 4-electrode bioelectrical impedance analysis showed statistical equivalence with the criterion using a 10% Equivalence Interval with absolute mean differences ranging from 1.0%-4.9% body fat. Mean Absolute Percent Error ranged from 11.7% using skinfold thickness to 21.9% using Omron (hand-held) bioelectrical impedance analysis. Limits of Agreement were heteroscedastic across the range of mean scores compared to dual-emission x-ray absorptiometry, with greater mean differences observed at higher levels of percent body fat. Hydrostatic weighing, skinfold thickness, air displacement plethysmography, and 4-electrode bioelectrical impedance analysis showed strong evidence for statistical equivalence with dual-emission x-ray absorptiometry in a sample of college students.

Introduction

Body composition is a health-related fitness domain that correlates with measures of cardiometabolic health [1,2]. Higher levels of fat mass have been associated with increased incidence

chronic diseases and higher mortality rates [3,4]. Because of the links to health outcomes, health-related fitness testing batteries include body composition as a primary assessment domain [5,6]. Thus, it is an important public health priority to establish the psychometric characteristics of various body composition assessments.

Throughout the past several decades, there has been numerous research studies exploring the reliability and validity of body composition assessments [7,8]. At the population level, indices such as Body Mass Index (BMI) are widely used because of its ease of use and ease of interpretation [9], however at the individual level BMI may not be the most valid assessment because of its inability to distinguish between fat mass and fat-free mass [10]. Therefore, more direct assessments of body composition are needed. Unfortunately, many studies that examine agreement among body composition assessments only examine a limited number of lab and/ or field assessments, making it difficult to determine psychometric characteristics across a range of modalities [11].

Historically, hydrodensitometry (i.e., hydrostatic weighing) has been considered a criterion or reference assessment of body composition and the accuracy of surrogate assessments are often compared to body fat estimates measured using this modality; however recently, multi-compartment models are used as suitable criterion methods [12, 13]. Dual-emission x-ray absorptiometry DXA is often considered a reference measure because of its 3-compartment methodology and high precision and accuracy compared to other 2- and 3-compartment body composition assessments [14, 15]. In this study, DXA was used as the criterion because it uses a 3-compartment model (bone, protein/muscle, and fat) compared to hydrostatic weighing that uses a 2-compartment model (fat-free mass, fat mass).

Many validation studies comparing surrogate assessments to a criterion use a variation of the general linear model (i.e., t-tests, ANOVA) within the data analytic plan to examine group level agreement. However, use of the general linear model approach has limitations, especially as it pertains to decision-making based on obtained p-values [16–18]. For example, a study characterized as having a large sample size may find rejecting the null hypothesis (deciding significant differences exist between group means) relatively easy compared to smaller sample size studies [17]. Rejection of the null hypothesis may lead to erroneous decisions, especially when absolute differences are relatively small [18]. Conversely, studies having small sample sizes may erroneously conclude no differences exist between group means based on obtained p-values, even when absolute differences are relatively large [18]. Furthermore, failure to reject the null hypothesis of "no difference" does not necessarily provide evidence for "equivalence" [16].

Equivalence testing is an alternative method to assess measurement agreement. Equivalence testing tests the null hypothesis of "non-equivalence", or presence of effects large enough to be worthwhile. If differences between two hypothetical group means is sufficiently small, the null is rejected and it is concluded that there is evidence for equivalence between two group means [16]. This "reverse null hypothesis" is practically impossible to reject, as random error will always yield some difference between groups means; therefore, for equivalence testing, researchers define Equivalence Intervals. If an observed mean difference 90% Confidence Interval falls within the designated Equivalence Interval, the null hypothesis is rejected and it is concluded that two measurements are statistically equivalent [16]. For health-related fitness assessments, a 10% Equivalence Interval has been used in the past to examine group-level agreement [19]. As stated in Dixon et al. [16] and Saint-Maurice et al. [19], setting Equivalence Intervals is inherently subjective and there is no universal standard, however a 10% Equivalence Interval is relatively conservative and is an interval that has also been recommended by Robinson et al. [20] for model validation.

Although there are numerous studies examining agreement among body composition assessments, the number of assessments examined within each study are usually limited and

the statistical methodology used to examine group level absolute agreement has employed a variation of the general linear model testing group mean differences, which unfortunately, as stated previously, has limitations. There has been a paucity of work examining agreement in percent body fat estimates across a variety of lab and field-based assessments of body composition using equivalence testing. Therefore, the purpose of this study was to examine the agreement in percent body fat estimates among 7 laboratory and field assessments against DXA via use of equivalence testing.

Materials and methods

Participants

An a priori power analysis was conducted in STATA v15.0 for a paired-sample mean differences test (t-test) with a conservative small effect size, a paired correlation of $r = 0.50$, and a two-sided alpha level of 0.05; in order to achieve at least 80% statistical power, 199 students would need to be recruited. Participants were a convenience sample of college students (Mean age = 19.2 ± 0.6 years; $N = 437$; 301 females, 136 males) recruited from a research university in the western U.S. All participants were enrolled in an exercise science lab course offered during the Fall, Spring, or Summer semesters. All participants were free from physical injury or any psychological condition that would have precluded them from participating in health-related fitness testing. No students were taking diuretics or any medication that could have confounded body composition assessment. There were no exclusion criteria other than the students had to be enrolled in the exercise science laboratory section. All participant data were de-identified. Research reported in the paper was undertaken in compliance with the Helsinki Declaration.

Assessments

Hydrostatic weighing. Hydrodensitometry (i.e., hydrostatic weighing) was a surrogate lab assessment of body composition. The students entered a stainless-steel weighing tank, were instructed to sit on an underwater swing attached to a scale, and subsequently instructed to expel all air from their lungs. Measurements lasted 3–5 seconds. The test was repeated several times (≤ 5 depending on the participant) in order to obtain a stable underwater weight with the body fully submerged. Body volume was calculated as underwater weight divided by water density following a correction for estimated residual lung volume, which was estimated as a sex-specific proportion of spirometry-measured vital capacity (0.24 for males, 0.28 for females) [21]. Body density was calculated by dividing body mass by body volume. Body density was then converted to percent body fat using the Siri equation [22]. Hydrostatic weighing was found to be a reliable measure of body density with reliability coefficient scores > 0.99 [23].

Skinfolds thickness. Seven-site skinfolds thickness assessment was a surrogate field assessment of body composition. Skinfold sites for females and males included the chest, mid-axillary, triceps, subscapular, abdomen, suprailliac, and thigh. Skinfold measurements were collected using a Lange Skinfold Caliper (Lange; Ann Arbor, MI, USA) on the right side of the body. Each site was measured twice in a rotating order. If the two measurements differed by more than 2mm, a third measurement was taken. Body density was estimated using the skinfold sum using validated prediction algorithm [24–26], and test-retest reliability of the method has been established [25]. Percent body fat was then estimated using the Siri equation [22]. Trained graduate students performed the skinfolds thickness measurements.

Air displacement plethysmography. Air Displacement Plethysmography (ADP) was a surrogate lab assessment of body composition. For ADP, body composition was assessed using BOD POD (COSMED; Concord, CA, USA) and the associated standardized procedures to

measure body volume [27]. Calibration of the BOD POD was performed daily. Students were instructed to wear tight fitting clothing and a cap to attenuate potential for measurement error. Once a participant was seated inside the BOD POD chamber, two body volume measurements were taken. Using body weight and volume measurements, body density was calculated and converted to percent body fat using the Siri equation [22]. ADP has been shown to be a reliable assessment of body density with reliability coefficients > 0.99 [28].

Near infrared reactance. Near Infrared Reactance (IR) was a surrogate field assessment of body composition. The students' sex, weight, height and age were entered into the IR device (Futrex-6100 A/ZL, Futrex Inc.; Gaithersburg, MD, USA), which was then was zero-adjusted according to the manufacturer's instructions. Each student sat with their dominant arm relaxed in on an examination table while the light wand of the IR device was placed on the belly of the bicep at the mid-point between the elbow and the acromion process. Readings were determined using infrared light that penetrated 1 cm into the tissue. Scans were made over a range of wavelengths from 700–1,100 nm and the average of 6 optical density readings were used to obtained percent body fat. A light shield was used to block out any surrounding light which could affect the measurement. IR has shown to be a reliable assessment of percent body fat with reliability coefficients > 0.95 [29].

Omron bioelectrical impedance analysis. Two-electrode hand-to-hand Bioelectric Impendence Analysis (BIA) was administered using the Omron handheld device. Omron BIA was a surrogate field assessment of body composition. The students' height, weight, age, and sex were entered into a handheld OMRON Body Fat Analyzer (Model HBF-306; Lake Forest, IL, USA). The students then held the analyzer with arms extended, parallel to the floor until the device displayed the student's body fat percentage. Stability of hand-held Omron BIA devices has been established in college students with test-retest reliability coefficients > 0.97 [30].

Tanita bioelectrical impedance analysis. Tanita BIA was a surrogate field assessment of body composition. Two-electrode foot-to-foot BIA was administered using a Tanita Scale plus BIA (Model BF-556, Tanita; Arlington Heights, IL, USA). The students' height, weight, age, and sex were entered into the scale. The student's stood on the scale with shoes removed until a reading was obtained. Height (in meters), weight (in kilograms), and percent body fat were obtained using the Tanita Scale plus BIA. Tanita scales have shown to have excellent test-retest reliability with coefficients > 0.99 [31].

Valhalla bioelectrical impedance analysis. Four-electrode Valhalla BIA was a surrogate lab assessment of body composition using the RJL Quantum II, which is a four terminal single frequency (800 mA at 50 kHz) impedance plethysmograph (Valhalla Scientific Model 1990B; Clinton Twp., MI, USA). The calibration procedure uses an internal calibration system. Students wore light clothing and were barefoot (or removed the shoe and sock from the right foot). Students reclined in a supine position on an examination table with arms adjacent to the body, palms flat against the table, and legs adjacent to each other but not touching. Four surface self-adhesive spot electrodes were placed on the dorsal surface of the right hand and on the dorsal surface of the right foot. Prior to placement of electrodes the skin was wiped with alcohol at the 4 locations for electrode placement. Resistance and reactance values were determined on the right side of the body. Two trials were performed and recorded for each subject. The mean of these two trials was used in the calculation to estimate percent body fat. Valhalla BIA has produced consistent test-retest reactance values with differences less than 1% [32].

Dual-emission x-ray absorptiometry. DXA was the criterion lab assessment for body composition. DXA (Hologic Discovery W, software version 12.1, Hologic Inc.; Bedford, MA, USA) provides accurate and precise measurements of body bone mineral content and total fat mass with precision scores < 2% [33]. Body composition was divided into bone mass and soft

tissue mass. Soft tissue mass was further divided into fat mass and fat-free mass. Percent body fat was calculated by dividing the fat mass by total body mass. DXA procedures were carried out via a trained and certified administrator within a private screening room.

Estimated VO$_{2\text{ Peak}}$. The sub-maximal Astrand-Ryhming cycle ergometer test was used to estimate VO$_{2\text{Peak}}$ [34]. Participants completed the standard protocol for the sub-maximal Astrand-Ryhming cycle ergometer test. The Astrand-Ryhming sub-maximal cycle ergometer protocol was performed in six-minutes [34]. Heart rate was recorded using Polar Heart Rate Monitors (Polar Electro, Lake Success, NY, USA). The procedures for acquiring estimated VO$_{2\text{ Peak}}$ were aligned with standard procedures using the heart rate extrapolation method. Maximum heart rate was estimated using the equation 220 –age. Before testing, participants rested quietly for 5 minutes so that heart rate could lower to approximate resting levels.

Procedures

Assessments of body composition were administered during lab sections of an exercise science course. The body composition field assessments (skinfolds thickness, two-electrode BIA, IR) were collected first and the lab measures (hydrostatic weighing, DXA, ADP, 4-electrode BIA) were collected second. Testing order was assigned randomly for both lab and field assessments. Students were instructed to follow specific guidelines before reporting to the lab for body composition assessment. These guidelines included: 1.) avoiding large meals at least 2 hours prior to testing; 2.) avoiding vigorous physical activity or exercise at least 12 hours prior to testing; 3.) avoiding alcohol at least 48 hours prior to testing; 4.) consumption of liquids should be limited to 2 glasses of water at least 2 hours prior to testing; 5.) emptying bladder immediately prior to testing. Adherence to these guidelines were verbally confirmed by the student prior to testing. Aerobic capacity assessment took place during a separate lab section.

Statistical analysis

Data were screened for outliers using boxplots and z-scores. Seventeen cases were dropped because of extreme scores that were identified using boxplots that had a z-score $> + 3.0z$ (3.7% of sample). Differences between sexes on the descriptive variables were analyzed using independent t-tests, assuming unequal variances because of the discordance in group sample sizes. Effect sizes were calculated using Cohen's delta (d), where $d < 0.20$ indicating a small effect, $d = 0.50$ indicating a medium effect, and $d > 0.80$ indicating a large effect [35]. Relative agreement in assessment percent body fat estimates with DXA was analyzed using Intraclass Correlation Coefficients (ICCs) via two-way mixed models. ICC scores were computed for each surrogate assessment against DXA. Agreement was considered poor if ICC < 0.50, moderate if ICC $= 0.50$–0.74, good if ICC $= 0.75$–0.90, and excellent if ICC > 0.90 [36].

Group level agreement was examined using equivalence testing. Equivalence testing was employed to test the null hypothesis that there was no equivalence (non-equivalence) between the criterion (DXA) and surrogate assessments of percent body fat. A ±10% Equivalence Interval was employed to test the null hypothesis using the confidence interval method at an alpha level of 0.05 or 5%. As described in Dixon et al. [16], if an alpha = 5% test of equivalence is employed, a 90% Confidence Interval needs to be calculated for the difference in means. Therefore, mean differences between percent body fat measured using DXA and each surrogate assessment of percent body fat were reported along with 90% Confidence Intervals. If a calculated mean difference 90% Confidence Interval fell entirely within the Equivalence Interval (i.e., no values outside the Equivalence Interval), the null hypothesis was rejected and it was deemed that two observed assessments of body composition were statistically equivalent. Secondary analyses tested statistical equivalence against a 5% and a 15% Equivalence Interval.

Table 1. Descriptive statistics (means and standard deviations).

	Female (*n* = 301)	Male (*n* = 136)	Total Sample (*N* = 437)
Age (years)	19.5 (0.6)	18.9 (0.7)	19.2 (0.6)
Body Mass Index (kg/m^2)	24.0 (4.0)	**27.3**[†] **(4.3)**	24.8 (4.2)
Estimated VO$_{2\ Peak}$ (ml/kg/min)	41.8 (11.2)	42.4 (12.1)	42.0 (12.0)
DXA (% Body Fat)	**28.2%**[†] **(6.9%)**	20.8% (7.0%)	27.0% (7.6%)

Note: DXA stands for dual emission x-ray absorptiometry; Bold and † denotes statistical differences between sexes, $p < 0.05$.

Agreement at the individual level was assessed using the mean absolute percent error (MAPE). MAPE was calculated, using DXA as the criterion, for all assessments of body composition. Finally, Bland-Altman Plots were also used to examine individual-level agreement between each surrogate assessment with DXA using STATA's "batplot" command [37]. The 95% limits of Agreement were adjusted for heteroscedasticity to show the variability in mean differences across the range of scores [38]. This methodology can provide more precise information compared to conventional Bland-Altman Plots. The adjusted 95% Limits of Agreement were calculated by regressing the mean differences on to the means. Systematic bias was assessed correlating the residual (differences) against mean scores. All analyses were conducted using STATA v.15.0 statistical software package (College Station, Texas, USA).

Results

Descriptive statistics are reported in Table 1. Males had significantly higher BMI compared to girls (mean difference = 3.3 kg/m^2, $p < 0.001$, $d = 0.78$), however girls had higher percent body fat, measured using DXA (mean difference = 7.4%, $p < 0.001$, $d = 1.01$). Table 2 communicates the single measure and average ICC scores using two-way mixed methodology. Single measure ICC scores ranged from moderate-to-good and average ICC scores ranged from good-to-excellent across assessments. The assessment with the highest single measure ICC score with DXA was ADP and the assessment with the highest average ICC score with DXA was skinfolds thickness.

Table 3 presents the results of the employed equivalence testing using a 10% Equivalence Interval. The Equivalence Interval was set at +/- 10% of the criterion mean, which

Table 2. Intraclass correlation coefficients against dual emission X-ray absorptiometry (N = 437; ICC with 95% confidence Intervals).

Assessment	ICC (3,1)	ICC (3,2)
Hydrostatic Weighing	0.80 (0.76–0.83)	0.89 (0.86–0.91)
Skinfold Thickness	0.73 (0.66–0.79)	0.93 (0.90–0.94)
ADP	0.84 (0.81–0.87)	0.92 (0.90–0.93)
IR	0.79 (0.75–0.82)	0.89 (0.87–0.91)
Omron BIA	0.77 (0.73–0.81)	0.88 (0.86–0.90)
Tanita BIA	0.75 (0.71–0.79)	0.84 (0.81–0.88)
Valhalla BIA	0.77 (0.73–0.81)	0.86 (0.83–0.88)

Note: ICC stands for Intraclass Correlation Coefficient; (3,1) stands for two-way mixed single measure agreement; (3,2) stands for two-way mixed average agreement; ADP stands for air displacement plethysmography; IR stands for near infrared reactance; BIA stands for bioelectrical impedance analysis.

Table 3. Agreement in percent body fat estimates compared to dual-emission X-ray absorptiometry.

Assessment	Mean Difference (DXA–Surrogate; % Body Fat)	90% C.I. Mean Difference (% Body Fat)	10% Equivalence Interval (% Body Fat)	MAPE
Hydrostatic Weighing	**1.0%**[†]	0.6% - 1.4%	-2.7% - 2.7%	13.4%
Skinfold Thickness	**1.4%**[†]	1.1% - 1.7%	-2.7% - 2.7%	11.7%
ADP	**1.6%**[†]	1.3% - 1.9%	-2.7% - 2.7%	14.5%
IR	4.2%	3.7% - 4.7%	-2.7% - 2.7%	18.7%
Omron BIA	4.9%	4.5% - 5.3%	-2.7% - 2.7%	21.9%
Tanita BIA	3.5%	3.0% - 4.0%	-2.7% - 2.7%	17.2%
Valhalla BIA	**1.6%**[†]	1.1% - 2.1%	-2.7% - 2.7%	17.0%

Note: Criterion is percent body fat measured from dual-emission x-ray absorptiometry; 90% C.I. stands for the 90% Confidence Interval; Equivalence is denoted if the 90% C.I. falls completely within the Equivalence Interval

bold and † denotes statistical significance, $p < 0.05$

mean difference statistical significance denotes the null hypothesis rejection of non-equivalence; MAPE stands for mean absolute percent error; ADP stands for air displacement plethysmography; IR stands for near infrared reactance; BIA stands from bioelectrical impedance analysis; DXA stands for dual emission x-ray absorptiometry.

https://doi.org/10.1371/journal.pone.0214029.t003

corresponded to +/- 2.7% body fat. Mean differences with DXA ranged from just 1% body fat using hydrostatic weighing to 4.9% body fat using Omron BIA. Rejection of the null hypothesis of non-equivalence was observed for four assessments: skinfolds thickness, ADP, 4-electrode Valhalla BIA, and hydrostatic weighing. The remaining assessments were determined non-equivalent compared to DXA, as the 90% Confidence Intervals did not fall entirely within the 10% Equivalence Interval. An error bar graph (Fig 1) is communicated to visually communicate the relation of 90% Confidence Intervals compared to various Equivalence Intervals obtained from DXA. Secondary analyses indicated that no surrogate assessment was statistically equivalent with DXA using a 5% Equivalence Interval and all surrogate assessments, except for Omron BIA and IR, were statistically equivalent with DXA using a 15% Equivalence Interval. Agreement with DXA at the individual level was quantified using MAPE (Table 3). MAPE ranged from 11.7% using skinfold thickness to 21.9% using the Omron BIA.

Using Bland-Altman plots, there was evidence for heteroscedasticity when comparing DXA with all surrogate assessments, as there was a correlation between the mean differences and mean scores and the 95% Limits of Agreement were not homogenous across the range of values. Table 4 displays the correlation coefficients between mean differences and mean scores. Fig 2 and Fig 3 visually display the Bland-Altman Plots for each lab and field surrogate assessment against DXA, respectively. The grey shaded area is the 95% Limits of Agreement adjusted for heteroscedasticity. As seen by the varying 95% Limits of Agreement, differences between DXA and each surrogate assessment increased as mean percent body fat increased.

Discussion

The purpose of this study was to examine the agreement in percent body fat estimates among 7 lab and field assessments in a sample of college aged students using equivalence testing. The results support that skinfolds thickness, ADP, 4-electrode Valhalla BIA, and hydrostatic weighing yielded statistically equivalent estimates in body fat using a 10% Equivalence Interval relative to DXA. Previous studies have assessed body composition using a limited number of assessments and the general linear model to assess group level agreement. The results support the use of specific lab and field assessments; specifically, the lab assessments of ADP, 4-electrode BIA (Valhalla), hydrostatic weighing, and the skinfolds thickness field assessment.
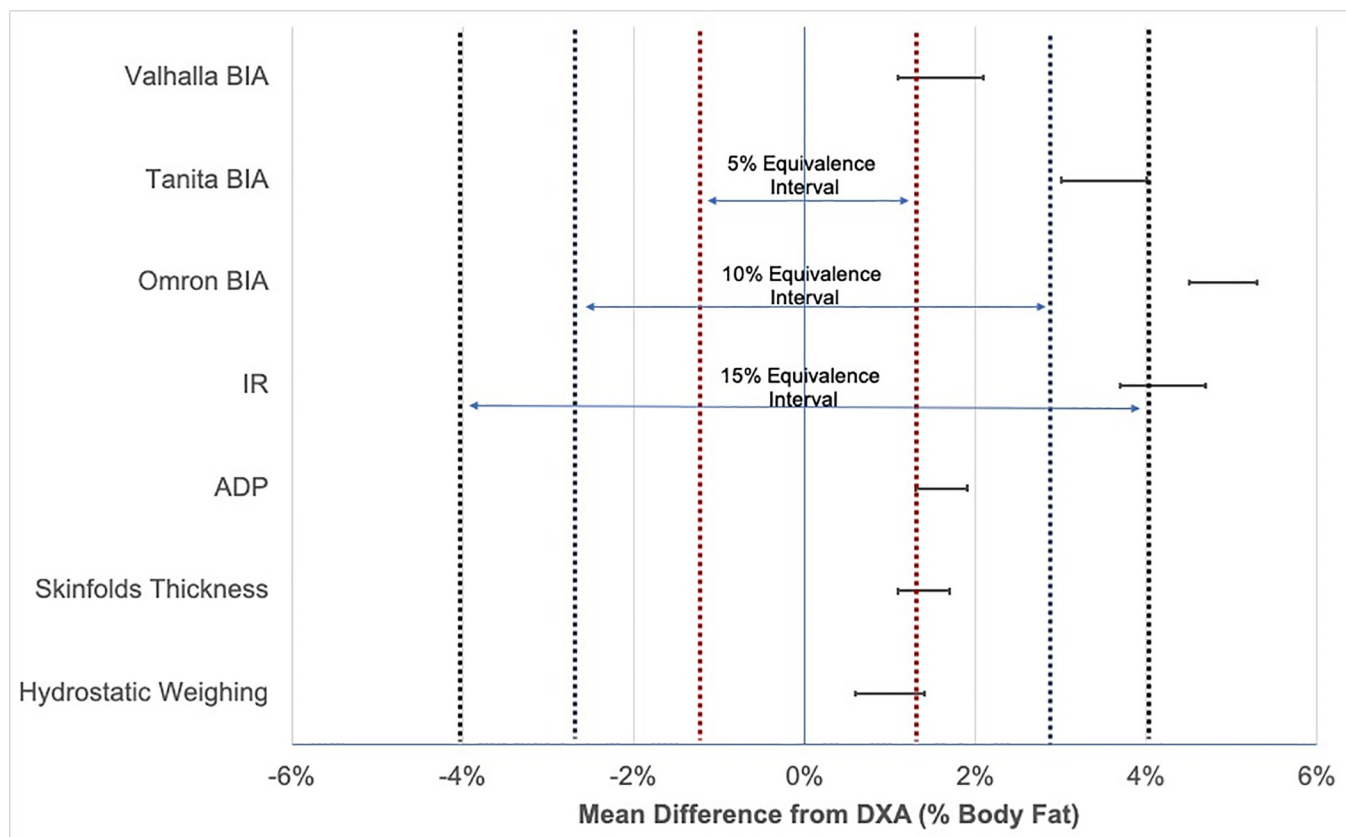
Fig 1. Error bar chart showing the relation of mean difference 90% confidence Intervals to various equivalence intervals. DXA stands for dual emission x-ray absorptiometry; BIA stands from bioelectrical impedance analysis; IR stands for near infrared reactance; ADP stands for air displacement plethysmography; x-axis is mean difference from percent body fat measured using hydrostatic weighing; upper and lower bounds of the Equivalence Intervals are denoted by dashed vertical lines; red dashed lines is the 5% Equivalence Interval; blue dashed lines is the 10% Equivalence Interval; black dashed lines is the 15% Equivalence Interval; Equivalence denoted by a respective 90% Confidence Interval falling within Equivalence Interval.

https://doi.org/10.1371/journal.pone.0214029.g001

MAPE ranged between 10%-20% across most assessments, however there was evidence of heteroscedasticity across the range of mean percent body fat scores, suggesting greater individual error at higher levels of percent body fat.

Table 4. Correlation coefficients between mean differences and mean scores for each surrogate assessment against dual emission X-ray absorptiometry.

Assessment	Pearson *r*	*p*-value
Hydrostatic Weighing	**0.23**[†]	< 0.001
Skinfold Thickness	**0.25**[†]	< 0.001
ADP	**0.25**[†]	< 0.001
IR	**0.33**[†]	< 0.001
Omron BIA	**0.39**[†]	< 0.001
Tanita BIA	**0.28**[†]	< 0.001
Valhalla BIA	0.04	0.784

Note: ADP stands for air displacement plethysmography; IR stands for near infrared reactance; BIA stands for bioelectrical impedance analysis

bold and [†] indicates statistical significance.

https://doi.org/10.1371/journal.pone.0214029.t004
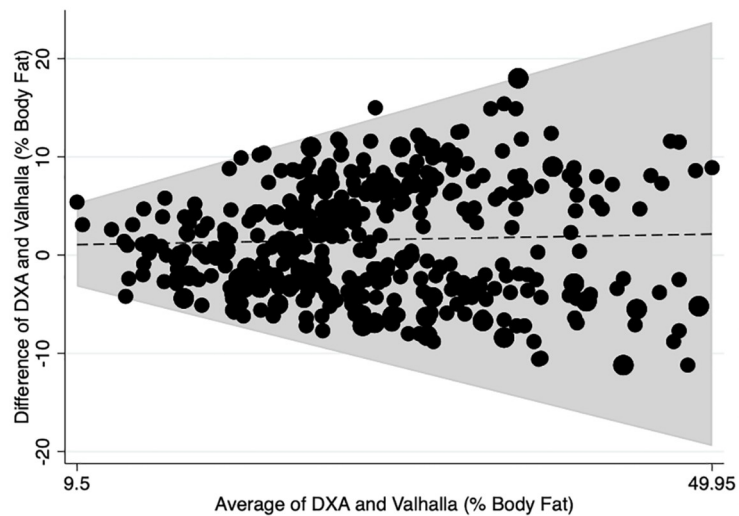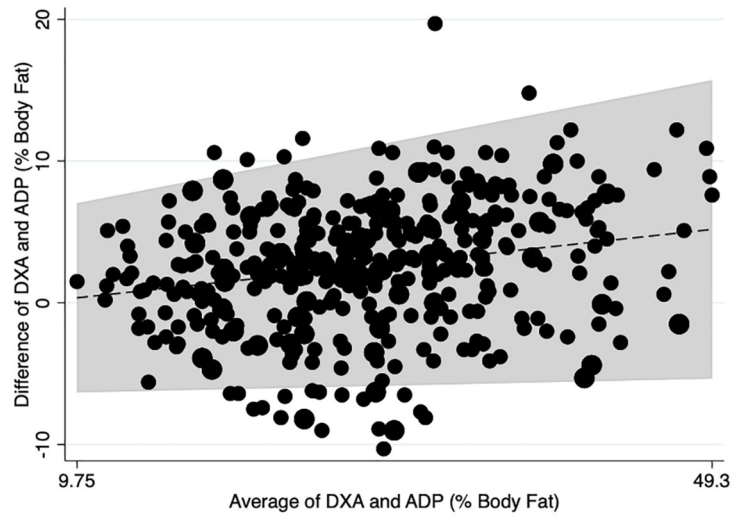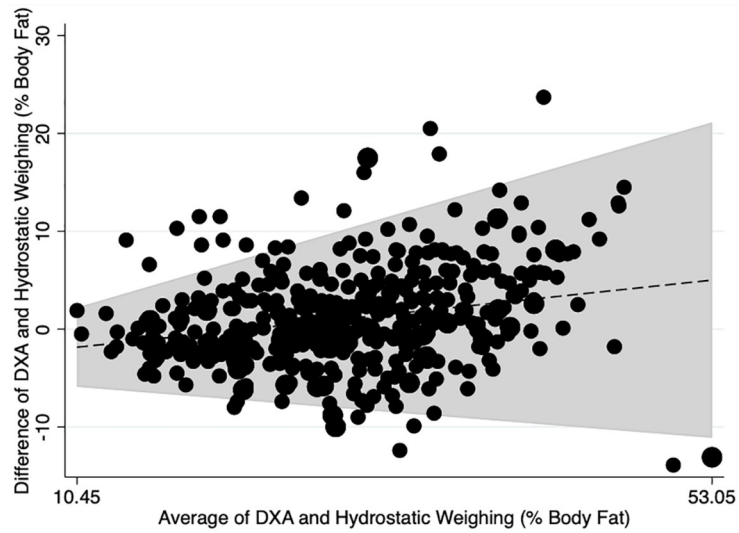
Fig 2. Bland-Altman Plots showing individual-level agreement between each lab surrogate assessment and percent body fat measured using dual emission x-ray absorptiometry. DXA stands for dual emission x-ray absorptiometry; ADP stands for air displacement plethysmography; shaded area is the 95% Limits of Agreement adjusted for heteroscedasticity across the range of mean values.

Valid body composition assessment is of importance to both researchers and practitioners within the fields of nutrition, exercise science, and public health [39]. This is especially important when assessing individuals, where the popular index for assessing body composition, BMI, is characterized as having inherent limitations [40]. Of the surrogate lab assessments observed in the current study, 4-electrode BIA, ADP, and hydrostatic weighing were all determined to be statistically equivalent to percent body fat measured using DXA.

Even though hydrostatic weighing has been considered the criterion in body composition assessment [41], administering the test is cumbersome, requires significant lab space, needs a high degree of administer training, and places a high degree of burden on the participant due to having to be completely submerged underwater [42]. More practical lab assessments throughout recent decades have become alternatives to hydrostatic weighing, such as ADP and DXA. Both of these lab assessments do require some administer training and are monetarily expensive, but subject burden is somewhat attenuated [43]. Hydrostatic weighing's error with DXA was 1%, which is similar in previous studies, however some studies do show greater bias [44]. ADP's absolute error with DXA was at 1.6%, which is again similar compared to other studies using younger adult samples where error ranges from 2–3% body fat [45]. Valhalla 4-electrode BIA absolute error was at 1.6%, which was within the +/- 2.0% body fat observed in other studies compared to a reference method [32, 46].

Of the 4 field assessments examined, percent body fat estimated using skinfolds thickness yielded the lowest percent error and was statistically equivalent with DXA. Skinfolds thickness requires some administer training, but is characterized as having low participant burden and is inexpensive [47, 48]. Health-related fitness test batteries often use skinfolds thickness to assess body composition across a variety of populations [49]. MAPE scores were the smallest using skinfolds thickness with an absolute measurement error of 1.4% compared to DXA. The minimal error observed for skinfolds thickness could be further attenuated using quality assessment training, ensuring that the participant is hydrated and in a fasted state, and ensuring that the assessment precedes any vigorous physical activity, which can cause sub-cutaneous water retention variability that may alter measurement.

Despite the positive findings yielded using skinfolds thickness, other fields assessments including 2-electrode BIA and IR did not show statistical equivalence with DXA. Two-electrode BIA and IR are practical, carrying relative low participant burden and are relatively inexpensive [50]; however, large measurement error is often observed using commercial BIA field assessments, especially in individuals with excess adiposity [51–53]. Two-electrode BIA may be more practical for personal use compared to 4-electrode BIA [52], however tetrapolar BIA may be more accurate. BIA is a method that measures resistance of electrical current through the body [54]. From obtained resistance scores, total body water and fat-free mass can be calculated [55]. Despite BIA's practicality, dehydration can lead to overestimation of percent body fat and overhydration can lead to underestimation of percent body fat [55]. There may also be error related to the specific Tanita and Omron brand algorithms used to estimate body fat. Improvements in BIA accuracy may manifest from further calibration studies against reference assessments [52]. IR also carries with it several sources of measurement error with questionable validity against reference methods [56, 57]. Because of the lack of statistical equivalence and large MAPE, use of 2-electrode BIA and IR to estimate body composition should be interpreted with caution.
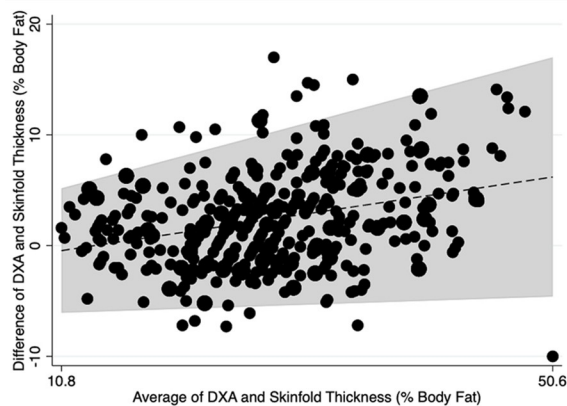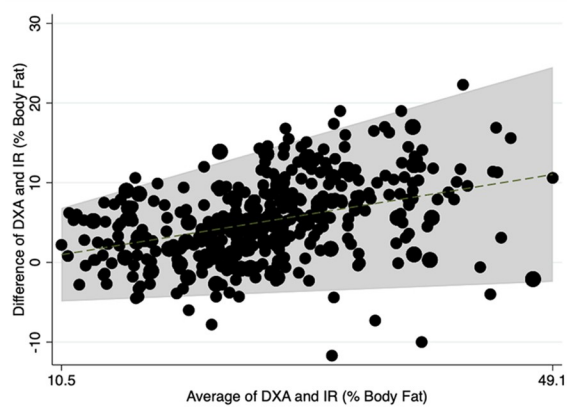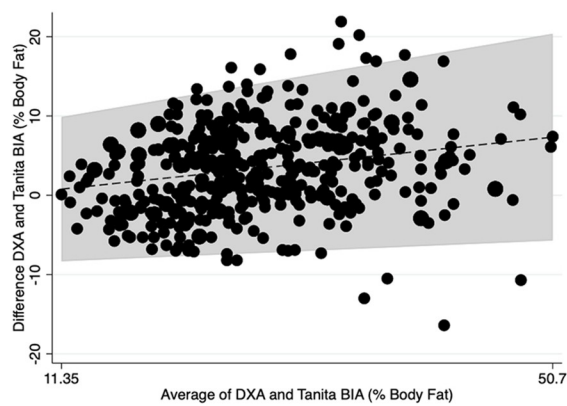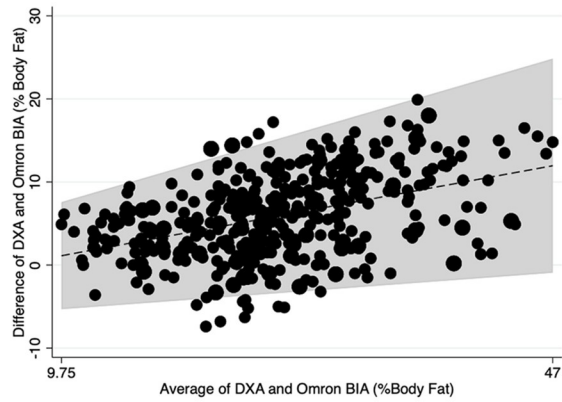
Fig 3. Bland-Altman Plots showing individual-level agreement between each field surrogate assessment and percent body fat measured using dual emission x-ray absorptiometry. DXA stands for dual emission x-ray absorptiometry; BIA stands for bioelectrical impedance analysis; IR stands for near infrared reactance; shaded area is the 95% Limits of Agreement adjusted for heteroscedasticity across the range of mean values.

https://doi.org/10.1371/journal.pone.0214029.g003

Results of this study yield important practical implications. Only one of the 4 field assessments of body composition were determined to be statistically equivalent with DXA. Researchers and practitioners should therefore use caution when employing these devices. Despite ease of use, absolute and relative error using BIA and IR may be practically significant. Because of the non-equivalence and relatively larger MAPE scores, assessment of individual health status based off of these field assessments is precluded. However, it is encouraging the degree of agreement of skinfolds thickness compared to DXA. The use of skinfolds thickness may be preferable over lab-based assessments because of lower subject burden and lower cost of performing the assessment. Therefore, within clinics where resources are limited, the use of skinfolds thickness performed by a trained technician may be very cost-effective.

There are several strengths to this study including the use of a relatively large sample of college students and the use of several lab and field assessments of body composition. The salient strength and novelty of the study was the use of equivalence testing. Additionally, we compared agreement across a range of equivalence intervals. Given these strengths however, there are limitations. First, the sample included college-aged students located from one university located in the western US; therefore, the results do not generalize to younger or older age groups. Second, college students were enrolled in exercise science courses; therefore, most of the participants had a good level of health-related fitness. Third, to maintain a relatively large sample size, results were not stratified by sex. Fourth, hydration status was not completely controlled for, which may decrease the validity of the BIA scores. Fifth, the observed results may have varied by body composition status (e.g., overweight/obese), this should be a focus for future research. Finally, all assessments do not measure percent body fat directly but utilize prediction algorithms. Results may have differed if other prediction equations were used across each of the assessments.

Conclusions

In conclusion, the results support that skinfolds thickness, ADP, 4-electrode Valhalla BIA, and hydrostatic weighing yielded statistical equivalence in estimates of percent body fat relative to DXA. The results support the use the lab assessments of ADP, 4-electrode BIA (Valhalla), hydrostatic weighing, and the skinfolds thickness field assessment. Given the results from the current study, these assessments were the most valid body composition assessments in college-aged participants. Researchers and practitioners should use caution when assessing body composition and subsequent health risk when using 2-electrode BIA and IR field assessments. Valid assessments of body composition are essential for classifying health risk in any population. The current study provides important information on the relative and absolute agreement of various lab and field assessments that can be used in both researchers and clinicians.

Supporting information

S1 Data. De-identified data file.
(CSV)

Acknowledgments

Author Contributions

Conceptualization: Ryan D. Burns, Nora Constantino.

Data curation: You Fu, Nora Constantino.

Formal analysis: Ryan D. Burns.

Investigation: Ryan D. Burns.

Methodology: Ryan D. Burns.

Project administration: Nora Constantino.

Resources: You Fu, Nora Constantino.

Supervision: Nora Constantino.

Writing – original draft: Ryan D. Burns, You Fu, Nora Constantino.

Writing – review & editing: Ryan D. Burns, You Fu, Nora Constantino.

References

1. Bosy-Westphal A, Braun W, Geisler G, Norman K, Muller M. Body composition and cardiometabolic health: The need for novel concepts. Eur J Clin Nutr. 2018; 72:638–644. https://doi.org/10.1038/s41430-018-0158-2 PMID: 29748654

2. Kopelman PG. Obesity as a medical problem. Nature. 2018; 404:635–643. https://doi.org/10.1038/35007508 PMID: 10766250

3. Abdelaal M, le Roux CW, Docherty NG. Morbidity and mortality associated with obesity. Annals of Transl Med. 2017; 5:161. https://doi.org/10.21037/atm.2017.03.107 PMID: 28480197

4. Maffeis C, Tato L. Long-term effects of childhood obesity on morbidity and mortality. Hormone Res. 2001; 55:42–45. https://doi.org/10.1159/000063462 PMID: 11408761

5. Suni JH, Oja P, Miilunpalo SI, Pasanen ME, Vuori IM, Bos K. Health-related fitness test battery for middle-aged adults: Associations with physical activity patterns. Int J Sports Med. 1999; 20:83–91. https://doi.org/10.1055/s-1999-970287

6. Welk GJ, De Saint-Maurice Maduro PF, Laurson KR, Brown DD. Field evaluation of the new FITNESS-GRAM criterion-referenced standards. Am J Prev Med. 2011; 41:S131–S142. https://doi.org/10.1016/j.amepre.2011.07.011 PMID: 21961613

7. Aguirre CA, Salazar GDC, Lopez de Romana DV, Kain JA, Uauy RE. Evaluation of simple body composition methods: Assessment of validity in prepubertal Chilean children. Eur J Clin Nutr. 2014; 69:269–273. https://doi.org/10.1038/ejcn.2014.144 PMID: 25097002

8. Lukaski H. Methods for the assessment of human body composition: Traditional and new. Am J Clin Nutr. 1987; 46:537–556. https://doi.org/10.1093/ajcn/46.4.537 PMID: 3310598

9. Blackburn H, Jacobs D. Origins and evolution of body mass index (BMI): Continuing saga. Int J Epidemiol. 2014; 43:665–669. https://doi.org/10.1093/ije/dyu061 PMID: 24691955

10. Blundell JE, Dulloo AG, Salvador J, Fruhbeck G. Beyond BMI—phenotyping the obesities. Obes Facts. 2014; 7:322–328. https://doi.org/10.1159/000368783 PMID: 25485991

11. Duren DL, Sherwood RJ, Czerwinski SA, Lee M, Choh AC, Siervogel RM, et al. Body composition methods: Comparisons and interpretation. J Diabet Sci Technol. 2008; 2:1139–1146. https://doi.org/10.1177/193229680800200623 PMID: 19885303

12. Van der Ploeg GE, Withers RT, Laforgia J. Percent body fat via DEXA: Comparison with a four-compartment model. J Appl Physiol. 2003; 94:499–506. https://doi.org/10.1152/japplphysiol.00436.2002 PMID: 12531910

13. Schubert MM, Seay RF, Spain KK, Clarke HE, Taylor JK. Reliability and validity of various laboratory methods of body composition assessment in young adults. Clin Physiol Funct Imaging. 2018. https://doi.org/10.1111/cpf.12550 PMID: 30325573

14. Bilsborough JC, Greenway K, Opar D, Livingstone S, Cordy J, Coutts AJ. The accuracy and precision of DXA for assessing body composition in team sport athletes. J Sports Sci. 2014; 32:1821–1828. https://doi.org/10.1080/02640414.2014.926380 PMID: 24914773

15. Norcross J, Van Loan MD. Validation of fan beam dual energy x-ray absorptiometry for body composition assessment in adults aged 18–45 years. Br J Sports Med. 2004; 38:472–476. https://doi.org/10.1136/bjsm.2003.005413 PMID: 15273189

16. Dixon PM, Saint-Maurice PF, Kim Y, Hibbing P, Bai Y, Welk GJ. A primer on the use of equivalence testing for evaluating measurement agreement. Med Sci Sports Exerc. 2018; 50:837–845. https://doi.org/10.1249/MSS.0000000000001481 PMID: 29135817

17. Zhu W. Sadly, the earth is still round (p < 0.05). J Sport Health Sci. 2012; 1:9–11. https://doi.org/10.1016/j.jshs.2012.02.002

18. Zhu W. p < 0.05, < 0.01, < 0.001, < 0.0001, < 0.00001, < 0.000001, or < 0.0000001 . . .J Sport Health Sci. 2016; 5:77–79. https://doi.org/10.1016/j.jshs.2016.01.019 PMID: 30356881

19. Saint-Maurice PF, Welk GJ, Finn KJ, Kaj M. Cross-validation of a PACER prediction equation for assessing aerobic capacity in Hungarian youth. Res Q Exerc Sport. 2015; 86:S66–S73. https://doi.org/10.1080/02701367.2015.1043002 PMID: 26054958

20. Robinson AP, Duursma RA, Marshall JDA. Regression-based equivalence test for model validation: Shifting the burden of proof. Tree Physiol. 2005; 25:903–913. https://doi.org/10.1093/treephys/25.7.903 PMID: 15870057

21. Wilmore J. The use of actual, predicted and constant residual volumes in the assessment of body composition by underwater weighing. Med Sci Sports. 1969; 1:87–90. https://doi.org/10.1249/00005768-196906000-00006

22. Siri W. The gross composition of the body. Advances Bio Med Phys. 1956; 4:239–280. https://doi.org/10.1016/B978-1-4832-3110-5.50011-X

23. Warner JG Jr, Yeater R, Sherwood L, Weber KA. Hydrostatic weighing method using total lung capacity and a small tank. Br J Sports Med. 1986; 20:17–21. https://doi.org/10.1136/bjsm.20.1.17 PMID: 3697596

24. Jackson AS, Pollock ML. Generalized equations for predicting body density of men. Brit J Nutrit. 1978; 40:497–504. https://doi.org/10.1079/BJN19780152 PMID: 718832

25. Jackson AS, Pollock ML, Ward A. Generalized equations for predicting body density of women. Med Sci Sports Exerc. 1980; 12:175–181. https://doi.org/10.1249/00005768-198023000-00009 PMID: 7402053

26. Pescatello LS. ACSM's guidelines for exercise testing and prescription. 9th ed. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health; 2014.

27. McCrory MA, Gomez TD, Bernauer EM, Mole PA. Evaluation of a new air displacement plethysmograph for measuring human body composition. Med Sci Sports Exerc. 1995; 27:1686–1691. https://doi.org/10.1249/00005768-199512000-00016 PMID: 8614326

28. Tucker LA, Lecheminant JD, Bailey BW. Test-retest reliability of the Bod Pod: The effect of multiple assessments. Percept Mot Skills. 2014; 118:563–570. https://doi.org/10.2466/03.PMS.118k15w5 PMID: 24897887

29. Schreiner PJ, Pitkaniemi J, Pekkanen J, Salomaa VV. Reliability of near-infrared interactance body fat assessment relative to standard anthropometric techniques. J Clin Epidemiol. 1995; 48:1361–1367. https://doi.org/10.1016/0895-4356(95)00052-6 PMID: 7490599

30. Hart PD. Test-retest stability of four common body composition assessments in college students. J Phys Fit Med Treat Sports. 2017. https://doi.org/10.19080/JPFMTS.2017.01.555561

31. Kabiri LS, Hernandez DC, Mitchell K. Reliability, validity, and diagnostic value of a pediatric bioelectrical impedance analysis scale. Child Obes. 2015; 11:650–655. https://doi.org/10.1089/chi.2014.0156 PMID: 26332367

32. Nicols J, Going S, Loftin M, Stewart D, Nowicki E, Pickrel J. Comparison of two bioelectrical analysis instruments for determining body composition in adolescent girls. Int J Body Comp Res. 2006; 4:153–160.

33. Jensen MD, Kanaley JA, Roust LR, O'Brian PC, Braun JS, Dunn WL. Assessment of body composition with use of dual-energy x-ray absorptiometry evaluation and comparison with other methods. Mayo Clin Proc. 1993; 68:867–873. https://doi.org/10.1016/S0025-6196(12)60695-8 PMID: 8371605

34. Astrand PO, Ryhming I. A nomogram for calculation of aerobic capacity (physical fitness) from pulse rate during sub-maximal work. J Appl Physiol. 1954; 7:218–221. https://doi.org/10.1152/jappl.1954.7.2.218 PMID: 13211501

35. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale NJ: L. Erlbaum Associates; 1988.

36. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiroprac Med. 2016; 15:155–163. https://doi.org/10.1016/j.jcm.2016.02.012

37. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986; 327:307–310. https://doi.org/10.1016/j.ijnurstu.2009.10.001

38. Mander A. (2012, June 17). BATPLOT: Stata module to produce Bland-Altman plots accounting for trend. Retrieved from https://econpapers.repec.org/RePEc:boc:bocode:s448703.

39. Borga M, West J, Bell JD, Harvey NC, Romu T, Heymsfield SB, et al. Advanced body composition assessment: From body mass index to body composition profiling. J Investig Med. 2018; 66:1–9. https://doi.org/10.1136/jim-2018-000722 PMID: 29581385

40. Gallagher D, Heymsfield SB, Heo M, Jebb SA, Murgatroyd PR, Sakamoto Y. Healthy percentage body fat ranges: An approach for developing guidelines based on body mass index. Am J Clin Nutr. 2000; 72:694–701. https://doi.org/10.1093/ajcn/72.3.694 PMID: 10966886

41. Wagner DR, Heyward VH. Techniques of body composition assessment: A review of laboratory and field methods. Res Q Exerc Sport. 1999; 70:135–149. https://doi.org/10.1080/02701367.1999.10608031 PMID: 10380245

42. Claros G, Hull HR, Fields DA. Comparison of air displacement plethysmography to hydrostatic weighing for estimating total body density in children. BMC Pediatrics. 2005; 5:37. https://doi.org/10.1186/1471-2431-5-37 PMID: 16153297

43. Ackland TR, Lohman TG, Sundgot-Borgen J, Maughan RJ, Meyer NL, Stewart AD, et al. Current status of body composition assessment in sport: Review and position statement on behalf of the ad hoc research working group on body composition health and performance, under the auspices of the I.O.C. Medical Commission. Sports Med. 2012; 42:227–249. https://doi.org/10.2165/11597140-000000000-00000 PMID: 22303996

44. Fogelholm M, van Marken Lichtenbelt W. Comparison of body composition methods: A Literature review. Eur J Clin Nutr. 1997; 51:495–503. https://doi.org/10.1038/sj.ejcn.1600448 PMID: 11248873

45. Fields DA, Goran MI, McCrory MA. Body composition assessment via air displacement plethysmography in adults and children: A Review. Am J Clin Nutr. 2002; 75:453–457. https://doi.org/10.1093/ajcn/75.3.453 PMID: 11864850

46. Kremer MM, Latin RW, Berg KE, Stanek K. Validity of bioelectrical impedance analysis to measure body fat in air force members. Milit Med. 1998; 163:781–785. https://doi.org/10.1093/milmed/163.11.781

47. Barreira TV, Renfrow MS, Tseh W, Kang M. The validity of 7-site skinfold measurements taken by exercise science students. Int J Exerc Sci. 2013; 6:20–28.

48. Bacchi E, Cavedon V, Zancanaro C, Moghetti P, Milanese C. Comparison between dual-energy x-ray absorptiometry and skinfold thickness in assessing body fat in overweight/obese adult patients with type-2 diabetes. Sci Rep. 2017; 7:17424. https://doi.org/10.1038/s41598-017-17788-y PMID: 29234125

49. Castro-Pinero J, Artero EG, Espana-Romero V, Ortega FB, Sjostrom M, Suni J, et al. Criterion-related validity of field-based fitness tests in youth: A systematic review. Brit J Sports Med. 2010; 44:934–943. https://doi.org/10.1136/bjsm.2009.058321 PMID: 19364756

50. Dehgahn M, Merchant AT. Is bioelectrical impedance accurate for use in large epidemiological studies. Nutr J. 2008; 7:26. https://doi.org/10.1186/1475-2891-7-26 PMID: 18778488

51. Gonzalez-Ruiz K, Medrano M, Correa-Bautista JE, Garcia-Hermoso A, Prieto-Benavidas DH, Tordecilla-Sanders A et al. Comparison of bioelectrical impedance analysis slaughter skinfold-thickness equations, and dual-energy x-ray absorptiometry for estimating body fat percentage in Colombian children and adolescents with excess of adiposity. Nutrients. 2018; 10:1086. https://doi.org/10.3390/nu10081086 PMID: 30110944

52. Lee SY, Ahn S, Kim YJ, Ji MJ, Kim KM, Choi SH, et al. Comparison between dual-energy x-ray absorptiometry and bioelectrical impedance analyses for accuracy in measuring whole body muscle mass and appendicular skeletal muscle mass. Nutrients. 2018; 10:738. https://doi.org/10.3390/nu10060738 PMID: 29880741

53. Achamrah N, Colange G, Delay J, Rimbert A, Folope V, Petit V, et al. Comparison of body composition assessment by DXA and BIA according to the body mass index: A retrospective study on 3655 measures. PLOS One. 2018; 13:e0200465. https://doi.org/10.1371/journal.pone.0200465 PMID: 30001381

54. Kyle UG, Bosaeus I, De Lorenzo AD, Deurenberg P, Elia M, Gomez JM, et al. Composition of the ESPEN Working Group. Bioelectrical impedance analysis—part 1: Review of principles and methods. Clin. Nutr. 2004; 23:1226–1243. https://doi.org/10.1016/j.clnu.2004.06.004 PMID: 15380917

55. O'Brian C, Young AJ, Sawka MN. Bioelectrical impendence to estimate changes in hydration status. Int J Sports Med. 2002; 23:361–366. https://doi.org/10.1055/s-2002-33145 PMID: 12165888

56. Klimis-Tavantzis D, Oulare M, Lehnhard H, Cook RA. Near infrared reactance: Validity and use in estimating body composition in adolescents. Nutr Res. 1992; 12:427–439. https://doi.org/10.1016/S0271-53(05)80014-4

57. Wilmore KM, McBride PJ, Wilmore JH. Comparison of bioelectric impedance and near-infrared interactance for body composition assessment in a population of self-perceived overweight adults. Int J Obes Relat Metab Dis. 1994; 18:375–381. https://doi.org/10.1249/00005768-199205001-00036