Development and validation of skinfold-thickness prediction equations with a 4-compartment model1–3

Matthew J Peterson, Stefan A Czerwinski, and Roger M Siervogel

ABSTRACT

Background: Skinfold-thickness measurements are commonly obtained for the indirect assessment of body composition.

Objective: We developed new skinfold-thickness equations by using a 4-compartment model as the reference. Additionally, we compared our new equations with the Durnin and Womersley and Jackson and Pollock skinfold-thickness equations to evaluate each equation’s validity and precision.

Design: Data from 681 healthy, white adults were used. Percentage body fat (%BF) values were calculated by using the 4-compartment model. The cohort was then divided into validation and cross-validation groups. Equations were developed by using regression analyses and the 4-compartment model. All equations were then tested by using the cross-validation group. Tests for accuracy included mean differences, R², and Bland-Altman plots. Precision was evaluated by comparing root mean squared errors.

Results: Our new equations’ estimated means for %BF in men and women (22.7% and 32.6%, respectively) were closest to the corresponding 4-compartment values (22.8% and 32.8%). The Durnin and Womersley equation means in men and women (20.0% and 31.0%, respectively) and the Jackson and Pollock mean in women (26.2%) underestimated %BF. All equations showed a tendency toward underestimation in subjects with higher %BF. Bland-Altman plots showed limited agreement between Durnin and Womersley, Jackson and Pollock, and the 4-compartment model. Precision was similar among all the equations.

Conclusions: We developed accurate and precise skinfold-thickness equations by using a 4-compartment model as the method of reference. Additionally, we found that the skinfold-thickness equations frequently used by clinicians and practitioners underestimate %BF.

KEYWORDS Body composition, skinfold thickness, body compartments, anthropometry, adipose tissue, nutritional assessment, body fat, obesity

INTRODUCTION

Higher amounts of body fat and obesity are associated with increased risks of adverse health events and greater mortality (1, 2). Skinfold thicknesses are commonly measured in clinical and field settings for the assessment of percentage body fat (%BF) because this method is simple to perform and low in cost (3). Two of the most widely used skinfold-thickness equations are those developed by Durnin and Womersley and Jackson and Pollock (4–6).

The Durnin and Womersley equation and the Jackson and Pollock equation were developed and validated by using a 2-compartment (2C) model. The 2C model separates the composition of the body into fat mass and fat-free mass. Using Siri’s equation (7), the 2C model is written as:

\[ %BF = \frac{(4.95/BD) - 4.5}{100} \]  

where BD is whole-body density in g/cc. Use of the 2C-model equation requires the assumptions that body hydration level and bone mineral content are both stable. Unfortunately, these assumptions are often violated because of significant variations in hydration levels and mineral contents between ages, sexes, and races (8–10). The inability of the 2C model to detect these differences can lead to potentially large errors in estimates of %BF in persons for whom these assumptions are violated. Thus, use of the 2C model for development and validation of skinfold-thickness prediction equations is less than ideal (10–14).

Recent advances in in vivo measurement, including 4-compartment (4C) body composition assessment, have allowed researchers to assess variations in hydration levels and bone mineral contents that hydrodensitometry alone cannot measure (13). The 4 compartments of the 4C model are fat, mineral, water, and residual. By adding measurements of total body water (TBW), bone mineral density [via dual-energy X-ray absorptiometry (DXA)], and BD (via hydrodensitometry), it is possible to measure individual variances in mineral and water, thus, in theory, leading to more accurate measurement of %BF. It has been theorized that multiple measurements could lead to the propagation of additive errors when using the 4C model, however, researchers have shown negligible additive errors with this model (11). In addition, the 4C model has been used as the criterion method in studies with children, younger and middle-aged adults, and the elderly (3, 10, 11, 15–18).
We developed a new skinfold-thickness equation (%BF_{new}) to predict %BF by using the 4C model as the reference in younger and middle-aged adults (18–55 y). We evaluated the performance of %BF_{new} compared with %BF estimates derived from 2 commonly used skinfold-thickness equations, namely those of Durnin and Wormersley (%BF_{DW}) and Jackson and Pollock (%BF_{JP}). Because technology that allows us to use a 4C model for validation studies is now accessible, we decided it was important to examine the possibility of providing clinicians and practitioners with improved skinfold-thickness prediction equations.

**SUBJECTS AND METHODS**

**Subjects**

The subjects in the present study are a subset of the participants in the Fels Longitudinal Study. They were enrolled in the Fels Longitudinal Study between 1929 and the present, typically soon after their birth. Most of the Fels participants are white, resided in southwestern Ohio at the time of their enrollment, and were selected for enrollment because their parents were willing to allow them to participate in a long-term, serial study. To date, there have been >1200 participants in the study; they are examined at regular intervals throughout their life span. Of these 1200, a total of 900 are participating in the currently funded National Institutes of Health study. All subjects provided written informed consent.

The Fels Longitudinal Study and its participants have been described in detail previously (19). The standard body-composition exam consists of the anthropometric assessment of weight, stature, and several skinfold thicknesses (chest and abdomen skinfold thicknesses are not currently measured). Many other measures of body composition, including DXA, hydrodensitometry, and bioelectrical impedance, are also obtained. The protocol for this study is reviewed and approved annually by the Wright State University Institutional Review Board. The study sample is representative of the national population in terms of its socioeconomic composition, except for a slight underrepresentation of the lowest socioeconomic group in recent decades. After excluding participants without all of the data needed to derive a %BF reference from the 4C model (see below), there were a total of 681 healthy, white male (n = 360) and female (n = 321) participants available for derivation of the %BF_{new} individual values. This larger sample was then randomly divided into a validation sample (274 men and 230 women) to develop the %BF_{new} equations and a cross-validation sample (86 men and 91 women) used to compare all the equations.

**Development of %BF_{new} values**

DXA scans were performed with a Lunar DPX (Lunar Co, Madison, WI) with total body scan software version 3.6z. The Lunar DPX uses a constant X-ray source and a K-edge filter to achieve a congruent beam of dual-energy radiation. Tissue mass and bone mass are calculated, and tissue mass is further divided into fat-free nonskeletal mass and fat mass. Total body bone mineral content is calculated relative to DXA-determined fat-free mass. Deuterium oxide dilution methods were used to determine TBW according to the procedures described by Schoeller et al (20). Briefly, this method involves measuring the degree of dilution of a known dose of deuterium after it has equilibrated with TBW. Participants emptied their bladders before the collection of a 6-mL baseline saliva sample and then drank 15 mL deuterium oxide (99.9% purity; Cambridge Isotope Laboratories, Woburn, MA) in a solution of 75 mL deionized water. A second saliva sample was collected exactly 2 h later. The 2 samples were centrifuged and the supernate was collected, sealed, frozen, and transported to the Kettering-Scott Magnetic Resonance Laboratory (Kettering, OH) for analysis. Nuclear magnetic resonance spectrometry was used to determine the TBW from each sample. These values were then divided by 1.04 to account for the estimated 4% H\(^{18}\) nonaqueous exchange (21).

The hydrodensitometry method was used to measure BD according to the method described by Siri (7). This method determines BD by using standardized hydrostatic weighing with correction for residual volume. Residual volume was measured twice on land to the nearest 0.1 L by nitrogen washout, using a computerized spirometer (Solatron, Dayton, OH). Underwater weight was determined with the participant sitting in a chair suspended from 4 load cells in a tank of water at 35 °C. When the participant was completely submerged and at maximal exhalation, underwater weights were recorded to the nearest 0.002 kg from a digital display. Ten repeated underwater weights were completed, and the average of the highest 3 weights (because these are indicative of maximal exhalation) was used to calculate BD corrected for residual volume.

The methods described above can be combined to estimate %BF by using the equation for the %BF_{AC} model (11):

\[
{\%BF}_{AC} = \frac{2.559}{BD} - 0.734 (TBW/weight) + 0.983 (TBBM/weight) - 1.841 \times 100
\]

where BD is whole-body density in g/cc measured with densitometry, TBW is total body water measured with \(D_2O\) (in L), TBBM is total-body bone mineral content measured with DXA (in kg), and weight is body weight (in kg).

**Development of %BF_{new} equations**

We identified the set of skinfold-thickness and anthropometric sites that best predicted %BF from the 4C model, and we developed prediction equations using %BF obtained from the 4C model as the reference value. The equation originally included several skinfold-thickness and circumferential sites and various combinations of skinfold-thickness sites. Circumferential measurements were included because these measurements are directly affected by increasing and decreasing regional adiposity and thus they may aid in developing a more precise estimate of %BF (5, 6). In addition, several variables known to be associated with body composition, including age, height, and weight, were included in the equations. Body mass index (BMI, in kg/m\(^2\)) was also included because this is easily measured (using height and weight) and is commonly used as an index of obesity. The anthropometric sites initially included in the new equations were the triceps, subscapular, biceps, midaxillary, suprailliac, midthigh, and lateral calf skinfold thicknesses and the circumferences of the abdomen, hip, thigh, calf, and upper arm. Although it would have been most desirable to also include an abdominal skinfold thickness as a measure of central adiposity in the initial equations, the Fels Longitudinal Study data collection protocol does not currently include such a measure. However, we thought that the inclusion of a total of 7 skinfold-thickness sites and 5 circumferences in the original equation would result in a strong prediction equation. Two highly skilled technicians measured each anthropometric site using the techniques described in the Anthropometric Standardization Reference Manual (22) with Holtain skinfold calipers (Holtain Ltd,
Crosswell, Crymych, United Kingdom) and a metal or cloth (for upper arm only) tape measure.

Statistical analyses

The physical characteristics of the validation and cross-validation groups are presented as means with SDs and ranges. Unpaired t tests were used to determine significant differences between mean values within each sex between groups (validation and cross-validation groups) and between the sexes within groups. Sex-specific stepwise regression analyses were used to develop the %BFnew equations in the validation group. Variables included in the initial analyses included age, height (cm); weight (kg); BMI; the sum of the triceps, subscapular, suprailiac, and midthigh skinfold thicknesses (sum4); and the sum of the triceps, subscapular, suprailiac, and midthigh skinfold thicknesses (sum4). Squares of sum3 (sum32) and sum4 (sum42) were also included to test for a curvilinear association between skinfold thicknesses and %BF.

In the cross-validation group, the accuracy of the %BFnew, %BFJP, and %BFnew equations was tested by calculating correlation coefficients. In addition, Bland-Altman plots (23) were used to determine bias and limits of agreement between equations. Bias was defined as (predicted %BF – %BFtrue) plotted on the y axis and the mean of %BFtrue and predicted %BF plotted on the x axis. Limits of agreement were defined as upper and lower 95% CIs and were determined by the mean differences ± 2 SD. To determine each equation’s precision, root mean square error (RMSE) values were obtained by using univariate analysis. In all the analyses, %BFtrue was used as the method of reference. As described previously, chest skinfold data were not available for men, so analyses with Jackson and Pollock equations are for women only. Alpha was set at $P \leq 0.05$. We used a Bonferroni adjustment for multiple comparisons of %BF methods and set $\alpha$ at $P < 0.025$ for men and $P < 0.0167$ for women. All analyses were performed with SAS version 8e (SAS Institute Inc, Cary, NC).

RESULTS

%BFnew equations

The physical characteristics of subjects in the validation and cross-validation groups are shown in Table 1. In both groups, men and women differed significantly in height, weight, and BMI, but age was not significantly different between the sexes. Men in the validation group were heavier than were men in the cross-validation group. Physical characteristics did not differ significantly between women in the validation and cross-validation groups.

Regression analyses in the validation group (Table 2) revealed that the sum of the triceps, subscapular, suprailiac, and midthigh skinfold thicknesses (sum4) explained ~56% (partial $R^2 = 0.56$) of the variance of %BF in men and ~65% (partial $R^2 = 0.64$) of the variance of %BF in women. In men, age, height, and sum42 were also significant contributors to the model, but to a lesser degree, with no remaining variable’s partial $R^2$ being $> 0.027$. In women, in addition to the same set of variables that were significant in the model for men, BMI was also a significant predictor of %BF, although its effect size was relatively small (partial $R^2 = 0.024$). Interestingly, sum3, sum32, circumferential measures, and weight were not significant predictors of %BF in either model. The final equations, which were used in subsequent analyses, are as follows for men and women:

For men: $%BF_{\text{new}} = 20.94878 + (\text{age} \times 0.1166) - (\text{height} \times 0.1166) + (\text{sum4} \times 0.42696) - (\text{sum4}^2 \times 0.00159)$ (3)

For women: $%BF_{\text{new}} = 22.18945 + (\text{age} \times 0.06368) + (\text{BMI} \times 0.60404) - (\text{height} \times 0.14520) + (\text{sum4} \times 0.30919) + (\text{sum3}\times 0.0009562)$ (4)

where height is in cm and sum4 is the sum of the triceps, subscapular, suprailiac, and midthigh skinfold thicknesses.
TABLE 3
Percentage body fat (%BF) estimates and measurements in the cross-validation group

<table>
<thead>
<tr>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Men (n = 86)</strong></td>
</tr>
<tr>
<td>%BF&lt;sub&gt;new&lt;/sub&gt;</td>
</tr>
<tr>
<td>22.7 ± 7.1 (11.1–36.6)</td>
</tr>
<tr>
<td>%BF&lt;sub&gt;4C&lt;/sub&gt;</td>
</tr>
<tr>
<td>22.8 ± 6.5 (4.6–36.4)</td>
</tr>
<tr>
<td>%BF&lt;sub&gt;JP&lt;/sub&gt;</td>
</tr>
<tr>
<td>20.0 ± 7.0 (7.6–34.8)</td>
</tr>
<tr>
<td><strong>Women (n = 91)</strong></td>
</tr>
<tr>
<td>%BF&lt;sub&gt;new&lt;/sub&gt;</td>
</tr>
<tr>
<td>32.6 ± 6.2 (21.7–34.2)</td>
</tr>
<tr>
<td>%BF&lt;sub&gt;4C&lt;/sub&gt;</td>
</tr>
<tr>
<td>32.8 ± 6.2 (14.1–46.2)</td>
</tr>
<tr>
<td>%BF&lt;sub&gt;JP&lt;/sub&gt;</td>
</tr>
<tr>
<td>26.2 ± 5.0 (13.6–39.8)</td>
</tr>
<tr>
<td>%BF&lt;sub&gt;DW&lt;/sub&gt;</td>
</tr>
<tr>
<td>31.0 ± 5.5 (20.4–34.9)</td>
</tr>
</tbody>
</table>

<sup>1</sup> ± SD; range in parentheses. 4C, 4-compartment; JP, Jackson and Pollock; DW, Durnin and Wormersley.

<sup>2</sup> Significantly different from %BF<sub>4C</sub> within the same sex, P < 0.05 (Bonferroni-adjusted P; unpaired t test).

Comparison of all equations

The results of the predicted and measured %BF means, SDs, and ranges for each equation or method are shown in Table 3 for the cross-validation group. As the reported mean values show, the %BF<sub>new</sub> values are closest to the %BF<sub>4C</sub> values for both men and women, with slight overestimations of %BF (0.1% and 0.2% in men and women, respectively). Jackson and Pollock underestimated %BF in women when compared with the 4C model, with a mean underestimation of 6.6%. In men and women, Durnin and Wormersley also underestimated %BF when compared with the 4C model, with mean underestimations of 3.1% and 2.4%, respectively. When considering differences in means between the predicted and %BF<sub>4C</sub> values, the %BF<sub>new</sub> equations were the most accurate in the cross-validation group.

Bland-Altman plots were used to further assess the accuracy of each prediction equation. In Figures 1–5, we have shown the bias and limits of agreement in each prediction equation by plotting the difference between the measured and predicted values (y axis) against the mean of the measured and predicted values (x axis).

Differences in %BF<sub>4C</sub> and %BF<sub>new</sub> in men and women were plotted against %BF<sub>4C</sub> in Figures 1 and 2. The limits of agreement (mean difference ± 2 SD) were approximately ± 9.5 %BF, with the exception of the lower confidence limits (−10.1%) for women. This is interpreted as 95% of individuals’ predicted %BF value could be up to ±10% above or below their actual value. Mean differences were close to 0, at −0.18% and −0.25% (P > 0.05) in men and women, respectively. In Figures 3 and 4, relations between %BF<sub>4C</sub> and %BF<sub>DW</sub> differences and their means in men and women are shown. Limits of agreement and bias tend to favor an underestimation by the Durnin and Wormersley equation, with upper limits at ±7.2% and 8.0% in men and women, respectively and lower confidence limits at −12.9% and −11.4% in men and women, respectively. There is also an underestimation bias in men and women when using the Durnin and Wormersley equation, with mean differences at −2.8% and −1.8% (P < 0.05) in men and women, respectively. Differences between %BF<sub>JP</sub> and %BF<sub>4C</sub> and

FIGURE 1. Bland-Altman plot showing the limits of agreement between percentage body fat calculated with the 4-compartment-model equation (%BF<sub>4C</sub>; the reference equation) and percentage body fat calculated with the new equation developed in the current study (%BF<sub>new</sub>) in men.

FIGURE 2. Bland-Altman plot showing the limits of agreement between percentage body fat calculated with the 4-compartment-model equation (%BF<sub>4C</sub>; the reference equation) and percentage body fat calculated with the new equation developed in the current study (%BF<sub>new</sub>) in women.

FIGURE 3. Bland-Altman plot showing the limits of agreement between percentage body fat calculated with the 4-compartment-model equation (%BF<sub>4C</sub>; the reference equation) and percentage body fat calculated with the Durnin and Wormersley equation (%BF<sub>DW</sub>) in men.
their means in women are shown in Figure 5. Bias with the Jackson and Pollock equation is seen, with a mean difference of −6.6% (P < 0.05). Limits of agreement follow the same tendency of underestimation, with upper and lower confidence limits at 3.3% and −16.5%, respectively. Relations are evident between the mean of the measured and predicted values and the individual differences in %BF in both sexes with all equations. In other words, the relations seen in all equations are a function of %BF, with all equations having a tendency to produce larger underestimations as %BF increased.

To address issues of precision, RMSEs of predicted %BF in the cross-validation groups were investigated. %BFnew had a RMSE for men of 4.6% when predicting %BF, and the corresponding value for %BF Dw was 4.9%. In women, %BFnew equations had a RMSE of 5.0%. The corresponding values for %BF Dw and %BFJP in women were both 4.9%. Overall, RMSEs were quite similar among all the prediction equations.

DISCUSSION

We developed new skinfold-thickness prediction equations and compared them with the equations of Durnin and Wormersley and Jackson and Pollock, and with a 4C model, which we used as the reference equation. To our knowledge, no previous study has used a 4C model to cross-validate several skinfold-thickness prediction equations. We chose the Durnin and Wormersley and Jackson and Pollock equations because of their popularity in the field and in research settings (18, 24–31). Previous studies have explored the predictive accuracy and precision of the Jackson and Pollock and Durnin and Wormersley equations in different populations (18, 24–31). Our study is also unique in that we present data from healthy, white men and women who had characteristics comparable to those of the groups used by Durnin and Wormersley and Jackson and Pollock.

The skinfold-thickness equations developed in our laboratory had no significant mean differences with the %BFac model in men or women in the cross-validation group. In this cohort, %BF Dw and %BFJP means were significantly lower than the %BFac mean, which may need to be considered when using the Durnin and Wormersley and Jackson and Pollock equations to predict %BF in groups similar to the one presented here. Bland-Altman plots provided additional information regarding the underestimation of %BF by the Durnin and Wormersley and Jackson and Pollock equations. With the Durnin and Wormersley equations, the lower limits of agreement (mean differences − 2 SD) were −12.9% and −11.4% in men and women, respectively. These lower limits of agreement are considerably larger than the upper limit values, at 7.2% and 7.9% in men and women, respectively. On the basis of these data, a man or woman is more likely to have their %BF underestimated using the Durnin and Wormersley equations. This tendency to have the limits of agreement shifted toward underestimation is even more pronounced in women using the Jackson and Pollock equation, with upper and lower limits of agreement of 3.3% and −16.5%, respectively. This indicates that a woman could have her %BF underestimated by as much as 16.5%. These underestimations are most likely a result of systematic overestimations of BD. For example, the mean BD value with the Durnin and Wormersley equation was higher, at 1.032 g/mL, than the value of 1.028 g/mL that was measured by hydrodensitometry. Although 0.004 g/mL may initially appear to be a small difference in BD estimates, when using these 2 values to predict %BF, the differences in mean predicted %BF values are almost 2%. A relatively small under- or overestimation of BD can result in considerable errors in predicted %BF values.

%BFnew, %BFJP, and %BF Dw all had similar estimates of precision, as shown by the similar RMSE values. A prediction equation’s RMSE is certainly important, however, it should not be considered the only criterion when choosing a proper equation. Although %BF Dw and %BFJP values were similar in terms of RMSE values and precision, we showed that accuracy was less than desirable in the Durnin and Wormersley and Jackson and Pollock equations in a cohort similar in age and physical characteristics to their derivation samples.

Providing precise and accurate body-composition information is of great importance. The comparisons provided in this paper should aid in making a more informed decision, considering precision and accuracy, when choosing among popular equations for predicting %BF. In summary, we have shown that our newly developed generalized equations, which use the 4C model as a reference, can accurately predict %BF in healthy, white adults. Our
newly developed skinfold-thickness equations are important for 2 reasons. First, the \%BF\textsubscript{new} equations were validated and cross-validated with large samples of men and women spanning a wide age range (18–55 y). The validation sample used by Durnin and Wormersley was similar to ours in size, but they did not cross-validate their equations. Jackson and Pollock did cross-validate, but their cross-validation did not provide an analysis of their equations’ potential to over- or underestimate \%BF. Second, the \%BF\textsubscript{new} equations were developed and validated by using a 4C model as a reference. The 4C model has been shown to be more accurate in measuring \%BF than the 2C method, which was the criterion method for Durnin and Wormersley and Jackson and Pollock. When validated against the 4C model, the Durnin and Wormersley equations underestimated \%BF in men and women and the Jackson and Pollock generalized equations underestimated \%BF in women to a relatively high degree. We recommend considering an equation’s accuracy and precision to fully understand its strengths and weaknesses. This may require finding the original source for the derivation of the equation and objectively analyzing the information presented to find the best equation for a specific population.

We thank the participants of the Fels Longitudinal Study and the data collection staff, without whom this work would not have been possible. We also thank Miriam Morey for her thoughtful review of drafts and for her support of MJP throughout the writing and revision of this work. MJP, SAC, and RMS were all involved in the study design and writing of the manuscript. Data analyses were performed by MJP. The authors had no financial or personal interests in the sponsoring organization.

REFERENCES