Reliability and Validity of Commercially Available Low-Cost Bioelectrical Impedance Analysis

Kerri L. Vasold, Andrew C. Parks, Deanna M.L. Phelan, Matthew B. Pontifex, and James M. Pivarnik

Michigan State University

Research comparing portable body composition methods, such as bioelectrical impedance analysis (BIA), to air displacement plethysmography (ADP) is limited. We assessed reliability and validity of predicting fat-free mass (FFM) by the RJL, Omron, and Tanita BIA machines using ADP via BodPod as a criterion. FFM (kg) was assessed twice in college students (N=77, 31 males and 46 females; age = 19.1 ± 1.2 years) using ADP, RJL, Omron, and Tanita BIA. Reliability was assessed using analysis of variance to obtain an intraclass correlation statistic (R_{ic}). Validity was assessed using Pearson correlation (r) coefficient. FFM averaged 75.6 ± 9.4 kg in men and 59.8 ± 7.6 kg in women. Reliability was high in both genders RJL (R_{ic} = .974–.994), Omron (R_{ic} = .933–.993), and Tanita (R_{ic} = .921–.991). Validity within males was also high: RJL (r = .935), Omron (r = .942), and Tanita (r = .934), and only slightly lower in females: RJL (r = .924), Omron (r = .897), and Tanita (r = .898). The RJL, Omron, and Tanita BIA machines appear to be both reliable and valid for predicting FFM of male and female college students. Therefore, any of these three BIA devices is appropriate to use for body composition assessment in a healthy adult population.

Keywords: body composition measurement, concurrent validity, test–retest reliability

Body composition is a key component of health-related fitness and often used as a tool for risk assessment or as a measure of change in physical activity or diet. A variety of situations require the use of accurate body composition measurement techniques. There are various methods used to assess body composition using the two-compartment model (fat mass [FM] and fat-free mass [FFM]), and the limitations and feasibility of use vary with each method. Two well-accepted measurement methods used today include air displacement plethysmography (ADP) and bioelectrical impedance analysis (BIA).

Air displacement plethysmography has shown high reliability and validity for evaluating body composition in many populations, and it has been used frequently as a criterion measure for field techniques in the past two decades, as other methods such as hydrodensitometry and dual-energy X-ray absorptiometry are not often available (Fields et al., 2000; McCrory et al., 1995; Tucker et al., 2014). Tucker and colleagues (2014) found high test–retest reliability with intraclass correlation coefficients from .991 to .998 in a sample of women (N=283). Concerning validity, studies by Fields et al. (2000) and McCrory et al. (1995) found high validity between ADP and hydrostatic weighing with r^2 of .94 and .93, respectively. Both ADP and BIA have their own strengths and limitations, but a primary concern when choosing a method for body composition analysis is the feasibility of testing large groups. Whereas ADP is not easily transported and is significantly more expensive than BIA devices, popular BIA devices, such as the RJL, Omron, and Tanita, are small, easy to transport, and relatively inexpensive when testing large groups. With the ability to administer assessments outside of the laboratory setting more easily, the use of BIA in research and clinical settings has increased (Chumlea & Guo, 1994; Lukaski et al., 1985). Therefore, it is important to ensure that these commonly used devices are reliable and valid when compared to laboratory criterion methods. Previous research on the validity of BIA has shown that while correlations are relatively strong (r = .81–.98), investigators have found a wide range of error (1.8–18.3%) (Fornetti et al., 1999; Gibson et al., 2000; Kelly & Metcalfe, 2012; Khaled et al., 1988; Pribyl et al., 2011). Thus, for a given individual, all instruments are not equal when it comes to predictive accuracy. In addition, few studies have estimated the validity of more than one type of BIA machine simultaneously with a criterion measure or evaluated validity and reliability within the same study. Because there are numerous modalities available for body composition assessment, there is a need for analysis to determine whether devices are equal in measurement capabilities or if specific tools should be utilized over others.

Therefore, our purpose was to assess the reliability and validity of FFM estimation using the RJL, Omron, and Tanita BIA devices using ADP as a criterion measure. We hypothesized that all BIA devices would demonstrate high reliability and validity, but the most reliable and valid device would be the RJL BIA due to the measurement of resistance and reactance, rather than relying on a proprietary equation for the Omron and Tanita BIA devices.

Methods

Participants

Analyses were conducted on a convenience sample of 77 (46 females) college-aged adults (mean age: 19.1 ± 1.2 years). All participants provided written informed consent in accordance with the institutional review board at Michigan State University and reported being free of any physical disabilities. Table 1 summarizes the demographic data for all participants.
Body Composition Assessments

Body composition was assessed using a criterion measure of ADP and three methods of BIA. Prior to completing these assessments, anthropometric measurements of height and weight were collected for each participant.

Air displacement plethysmography. The BodPod GS (COSMED, Chicago, IL) was used for ADP and calibrated based on the manufacturer’s instructions. Participants wore Lycra caps, minimal clothing, and removed jewelry prior to measurement in accordance with the standard protocol for minimizing air displacement during measurement. Direct measurement of thoracic lung volume (L) was also measured for each participant according to the manufacturer’s breathing guidelines and procedures (COSMED, 2004). Fat-free mass (FFM) was estimated for all participants using the Siri Equation for percent body fat estimation, based on the two-compartment model (Riebe et al., 2018).

Bioelectrical impedance analysis. Each BIA device measured the resistance and reactance of the participant’s body tissues in reference to a small electrical current produced by the device. Each BIA device was designed to obtain the same information; however, the procedure for implementing each device varied. Therefore, they have each been described below separately.

The RJL Quantum II body composition analyzer (RJL Systems, Clinton Township, Macomb, MI) was used to obtain hand-to-foot BIA measures following protocol guidelines from the manufacturer. Participant age, height (cm), and sex were entered into the device before instructing the participant to step barefoot onto the scale with feet shoulder width apart. Participants were then instructed to hold the display unit with both hands and extend their arms parallel to the floor, while standing upright. A proprietary equation from the device provided measurements of body mass (BM) and percent body fat (%BF), which were then used to calculate the participants FFM (FFM = BM − [BM × (%BF/100)]).

The Tanita BC-534 InnerScan Body Composition Monitor (Tanita, Arlington Heights, IL) was used to obtain foot-to-foot BIA measures per the manufacturer’s guidelines with participants standing barefoot on the footplates. In addition to entering participant age, height (cm), and sex prior to asking the participant to step on the scale, participant activity level was also entered with all participants coded as having a moderately active level of physical activity. Similar to the Omron scale, a proprietary equation from the device provided measurements of BM and %BF, which were then used to calculate the participants’ FFM (FFM = BM − [BM × (%BF/100)]).

Procedure

Using a within-subjects design, participants were assessed at approximately the same time of day on two separate occasions (mean time between sessions: 6.0 ± 7.0 days; mean time of day difference, 0.1 ± 2.3 hr). Participants completed all requisite paperwork for participation upon entering the laboratory for the first session and were provided a brief overview of the assessments they would engage in throughout both sessions. Prior to attending their first session, all participants were instructed regarding the appropriate clothing guidelines but received no further pretest instructions. Minimal instructions were given in order to mimic typical conditions that would exist with the population chosen for the study. After ensuring the appropriate attire, participants then completed each of the body composition tests in the following order: ADP, RJL, Omron, and Tanita BIAs. During the second session, participants completed all body composition assessments in the same order as the first session.
Statistical Analysis

Outcome measures were analyzed as a total sample and also stratified by biological sex. Descriptive statistics including means and SDs were computed for variables of interest (see Table 1). Mean differences between the average of the two trials for each BIA and ADP criterion were analyzed by paired sample t tests. Data were assessed for normality using Shapiro–Wilk tests and graphing of Q–Q plots. Homoscedasticity was also assessed via plots.

The reliability of these measures was computed via intraclass correlation (using analysis of variance), where \( R_{xx} = (MS_R - MS_E)/(MS_R + [MS_C - MS_E/n]) \) for multiple trials and \( (MS_R - MS_E)/(MS_R + [k-1]MS_E + [k/n][MS_C - MS_E]) \) for single trial (Baumgartner & Jackson, 1987), where \( MS_R \) is the mean square for participants, \( MS_C \) is the mean square for FFM, \( n \) is the number of subjects, and \( R_{xx} \) is the reliability coefficient for the measures and standard errors of measurement (SEM). For completeness, both multiple and single-trial reliability were estimated. SEMs were calculated as \( SEM = SD/SQRT (1-R_{xx}) \), where \( SD \) is the SD of the measures. SEM values were calculated in absolute terms (kg FFM).

The validity of these measures was assessed using Pearson correlations comparing FFM estimated by each BIA device to ADP criterion measure. The average FFM measure from the two trials was used for validity calculations. Standard errors of estimate (SEE) were calculated using the formula \( SEE = SD/SQRT (1-r^2) \), where \( SD \) is the SD of the ADP measures and \( r \) is the validity coefficient. In addition, the SEE values were converted to %fat, as this is a more commonly used body composition value.

Results

The FFM and %BM measurements obtained on all assessment techniques are presented in Table 1. Analysis of mean differences with the overall sample indicated that the Omron BIA device significantly underpredicted FFM for participants (\( t=6.66, p < .001, d = 0.17, \) mean difference = 1.92), whereas the RJL and Tanita machines were found to have no significant differences. When data were stratified by biological sex, analysis of mean differences indicated that the Omron BIA device significantly underpredicted FFM for both men and women (males: \( t = 2.17, p = .04, d = 0.15, \) mean difference = 1.05; females: \( t = 7.50, p < .001, \) \( d = 0.55, \) mean difference = 2.52), whereas the RJL significantly overpredicted FFM for the women (\( t = -2.09, p = .04, \) \( d = 0.12, \) mean difference =0.60). In addition, test–retest reliability and concurrent validity analyses were conducted for all assessment techniques presented in Tables 2 and 3, respectively.

Validity

Validity assessment of FFM measures across all three BIA assessment techniques were high in comparison to ADP FFM assessments. For the overall sample, validity coefficients ranged from 0.972 to 0.979, with SEE values averaging 1.68 kg of FFM. The SEE values equate to a range of 2.9–3.8% fat (see Table 3). When stratified by biological sex, validity coefficients for males ranged from 0.934 to 0.942, with SEE values averaging 1.60 kg of FFM. Females were observed to have lower validity coefficients and SEE values compared with men, but this difference was negligible (\( r = .897–.924; \) SEE = 1.28 kg of FFM on average). The SEE values equate to a range of 1.8–2.5% fat for men and 2.1–3.1% fat for women (see Table 3).

Figure 1, 2 and 3 depict Bland–Altman plots for each BIA device compared with the criterion for the sample. The Omron clearly underpredicts FFM in the women in 43 of 46 study participants. Although the prediction error using the Tanita is somewhat random, it appears to overpredict study participants from 0.934 to 0.942, with SEE values averaging 1.60 kg of FFM. The SEE values were also low, averaging 2% of FFM for both men and women (see Table 2).

| Table 2 Reliability Assessment of Body Composition Methods Using Intraclass Correlation |
|---------------------------------|----------------|----------------|----------------|----------------|----------------|
|                                | Overall (\( n = 77 \)) | Men (\( n = 31 \)) | Women (\( n = 46 \)) |
|                                | \( R_{xx} \) | 95% CI | SEM (kg) | \( R_{xx} \) | 95% CI | SEM (kg) | \( R_{xx} \) | 95% CI | SEM (kg) |
| BodPod                         | .996 | .994–.997 | 0.32 | .985 | .969–.993 | 0.92 | .991 | .983–.995 | 0.47 |
| RJL                            | .994 | .990–.996 | 0.35 | .974 | .947–.988 | 1.09 | .979 | .959–.989 | 0.65 |
| Omron                          | .993 | .988–.995 | 0.33 | .983 | .964–.992 | 0.81 | .933 | .879–.963 | 1.03 |
| Tanita                         | .991 | .987–.995 | 0.30 | .973 | .944–.987 | 1.01 | .921 | .856–.956 | 0.90 |
| \( R_{xx} \)                   | \( R_{xx} \) | 95% CI | SEM (kg) | \( R_{xx} \) | 95% CI | SEM (kg) | \( R_{xx} \) | 95% CI | SEM (kg) |
| BodPod                         | .992 | .988–.995 | 0.45 | .970 | .939–.985 | 1.30 | .982 | .967–.990 | 0.67 |
| RJL                            | .987 | .980–.992 | 0.49 | .950 | .900–.976 | 1.52 | .959 | .921–.979 | 0.91 |
| Omron                          | .985 | .977–.991 | 0.49 | .966 | .931–.983 | 1.15 | .874 | .784–.928 | 1.42 |
| Tanita                         | .983 | .974–.989 | 0.42 | .947 | .895–.974 | 1.41 | .853 | .749–.916 | 1.22 |

Note. This table displays the reliability analysis results that include intraclass correlation coefficients, 95% CIs, and SEM. \( R_{xx} \) = the intraclass correlation coefficient for reliability of multiple trials; Single = the intraclass correlation coefficient for reliability of single trials; SEM = standard errors of measurement; CI = confidence interval.
overprediction in women, as the figures show only small deviation from the line of identity that is random in nature.

**Discussion**

It is critical to obtain accurate body composition measures when evaluating individuals for health and physical fitness. This study was designed to compare three popular BIA devices to an ADP criterion measure. Other instruments such as dual-energy X-ray absorptiometry and magnetic resonance imaging have also been used as criterion measures during body composition studies (Fornetti et al., 1999; Laddu et al., 2012). However, both are relatively expensive, and not as likely to be available to many research labs. Thus, we chose ADP as our criterion measure due to its known accuracy and affordability. In addition, we found no studies that evaluated both reliability and validity of these popular BIA machines with the same sample. With many different tools available for body composition analysis, there is a need to determine whether devices are equal in measurement capabilities or if specific tools should be utilized over others. Key findings from this study include high reliability and validity for all devices, with some differences found when stratifying by biological sex. The RJL BIA demonstrated the best FFM estimates for both men and women.

Mean differences between the RJL and Tanita BIA instruments and the criterion measure did not exceed 1-kg FFM. However, the Omron analyzer significantly underpredicted FFM.

**Table 3  Validity Assessment Between BodPod and Bioelectrical Impedance Analysis Measures**

<table>
<thead>
<tr>
<th></th>
<th>Overall (n = 77)</th>
<th>Males (n = 31)</th>
<th>Females (n = 46)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
<td>SEE (kg)</td>
<td>r</td>
</tr>
<tr>
<td>RJL</td>
<td>.979</td>
<td>1.5</td>
<td>.935</td>
</tr>
<tr>
<td>Omron</td>
<td>.974</td>
<td>1.8</td>
<td>.942</td>
</tr>
<tr>
<td>Tanita</td>
<td>.972</td>
<td>1.7</td>
<td>.934</td>
</tr>
</tbody>
</table>

*Note.* This table displays the validity analysis results with validity coefficients and SEE. *r* = validity coefficient; SEE = standard errors of estimate.
in participants (1.9 kg in total sample, 1.1 kg in men, 2.2 kg in women). Expressed in terms of body fatness, the underprediction of FFM was equivalent to an overprediction of 3.5–4.4% fat or more than twice the usual day-to-day variation found in ADP measures, which is similar to differences found previously between body composition methods (McCrory et al., 1995). With respect to reliability, Fornetti et al. (1999) found the RJL machine to be highly reliable in a sample of female college athletes. Specifically, their findings were similar to ours, with $R_{\text{c}} = .987$ and SEM = 1.1-kg FFM.

Study weaknesses include possible limited generalizability of the results to populations other than those included in this study. For example, future research should include individuals with a wider age and weight range. The sample was chosen out of convenience but contained both men and women whose physical characteristics represent apparently healthy college-aged adults in the United States, as well as individuals likely to be used in many research studies. Other populations may not necessarily fall into healthy body fat percentage ranges; therefore, this limits generalizability to underfat and overfat samples. In addition, sample size is also a limitation. Study strengths include the use of multiple BIA devices for comparison to the criterion ADP measure.

Although our results showed very high reliability and validity for all three BIA devices, compared with ADP criterion, there were some differences among them. The Omron device underpredicted FFM in women, whereas the prediction error using the Tanita device appeared random. Overall, the RJL BIA appears to show the best FFM estimates for both men and women. The RJL was the only device where we utilized a regression equation (i.e., Lohman Equation) based on previous validation studies, whereas the Omron and Tanita values were obtained from manufacturer derived equations not available to the general public. Perhaps, future research could improve the predictability of these two devices, if the manufacturer allowed investigators access to the actual data used in their equations.

**Acknowledgments**

The study was designed by K.L. Vasold, J. M. Pivarnik, A. C. Parks, M. B. Pontifex; data were collected by A. C. Parks and D. M. L. Phelan; data analysis was conducted by K. L. Vasold and J. M. Pivarnik; interpretation and manuscript preparation were undertaken by K. L. Vasold, J. M. Pivarnik, A. C. Parks, M. B. Pontifex, D. M. L. Phelan. All authors approved the final version of the paper. The authors report no conflicts of interest or funding sources associated with this manuscript. All participants provided written informed consent in accordance with the institutional review board at Michigan State University.

**References**


