The following is the abstract of the article discussed in the subsequent letter:

Clasey, J. L., J. A. Kanaley, L. Wideman, S. B. Heymsfield, C. D. Teates, M. E. Gutgesell, M. O. Thormer, M. L. Hartman, and A. Weltman. Validity of methods of body composition assessment in young and older men and women. J. Appl. Physiol. 86: 1728–1738, 1999.—We examined the validity of percent body fat (%Fat) estimation by two-compartment (2-Comp) hydrostatic weighing (Siri 2-Comp), 3-Comp dual-energy X-ray absorptiometry (DEXA 3-Comp), 3-Comp hydrostatic weighing corrected for the total body water (Siri 3-Comp), and anthropometric methods in young and older individuals (n = 78). A 4-Comp model of body composition served as the criterion measure of %Fat (Heymsfield 4-Comp; S. B. Heymsfield, S. Lichtman, R. N. Baumgartner, J. Wang, Y. Kamen, A. Aliprantis, and R. N. Pierson Jr., Am. J. Clin. Nutr. 52: 52–58, 1990.). Comparison of the Siri 3-Comp with the Heymsfield 4-Comp model revealed mean differences of $\pm 0.4$ %Fat, $r$ values $\geq r = 0.997$, total error values $\leq 0.85$ %Fat, and 95% confidence intervals (Bland-Altman analysis) of $\pm 1.7$ %Fat. Comparison of Siri 2-Comp, DEXA, and anthropometric models with the Heymsfield 4-Comp revealed that total error scores ranged from $\pm 4.0$ to $\pm 10.7$ %Fat, and 95% confidence intervals associated with the Bland-Altman analysis ranged from $\pm 5.1$ to $\pm 15.0$ %Fat. We conclude that the Siri 3-Comp model provides valid and accurate body composition data when compared with a 4-Comp criterion model. However, the individual variability associated with the Siri 2-Comp, DEXA 3-Comp, and anthropometric models may limit their use in research settings. The use of anthropometric estimation methods resulted in large mean differences and a considerable amount of interindividual variability. These data suggest that the use of these techniques should be viewed with caution.

Regarding the validity of methods of body composition assessment in young and older men and women

To the Editor: In their recent paper, Clasey and co-workers (3) compared several different methods of estimating body fat. The authors took as their reference method a 4-compartment model involving the measurement of body density, total body water, and total body bone ash (5). They then compared other methods of measuring body fat with this reference method using regression analysis and the Bland-Altman statistical analysis (2). Great care was taken in obtaining accurate measurements for all the methods compared. However, in Table 3 of the paper (3), I was concerned to see that comparison regressions for a second 4-compartment model against the reference model yielded correlation coefficients of 1.000. This set “statistical alarm bells” ringing in my head, as it is very unusual to get “perfect” correlations between independent measurements of physiological variables. On inspection of the source papers for the methods involved (1, 5) and after some arithmetic calculations, it became clear that the two methods being compared used the same variables, i.e. body density, total body water, and total body bone ash, in what could be reduced to essentially identical equations. Therefore, the excellent agreement between these methods is to be expected. Elsewhere in the paper, other methods that depend on body density (6, 7) were compared with the 4-compartment reference method. Here again, the variables are not completely independent and so any correlation between them is artifically enhanced. This made the comparisons involving methods measuring body density agree better with the reference method than with other methods involving dual-energy X-ray absorptiometry or anthropometric measurements. Although the comparisons may be true, I don’t think that the results of their study as they stand can be used to prove them.

In further comparisons, the authors correctly choose the Bland-Altman statistic; however, from the figures in the paper (3), it appears that this technique has been misapplied. In the Bland-Altman statistic, the difference between two methods of measuring the same entity is plotted as the dependent variable against the mean of the two methods as the independent variable. By plotting the results this way, it is possible to check for any bias that depends on the magnitude of the measurement. For instance, in the present case, does the agreement between the methods vary significantly as the subjects become fatter? If it does, a significant correlation would be found between the dependent and independent variables. Unfortunately, it appears that the authors have plotted the difference between the methods as the dependent variable against the reference method alone as the independent variable instead of the average of the two methods. Although this doesn’t seem to be a major crime, it can produce misleading correlations. If we have two methods of measuring fat resulting in pairs of data points, F1 and F2, where F1 represents the reference method data, then, if both methods are reasonably accurate, we would expect the results to correlate in a linear fashion

$$F_2 = m \cdot F_1 + c$$

where $m$ is the gradient and is close to unity and $c$ is the intercept and is close to zero.

If we take the difference of the two methods as per the Bland-Altman method, then

$$F_2 - F_1 = (m - 1)F_1 + c$$

It is now noticed that, even if $F_1$ and $F_2$ are unrelated to each other, i.e., $m = 0$, $F_2 - F_1$ will be related

To the Editor: I would like to offer some comments on the paper by Clasey et al. (8750-7587/00 $5.00$ Copyright © 2000 the American Physiological Society http://www.jap.org

2518

8750-7587/00 $5.00$ Copyright © 2000 the American Physiological Society

http://www.jap.org
to F1 with a gradient of −1 [in fact the only situation where (F2 − F1) is not related to F1 is if m = 1]. The bottom line is that plotting the difference between the two methods against the reference method alone can result in spurious correlations (4). If the authors have made this mistake, then they are not the only ones; I have certainly done so, and the medicscientific literature is peppered with the efforts of others who have also done this.

In summary, the authors have performed a pain-taking study involving many careful measurements; however, some of the conclusions they reach may be invalid because of misapplied statistical methods.

REFERENCES

W. S. Watson
Nuclear Medicine Department
Southern General Hospital
Glasgow G51 4TF, Scotland, United Kingdom
E-mail: wswatson.nmsgh@virgin.net

REPLY

To the Editor: We appreciate Professor Watson’s insightful comments regarding our recent paper comparing several methods of estimating body composition in younger and older adults (4). We agree that excellent agreement would be expected between the equations of Baumgartner et al. (1) and Heymsfield et al. (5). These equations were derived using the same variables. The difference between these two equations is that the Baumgartner et al. equation was derived on a sample of older adults and the Heymsfield et al. equation was derived using a wider age range. Either of these two equations could have been chosen as the criterion method. Because our sample included both younger and older adults, we chose the Heymsfield et al. equation as the criterion in our study.

Professor Watson is correct that several of the methods that were compared with the 4-compartment reference method use components of the 4-compartment method (i.e., body density, total body water) and as such the variables are not completely independent. One of the purposes of our study was to determine whether all 4 components were needed to reasonably estimate body composition. Although we reported the correlations between percent body fat measured using the criterion method and percent fat estimated using the prediction techniques, we did not base any of our conclusions on the strength of the reported correlations.

The issue regarding the use of the Bland-Altman statistic is one that we discussed at length before submitting the manuscript. Although it is correct that Bland and Altman (2, 3) argue that the use of the mean of the two methods being compared should be used as the independent variable, they recognize (3) that several groups have suggested that, when one method is regarded as a “gold standard,” it is presumably more accurate than the other method and therefore the difference should be plotted against the gold standard (6, 7). Furthermore, Bland and Altman also point out that the spurious correlation between the difference (predicted – criterion) and the criterion will be small when the methods themselves are highly correlated (3).

It should be realized that the choice of the criterion method or the mean of the two methods as the independent variable for the Bland-Altman plots does not affect the calculation of the mean difference or the 95% confidence intervals. Because this is the most critical information obtained from the analysis, we based our conclusions on these data. The choice of independent variable will affect the placement of the data points along the x-axis and as such could affect the regression line (bias). Because there is no consensus regarding choice of the independent variable (2, 3, 6, 7) and because our results were based on estimates of the

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>x</th>
<th>y</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baumgartner 4-Comp</td>
<td>y = −0.118 × x + 3.83</td>
<td>−0.25</td>
<td></td>
</tr>
<tr>
<td>Siri 3-Comp</td>
<td>y = −0.003 × x + 0.29</td>
<td>−0.05</td>
<td></td>
</tr>
<tr>
<td>DEXA</td>
<td>y = −0.0001 × x − 0.74</td>
<td>−0.0002</td>
<td></td>
</tr>
<tr>
<td>Siri 2-Comp</td>
<td>y = −0.001 × x − 1.73</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Jackson Pollock</td>
<td>y = 0.210 × x + 0.62</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td>Tran Weltman</td>
<td>y = 0.176 × x − 5.61</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>Williams et al.</td>
<td>y = 0.204 × x − 2.09</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent variable</td>
<td>x</td>
<td>y</td>
<td>r</td>
</tr>
<tr>
<td>Baumgartner 4-Comp</td>
<td>y = −0.119 × x + 3.84</td>
<td>−0.25</td>
<td></td>
</tr>
<tr>
<td>Siri 3-Comp</td>
<td>y = −0.001 × x + 0.23</td>
<td>−0.02</td>
<td></td>
</tr>
<tr>
<td>DEXA</td>
<td>y = −0.116 × x − 4.00</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>Siri 2-Comp</td>
<td>y = 0.126 × x − 5.26</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>Jackson Pollock</td>
<td>y = 0.357 × x − 4.09</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>Tran Weltman</td>
<td>y = 0.299 × x − 8.98</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>Williams et al.</td>
<td>y = 0.357 × x − 6.71</td>
<td>0.56</td>
<td></td>
</tr>
</tbody>
</table>

r, Correlation coefficient; Comp, compartment; DEXA, dual-energy X-ray absorptiometry.
confidence limits for agreement, we chose not to report bias. As can be seen from the regression models and correlation coefficients presented in Table 1, as expected, in some instances the choice of the independent variable did have an effect on the estimates of bias. This is reflected by higher values for slope and increased correlation coefficients for some variables when the criterion measure is used as the independent variable.

In the Clasey study, we based our conclusions on the mean differences and the 95% confidence intervals associated with the Bland-Altman analyses. As such, we believe that the conclusions are valid and stay within the limits of our data.

REFERENCES


Arthur Weltman
Department of Human Services
General Clinical Research Center
Exercise Physiology Laboratory
University of Virginia
Charlottesville, Virginia 22908
E-mail: alw2v@virginia.edu