Multicomponent Cross-Validation of Minimum Weight Predictions for College Wrestlers

R. RANDALL CLARK, JUDE C. SULLIVAN, CYNTHIA BARTOK, and DALE A. SCHOELLER

U.W. Hospital Sports Medicine Center, University of Wisconsin, Madison, WI

ABSTRACT

CLARK, R. R., J. C. SULLIVAN, C. BARTOK, and D. A. SCHOELLER. Multicomponent Cross-Validation of Minimum Weight Predictions for College Wrestlers. Med. Sci. Sports Exerc., Vol. 35, No. 2, pp. 342–347, 2003. In 1998, the National Collegiate Athletic Association (NCAA) adopted a rule requiring that skinfolds (SF) or hydrostatic weighing (HW) be used to estimate minimum weight (MW) in college wrestlers. Purpose: To cross-validate the NCAA methods for estimation of MW using a multicomponent criterion (4C). Methods: Criterion MW was calculated from body density (BD), bone mineral content (BMC), and total body water (TBW) using the 4C equation of Lohman (1992). BMC was measured by dual energy x-ray absorptiometry (DXA), TBW by deuterium dilution, and BD by HW. Subjects were Division I athletes from the University of Wisconsin (mean ± SD; N = 33, age = 19.5 ± 1.3 yr, height = 177.3 ± 7.8 cm, weight = 74.2 ± 9.3kg). Results: There was no significant difference between mean MW from HW (69.6 ± 8.5 kg) and SF (70.1 ± 8.3 kg) (P = 0.113), and between mean MW from HW (69.6 ± 8.5 kg) and 4C (69.5 ± 8.6 kg) (P = 0.46). A clinically small, yet significant difference was seen when comparing mean MW from SF to 4C (P = 0.013). The regression for the relationship between 4C and HW (y = 0.994 × HW + 0.294 kg, R² = 0.985) and 4C and SF (y = 1.019 × SF − 1.885 kg, R² = 0.797) did not significantly deviate from the line of identity. Pure error (PE) values of 1.04 kg and 1.35 kg were found for HW and SF, respectively. The difference between the methods was plotted as a function of the 4C criterion. The regression line for HW and 4C (y = −0.009x + 0.743, r = −0.07, P = 0.69) and SF and 4C (y = −0.038x + 3.259, r = −0.27, P = 0.13) suggest that no systematic differences in the prediction were associated with the size of the criterion. Conclusion: These data support the NCAA methods of HW and SF to predict MW when cross-validated using a 4C criterion in this sample. Key Words: BODY COMPOSITION METHODS, HYDROSTATIC WEIGHING, DUAL ENERGY X-RAY ABSORPTIOMETRY, TOTAL BODY WATER, SKINFOLDS

In 1998, after the tragic deaths of three college wrestlers while cutting weight, the National Collegiate Athletic Association (NCAA) adopted a rule requiring minimum weight (MW) testing for college wrestlers. Consistent with the guidelines set forth by the American College of Sports Medicine (19), the American Medical Association (2), and similar to programs used by state high school associations (8), the NCAA required assessment of body composition for each wrestler before the competitive season. The body composition value is used to calculate minimum wrestling weight (MW) at 5% body fat (%BF).

The NCAA Competitive Safeguards Committee selected the Lohman skinfold (SF) equation (17) as an effective, noninvasive field method to estimate body fat. Although it is a robust prediction equation with documented validity in high school wrestlers (6,7,27), it was developed without the benefit of modern advancements in body composition technology. As more sophisticated tools to analyze body composition become available, it is important to evaluate the accuracy of the method due to the implication for college wrestling.

Body density (BD) as determined by hydrostatic weighing (HW) is also approved by the NCAA for estimation of MW. Two-compartment models such as HW have been commonly used to evaluate body composition techniques in wrestlers (6,7,10,14,22,23,26,27). However, this approach has been questioned because of the inherent assumptions of the two-compartment technique (12,15). These assumptions include consistency in hydration and the density of the fat-free mass. The advent of more sophisticated and comprehensive multicomponent models, which include analysis of bone mineral content (BMC) and total body water (TBW), allow a more rigorous cross-validation (12,15). It

0195-9131/03/3502-0342/$3.00/0
MEDICINE & SCIENCE IN SPORTS & EXERCISE®
Copyright © 2003 by the American College of Sports Medicine
DOI: 10.1249/01.MSS.0000048832.87882.60

342
has been suggested that the four-compartment (4C) model be used as the "gold standard" in adults (15). This approach addresses the concerns of the two-component model because there is independent assessment of BD, TBW, and BMC.

Although this approach is preferred, it is impractical on a large scale because of the cost and instrumentation needed to measure TBW and BMC. Even though the advantages of the 4C approach are clear, this model has not been used to cross-validate the NCAA methods for wrestlers. Therefore, the purpose of this study was to cross-validate the NCAA methods to estimate minimum weight by using a 4C criterion.

**METHODS**

Subjects were 33 Division I athletes from the University of Wisconsin (Table 1). All measurements were taken during a single early-morning testing session under standardized conditions by the same investigators after an 8–12 h fast. A strict preparticipation protocol prohibited any workouts, ingestion of caffeine or alcohol, and encouraged normal fluid intake for 12 h before the evaluation. All participants were asked to void and defecate before beginning the procedures. The study was approved by the Clinical Science Center Human Subjects Committee at the University of Wisconsin, and informed written consent was obtained before initiating the testing protocols.

**Anthropometry.** Anthropometric measurements included height, weight, and SF. Height was measured with a stadiometer to the nearest 0.64 cm. Weight was measured on a beam-balance platform scale to the nearest 0.11 kg. SF sites were measured with a Lange caliper (Cambridge, MD). All measurements were taken on the right side of the body in serial fashion by the same experienced investigator. Reliability of the investigator has been previously documented (20). SF thickness was based on the median of three trials according to the NCAA protocol (18). BD was predicted by the NCAA Lohman equation (17). %BF was estimated from BD by using the equation of Brozek et al. (5). BW was calculated at 5%BF (fat-free body/0.95).

**Densitometry.** BD was determined by HW at residual volume (RV), as described by Behnke and Wilmore (4). Eight to 10 trials were obtained for each subject. The mean of the three heaviest trials was used as the underwater weight for calculating BD. RV was measured by the oxygen dilution method described by Wilmore (30), using a modified Collins 13.5-L respirometer (Braintree, MA) and a Med Science model 505 N2 analyzer (St. Louis, MO). The subject’s RV was measured outside the tank in a seated position simulating that used during HW. The mean of two trials within 75 mL was used as the RV. An additional correction of 100 mL, to account for gastrointestinal gas, was used in the HW calculation (4). %BF was estimated from BD using the equation of Brozek et al. (5) [%BF = ((4.57/BD) − 4.142) × 100] and MW was calculated at 5%BF.

**Dual energy x-ray absorptiometry (DXA).** Whole-body scans were performed using the Norland XR-36 whole-body bone densitometer (Norland Corporation, Ft. Atkinson, WI) with software version 3.7.4/2.1.0. The XR-36 x-ray tube operates at 100 kV and uses samarium filtration (K-edge at 46.8 keV) to produce energy peaks at maximum of 40 and 80 keV. Dual NaI detectors measure the attenuated x-ray using a pixel size of 6.5 × 13.0 mm and a scan speed 260 mm·s⁻¹. The coefficient of variation (CV) for total BMC is 0.9%. Subjects were scanned by the same investigator in the supine position after removing metal objects or clothing containing metal components. Each scan session was preceded by a calibration routine using multiple quality control phantoms that simulate soft tissue and bone.

**Total body water.** TBW was calculated using the deuterium dilution technique described by Schoeller (21). Saliva samples (10 mL) were collected before and 3 h after dosing with 4-g deuterium oxide. The dose was followed with a 100-mL rinse to ensure complete ingestion of the tracer. The samples were centrifuged to pack down particulate matter and then stored in cryogenic tubes at −40°C until further processing. Deuterium enrichment of saliva samples was quantified using isotope ratio mass spectrometry (Finnigan MAT Delta Plus, Bremen, Germany) with a chromium reduction system and an A 300 autosampler. TBW was corrected as dilution space/1.04 to account for hydrogen exchange with protein and carbohydrate during the 3-h equilibration period (21).

**Multicomponent model.** The criterion body fat was calculated with a 4C model by using BD from HW, TBW from deuterium labeled water, and BMC from DXA according to the equation modified from Lohman (3,15,16):

\[
\%BF = \left(\frac{(2.747/BD) - (0.714 \cdot W) + (1.146 \cdot B) - 2.0503}{100}\right)
\]

where BD is equal to the body density from HW, W is the TBW expressed as a fraction of body weight, and B is body mineral (corrected for the nonosseous mineral portion of the body (5) by using the estimate of bone mineral ash from DXA) expressed as a fraction of body weight. MW using %BF calculated from the 4C model was used as the criterion method.

**Statistical analysis.** Comparison of methods included the criterion referenced cross-validation methods suggested by Lohman (17), Guo and Chumlea (11), and the methods comparison of Altman and Bland (1). This included mean comparisons of MW using a one-way repeated ANOVA with Tukey follow-up test for significance. Statistical significance was set at \( P < 0.05 \). The means and Pearson product moment correlations are presented for descriptive purposes and for comparison with previous studies. Confidence intervals for the mean difference in MW by the

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean ± SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yr)</td>
<td>19.5 ± 1.3</td>
<td>18–22</td>
</tr>
<tr>
<td>Body weight</td>
<td>74.2 ± 9.3</td>
<td>55.9–91.6</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>177.3 ± 7.8</td>
<td>160.7–191.1</td>
</tr>
<tr>
<td>Body density (HW)</td>
<td>1.075 ± 0.008</td>
<td>1.056–1.094</td>
</tr>
<tr>
<td>Percent fat (4C)</td>
<td>11.0 ± 2.9</td>
<td>6.3–18.0</td>
</tr>
<tr>
<td>MW (4C)</td>
<td>69.5 ± 6.6</td>
<td>50.8–86.8</td>
</tr>
</tbody>
</table>

Body density is calculated by hydrostatic weighing (HW) and percent fat and minimum weight (MW) are calculated from the multicomponent criterion (4C).
TABLE 2. Mean ± SD percent body fat (%BF), minimum weight (MW), the SEE, and the pure error (PE).

<table>
<thead>
<tr>
<th>Method</th>
<th>%BF</th>
<th>MW</th>
<th>SEE</th>
<th>PE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4C</td>
<td>11.0 ± 2.9</td>
<td>68.5 ± 8.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HW</td>
<td>10.8 ± 3.2</td>
<td>66.6 ± 8.5</td>
<td>1.03</td>
<td>1.04</td>
</tr>
<tr>
<td>SF</td>
<td>10.2 ± 2.9</td>
<td>70.1 ± 8.3</td>
<td>1.25</td>
<td>1.35</td>
</tr>
</tbody>
</table>

Mean ± SD values for percent body fat (%BF), minimum weight (MW), the SEE, and the pure error (PE) are used to compare the MW from hydrostatic weighing (HW) and skinfolds (SF) with a multicomponent criterion (4C).

methods were calculated. The prediction was considered accurate if the regression between the methods had a slope not significantly different than one and intercept not significantly different from zero in this independent cross-validation sample (11). This tests the hypothesis that the regression of MW by the 4C criterion (y-axis) and the prediction methods (x-axis) do not significantly deviate from the line of identity. Analysis also included the R², the standard error of estimate (SEE) where SEE = \( \sqrt{\frac{\text{SSE}}{\text{df}}} \) and the pure error (PE) as described by Guo and Chumlea (11) where PE = \( \sqrt{\text{SSE}}/ \text{df} \) when \( Y_1 \) = the predicted value and \( Y_2 \) = the criterion value. To identify systematic differences between the prediction and the criterion, the difference between the methods (y-axis) was plotted as a function of the criterion (x-axis) according to Altman and Bland (1). The limits of agreement are indicated as the average difference between the methods ± 2 SD of the mean difference.

RESULTS

The physical characteristics of the subjects are summarized in Table 1. The means ± SD for MW and %BF, and the SEE and PE for each method are summarized in Table 2. There was a strong association (\( r = 0.99 \)) and no significant difference (\( P = 0.113 \)) between mean MW from HW (69.64 ± 8.54 kg) and SF (70.07 ± 8.30 kg). The mean difference in MW between HW and SF was −0.43 kg, with a confidence interval of −0.98 and 0.11 kg. There was a strong association (\( r = 0.99 \)) and no significant difference (\( P = 0.46 \)) between mean MW from HW (69.64 ± 8.54 kg) and the 4C criterion (69.50 ± 8.55 kg). The mean difference was 0.13 kg, with a confidence interval of −0.23 and 0.51 kg. A clinically small, yet significant difference (\( P = 0.013 \)) was seen when comparing mean MW from the NCAA SF prediction (70.07 ± 8.30 kg) to 4C (69.50 ± 8.55 kg). The mean difference was 0.57 kg, with a confidence interval of 0.13 and 1.01 kg.

Figure 1 illustrates the regression when 4C is the dependent variable (y-axis) and HW is the independent variable (x-axis). The technique was considered accurate if the regression between the methods did not deviate from the line of identity. The regression line between 4C and HW (\( y = 0.994 \times \text{HW} + 0.294 \text{ kg}, R^2 = 0.985 \)) had a slope not significantly different from one (\( P = 0.781 \)) and an intercept not significantly different than zero (\( P = 0.849 \)). Figure 2 illustrates the regression between 4C and SF. The regression line between 4C and SF (\( y = 1.019 \times \text{SF} - 1.885 \text{ kg}, R^2 = 0.979 \)) had a slope not significantly different from one (\( P = 0.486 \)) and an intercept not significantly different than zero (\( P = 0.323 \)).

HW explained 98.5% of the variance in MW from 4C with a SEE of 1.03 kg. SF explained 97.9% of the variance in MW from 4C with a SEE of 1.25 kg. The PE values for the comparison of 4C to HW and 4C to SF were 1.04 kg and 1.35 kg, respectively.

FIGURE 1—The agreement between HW estimation of MW and the 4C criterion. The 4C criterion is the dependent variable (y-axis) and HW is the independent variable (x-axis). The dashed line represents the line of identity. The regression analysis (slope and intercept) and squared multiple correlation are included. The slope and intercept of the regression line were not significantly different from one and zero, respectively.

FIGURE 2—The agreement between the SF estimation of MW and the 4C criterion. The 4C criterion is the dependent variable (y-axis), and SF is the independent variable (x-axis). The dashed line represents the line of identity. The regression analysis (slope and intercept) and squared multiple correlation are included. The slope and intercept of the regression line were not significantly different from one and zero, respectively.
Figures 3 and 4 evaluate systematic differences in the methods by plotting the difference between the methods as a function of the criterion. This approach examines the discrepancy between the technique in question and the criterion as a function of MW determined by 4C. In Figure 3 the regression line for the plot \( y = -0.009x + 0.743 \) illustrates that no systematic bias was observed between HW and 4C across the range of MW. A nonsignificant correlation \( (r = -0.07) \), with a slope not significantly different than zero \( (P = 0.692) \) and an intercept not significantly different than zero \( (P = 0.631) \), was found. In Figure 4, the regression line for the plot \( y = -0.038x + 3.259 \) illustrates that no systematic bias was observed between SF and 4C. A nonsignificant correlation \( (r = -0.27) \), with a slope not significantly different than zero \( (P = 0.133) \) and an intercept not significantly different than zero \( (P = 0.073) \), was found.

**DISCUSSION**

The purpose of this investigation was to cross-validate the methods used by the NCAA to predict MW in collegiate wrestlers. Cross-validation is the application of a predictive equation to a sample independent from the one used to construct the equation \( (11) \). The NCAA SF method and HW were evaluated against a 4C criterion by using the methods suggested by Lohman \( (17) \), Guo and Chumlea \( (11) \), and Altman and Bland \( (1) \). MW using a 4C approach was selected as the criterion because it utilizes actual measurement of BMC and TBW, two of the assumptions present in
densitometry. The results suggest the prediction methods were accurate because the regression lines did not significantly deviate from the line of identity. Large R² values and low SEE and PE values were seen. The difference in methods was not related to the size of the criterion across the range of MW. These data suggest that HW and the NCAA SF equation were accurate predictors of MW in this sample.

The Lohman SF equation (17) used by the NCAA is a robust prediction that has withstood the scrutiny of previous cross-validation studies in high school wrestlers (6,7,27). However, previous cross-validation studies have utilized densitometry as the criterion. Using densitometry in wrestlers may have limitations because of the assumptions of the two-component technique. These assumptions include consistent hydration and density of the fat-free mass. Wrestlers have been reported to exhibit dehydration during periods of weight cutting (19,24,28,29) and may have more advanced muscular development than their age-matched nonathletic counterparts.

The more comprehensive multicomponent models, which include analysis of BMC and TBW, allow a more complete evaluation for this group. Although multicomponent models are clearly not feasible on a large scale, it is important to cross-validate the field method that will determine a college athlete’s competitive weight class. Therefore, a logical extension of previous research (14,22,23,25,26) was to evaluate the NCAA methods (HW and SF) with a 4C criterion.

Guo and Chumlea (11) recommend the PE be used to measure the performance of a prediction equation on cross-validation. The PE has been described by some investigators as the best single variable for evaluating true differences between a prediction and criterion (27). The PE accounts for both the SEE and the mean difference between predicted and measured values (17). Lohman (17) describes it as an evaluation of a prediction that is superior to both the SEE and correlation. Only when the means are equal will SEE = PE. Whereas the SEE reflects deviations of predicted versus criterion MW about a regression line that fits the scatter plot, PE is reflective of the deviations around the ideal line with intercept zero and slope of one (17,27). PE values of approximately 2.0 kg have been considered excellent in samples similar to this one (26,27). The PE values in the present study were 1.04 kg and 1.35 kg for HW and SF, respectively. Although a PE value that would denote a successful cross-validation has not been set (11), these PE values were lower than previous studies (7,27) and less than the 2.0 kg reported as excellent. The smaller the PE, the greater the accuracy of the equation when applied to an independent sample (11).

An examination of systematic bias in the NCAA methods was evaluated by plotting the differences between the methods (y-axis) as a function of 4C (x-axis) according to Altman and Bland (1). Regression lines with slopes and intercepts not significantly different than zero indicate no systematic bias in the techniques across the range of MW seen in the subjects. No systematic bias was seen in the regression lines for the HW and 4C plot (y = −0.009x + 0.743, r = −0.07, P = 0.69) and the SF and 4C plot (y = −0.038x + 3.259, r = −0.27, P = 0.13) (see Figs. 3 and 4).

It is important to acknowledge the limitations of this study. Body fatness can only be estimated more accurately by using a multicomponent approach if BD, TBW, and BMC are accurately and precisely measured (15). Measuring each variable has inherent limitations. BD by HW requires subject compliance, total submersion in water, maximal exhalations, and accurate measurement of RV. Accurate assessment of TBW is subject to a number of factors including subject compliance during an equilibration period of three hours. BMC by DXA is subject to differences in the three manufacturer’s calibration standards and algorithms used to determine bone, fat, and lean from x-ray attenuation. Although, in theory a 4C model will provide a more complete evaluation of an individual’s body composition, it is not without error. It is important to remember that each technique is subject to biological and technical error. Whether the null hypothesis is rejected or not, there is always a question of whether the sample size was adequate to achieve appropriate testing power or estimation precision. Much confusion surrounds this issue in the applied literature. However, the current prevailing view in the mainstream statistical literature is that post hoc power calculations may not be appropriate and should be replaced with confidence intervals. An example of this view is Hoenig and Heisey (13), who present a rigorous mathematical demonstration of the limitations of post hoc power analyses, such as those advocated by Cohen (9), and the desirability of confidence intervals instead. Therefore, the most rigorous analysis has been applied in this study and results are presented with confidence intervals. In addition, these subjects are from a single university and a larger and more diverse sample of wrestlers from multiple universities will be helpful in identifying differences in the methods.

There was a strong association and no significant mean difference between MW from HW and SF. This is to be expected as the Lohman SF equation (17) was developed using HW as the criterion. There was a strong association and no significant mean difference between HW and 4C. A clinically small, but significant, mean difference was seen between 4C and SF. The 4C model yielded a larger %BF than SF, resulting in a significantly lower MW. Therefore, the corresponding estimate of MW from SF is slightly more conservative than the estimate from 4C. From a health standpoint, these results suggest that the difference between 4C and SF (albeit clinically small) is in the “safe” direction and this more conservative estimate may provide a margin of safety for wrestlers. This may reduce the potentially negative health effects associated with competing at an excessively low body weight.

In summary, this investigation cross-validated the methods used by the NCAA to estimate MW in collegiate wrestlers. The results suggest the methods were accurate because the regression lines did not significantly deviate from the line of identity. Low SEE and PE values and large R² values were seen in this cross-validation sample. Nonsignificant correlations and regression lines with slopes and intercepts...
not different than zero for the difference in the methods showed no systematic bias across the range of MW. This suggests the difference in methods was not related to the size of the criterion. Under the conditions of the study, these data support the NCAA methods as accurate predictors of MW in this cross-validation sample when evaluated using a 4C criterion.

REFERENCES


