A System for Evaluating Powerlifting and Other Multi-Event Performances

Running head: Evaluating Powerlifting Performance

Phillip A. Bishop¹, Tyler D. Williams², Alex N. Heldman², and Paul M. Vanderburgh³

¹ Liberty University, Lynchburg, VA, ²University of Alabama Exercise Science Laboratory, Tuscaloosa, AL
³ University of Dayton, Dayton, OH

Corresponding author:

Phillip Bishop
1971 University Blvd
Liberty University
Lynchburg, VA 24515
pbishop@Liberty.edu
mailto:pbishop@bama.ua.edu
(205) 799-7504

Conflict of Interest

No conflicts of interest are declared by the authors.

Disclosure of Funding

No funding was received for the study.
ABSTRACT
Currently the sport of Powerlifting discriminates against bench press scores and uses an empirical equation which is very difficult to update to determine the best lifter within sex. The purpose of this study was to provide a simple and statistically sound method for evaluating Powerlifting performance which could also be used with other multi-event sports. Data were collected from a large public database on the top 50 (men) or the top 30 (women) individual raw competition scores in each weight class for each of the three Powerlifting lifts. Means and standard deviations were then calculated for each lift and weight class by sex. Powerlifting performance was evaluated by the sum of the three z-scores for the three lifts divided by three (the highest score wins). The z-scores reflect a dimensionless number which can be used to evaluate each lift and the total of the 3 lifts across weight classes and between sexes. Although the means and standard deviations should be relatively stable, this computation is transparent and can be readily updated as lifters improve. This system overcomes the bias against bench press and many of the measurement challenges in evaluating athletes in multi-event sports, in multiple weight divisions, and between sexes. These findings can benefit coaches and athletes by allowing simpler evaluations of performance and improvement.

Key words: performance evaluation, z-scores, Wilks equation, multi-measurement scoring, sport comparisons
INTRODUCTION

In some athletic competitions, evaluations of, or comparisons among, athletes are relatively simple and straightforward. In other athletic competitions, comparisons among athletes in different age and weight classifications or between sexes, is more challenging. For example, Powerlifting is a sport providing opportunities for women and men of all sizes to train and compete in a strength-based sport. To establish best lifter in a weight class, the three lifts (bench press, squat and deadlift) are summed, disregarding that each pound of weight lifted for each lift is a different unit (e.g. a bench press pound lifted represents a larger increment than a pound lifted in deadlift or squat). This error induces a bias in favor of the deadlift, typically the heaviest lift of a powerlifting competitor. Consequently, it is recognized that lifters who win the overall category (sum of the three lifts) are particularly adept in the deadlift.

A second problem of Powerlifting and Olympic weightlifting has nothing to do with the difference in measurement of the three lifts, but rather is a result of the allometric relationship between body mass and strength. In brief, empirical research indicates that as muscle mass grows, strength grows at a slower rate (3-7, 9, 10). Therefore, the simple approach of dividing the weight lifted by bodyweight typically favors the lightest lifters, unfairly resulting in the lighter lifters almost always winning “best lifter”. Consequently the “Wilks formula” was derived and employed to determine, with less bias, the best lifter within a sex for a meet (8). However, the Wilks formula is mathematically complex, has no clear scientific basis, has never been updated to reflect changes in Powerlifting training and competition and only is applicable to Powerlifting.
A third problem illustrated in Powerlifting and Olympic weightlifting, but also true for many athletic competitions is that it is not possible to compare between sexes due to inherent physiological differences. For example, world record performances in most athletic competitions are different between men and women. Consequently, evaluations are typically restricted to within-sex comparisons. While this is widely understood and accepted, there are situations where comparisons between sexes would be useful and interesting.

The purpose of the present study was to develop a method of scoring Powerlifting which provides equal weighting for all three lifts, and allows for determination not only of the best lifter among weight classes, but also between women and men. This approach is intended to remove the favorable bias towards athletes with the highest deadlift, and make Powerlifting scoring equal among the three lifts and should benefit coaches and athletes by allowing simpler and more accurate performance evaluations.

METHODS

Experimental Approach to the Problem

Data for women and men for each of the three Powerlifting lifts were collected for the top American scores for each weight class from a public ranking database on www.powerliftingwatch.com, men’s on September 2\textsuperscript{nd}, 2015 and women’s on November 2\textsuperscript{nd}, 2015. This was done without regard for screening for the use of performance-enhancing drugs,
nor for any restrictions on the use of the equipment used in competition. Additionally, the sport of powerlifting is comprised of two distinct classifications: “raw” and “equipped” powerlifting. In equipped powerlifting, lifters are allowed to use hypercompressive garments, such as a bench shirt or squat suit, to aid in the performance of their lifts (2). These garments come in a single-ply or multi-ply construction that allows them to store elastic energy during the eccentric phase of the lift, which provides a significant performance enhancement during the concentric phase. Raw powerlifting does not allow the use of these supportive garments, but depending on the federation, may allow the use of knee wraps. Due to the popularity of raw powerlifting over equipped powerlifting, only lifts from raw powerlifting competitions were used for this analysis.

Subjects

No human subjects were tested, as these data were from raw Powerlifting competitions in the USA which were published in a public data base. The goal was to collect the highest 50 scores for each lift for each different group of athletes (males and females).

Procedures

Means and standard deviations (SD) were calculated for each lift in each weight class by sex from the top scores of the data base for the samples available.
Statistical Analyses

Z-scores were then calculated for men and women for each lift and weight class using the equation: $z = \frac{\text{lift-mean}}{\text{SD}}$. The highest z-score for each lift was then selected to provide quantitative information on the approximate percentile rank of the scores to identify outliers.

RESULTS

We were unable to secure scores for the top 50 lifts for women due to insufficient available data. In most weight classes the top 30 women’s scores were used. In two weight classes, not even 30 bench press scores were available (48 kg = 29 scores and 90 kg = 28 scores). Table 1 shows the mean and standard deviation for each weight class for women for each lift. The z-score for the highest lift is also shown. Table 2 shows the mean and standard deviation for each weight class for men for each lift. The z-score for the highest lift is also shown.

INSERT TABLE 1. HERE

INSERT TABLE 2. HERE

As can be seen, there were several men’s individual scores which exceeded 4 standard deviations above the mean. There were two unusual scores among the women, with one z-score above 8, representing an extreme performance. This exaggerated z-score also results from an unusually small standard deviation for that group and lift. These data also illustrate that the deadlift is the highest weight lifted for women of all weight classes and for men in the lightest weight classes.
In contrast the men in the heavier weight classes (125 kg and above) tended to have squatted heavier weights than they deadlifted, on average.

**DISCUSSION**

The purpose of this study was to resolve three problems with current Powerlifting scoring. We approached this by collecting data on top female and male raw Powerlifters across weight classes and computing a mean, standard deviation, and deriving a z-score for each lift. The resulting tables of means and standard deviations can be used to compute a composite z-score for any competitor using the appropriate sex, weight class, and type of lift. We do not have information on the individual training levels or training schedules of the athletes comprising the database we used, but these data were taken from raw competitions. We assume, for the most part, since these were data from formal competitions, that the lifters were well-trained, rested, and prepared for competition, and therefore represent optimal performance. The use of a relatively large sample of 30 or 50 for each weight class has the advantage of providing a relatively stable mean and standard deviation. Again, this is less true for those groups with fewer competitors. A re-examination every year or two should determine the need for z-score updates.

For any Powerlifting competition, regardless of the rules for that particular competition, a z-score can be computed for each individual lift completed by subtracting the mean for that weight class from each person’s score within that weight class and dividing the difference by the standard deviation (1). This z-score is a dimensionless unit. To obtain the person’s total for the
competition, the z-score for the three lifts are added without bias, then divided by 3. Instead of dividing each lifter’s total by 3, the total could be used. The advantage of expressing the score as a mean of the three z-scores is that it keeps the final results interpretable on a normal distribution wherein a z-score of zero equals the mean, and a score of 3 or above represents an outstanding score. To determine the best lifter between and within a sex, the average of the three z-scores can simply be compared, with the highest score winning. Accurate comparisons can be made among the scores for any lift regardless of weight class or sex.

The data in Table 2 are noteworthy on a few points. By using “raw” powerlifting data unrestricted by rules against performance-enhancement, the z-scores should be usable for all competitions with the understanding that more restrictive competitions will achieve lower z-scores. However, since all restrictions are equally applied to all competitors, the z-scores still permit a valid combination of lifts and comparisons among athletes at a given competition. E.g. in the case of equipped lifting competitions, the same table could be used with the understanding that the scores will be higher for equipped lifting than for raw. Relative performances among competitors should still be accurate.

The advantages of this approach are manifold. First, athletes can easily learn to convert their own raw scores to z-scores and understand how they compare to the best 50 (men) or 30 (women) unrestricted raw scores in that lift. A score less than zero indicates they are below average relative to the competitive population. A score over 3 suggests they are in the top 1% for that lift.
Some of the reported z-scores in Tables 1 and 2 are extremely high, in that a z-score of 3.99 represents the top 0.01% of that sample. Clearly z-scores above this would typically be unexpected; however, the present calculations are based upon only the top 30 and 50 data points, and these extremely high z-scores may reflect a small pool of competitors, unusual genetics, unusual training, or other atypical responses. In some cases, the total population of Powerlifters by sex in a class is small, particularly for women. Likewise, in all cases these do not always represent a perfect normal distribution. However, there is no reason to expect that these z-scores would not be useable, and updating the data is relatively easy and transparent to athletes.

The combined score indicates overall performance. The scores can be compared among all the lifters at any competition. One weakness might be that any particular class of lifters may have an exceptional group of individuals who bias the mean score for a lift or lifts, putting that class of lifters at some disadvantage. This is more likely to happen in the lighter weight categories where the population of competitors is smaller. Consequently, the top 30/50 should be reviewed periodically using the methods we described in order to ensure that the mean and standard deviation are reflective of the best 30/50 lifts. It should be expected that women’s scores will change more rapidly than men’s as illustrated by Whipp & Ward (11).

This approach negates the advantage of being strongest in deadlift, and the disadvantage of being strongest in the bench press, and provides a more transparent means for determining the best lifter by both lift type and over-all total, as well as between sexes. Whereas the Wilks formula is
Evaluating Powerlifting Performance

Currently the most common method of determining the overall best lifter in a competition, it has its disadvantages. First, it does not allow comparisons across sexes. Second, its adjustment standards are based entirely on world best performances, not on a justifiable theory such as allometry. Furthermore, the derivation of the Wilks Formula is not published or easily accessible. This leads to questions of updates and validity over time.

Allometric modeling, which has theoretical support, has the disadvantage of assuming that the body mass to strength relationship persists across all lifts and between sexes. In fact, the resulting allometric standard curves in previous research showed deviation from theory, probably due to the fact that the proportion of body mass being lifted is minimal in the bench press, moderate in the deadlift, and more significant in the squat (8). The different body mass exponents for each lift means that the resulting allometric scores are actually in different units and, therefore, cannot be added.

Finally, this simple approach is applicable in evaluating sport performance between sexes, and among classes, for any sport that uses clearly quantified results and the same competitive rules. For example, run or swim times for any given distance can be compared across age and sex through use of z-scores. This would allow comparisons of performance in the total population, and should be extendable to age groups in adolescent sport.
Likewise, this approach can be used for combining scores from different performances. Z-scores could be used to compare between different sports that can similarly be converted to a z-score. For example women’s heptathlon, and men’s decathlon could be scored using this method. Therefore, we recommend coaches and sport officials use this method for evaluating powerlifting and other multi-event sports.

**PRACTICAL APPLICATION**

Coaches and athletes can use the means and standard deviations provided by the Tables in this paper to accurately compare any Powerlifter’s performance, regardless of size and sex, to that of any other lifter. Furthermore these data show coaches and athletes how incremental increases in any lift will affect the overall score relative to all lifters. The z-score approach is simple and the Tables can be easily updated as more data become available and as athletes improve. This approach provides greater accuracy in determining the best lifter for each of the three lifts, and more clearly and fairly determines the best overall lifter without bias among the three lifts. Some strength coaches and athletes may use this method to combine any number of one-repetition maximums for different lifts, as well as other performances to determine a composite score. This approach should be widely adopted and updated as needed in the sport of Powerlifting. This z-score approach could also be applied to other multi-component sports.
ACKNOWLEDGMENTS

No funding was received for the study and no conflicts of interest are declared by the authors.

REFERENCES


### Table 1. Weight lifted in kilograms and z-scores for the squat, bench press, and deadlift of the top 30 women for each weight class.

<table>
<thead>
<tr>
<th>Weight class (kg)</th>
<th>Mean (SD)</th>
<th>Top z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Squat</td>
<td>Bench</td>
</tr>
<tr>
<td>44</td>
<td>84.1 (16.7)</td>
<td>52.5 (10.5)</td>
</tr>
<tr>
<td>48</td>
<td>108.8 (10.9)</td>
<td>66.1 (4.6)*</td>
</tr>
<tr>
<td>52</td>
<td>136.4 (16.9)</td>
<td>79.9 (8.0)</td>
</tr>
<tr>
<td>56</td>
<td>146.7 (12.0)</td>
<td>90.4 (9.0)</td>
</tr>
<tr>
<td>60</td>
<td>163.0 (14.2)</td>
<td>94.8 (7.4)</td>
</tr>
<tr>
<td>67.5</td>
<td>174.2 (12.0)</td>
<td>114.7 (11.7)</td>
</tr>
<tr>
<td>75</td>
<td>194.4 (12.2)</td>
<td>122.0 (9.4)</td>
</tr>
<tr>
<td>82.5</td>
<td>194.7 (16.2)</td>
<td>125.6 (11.4)</td>
</tr>
<tr>
<td>90</td>
<td>189.1 (18.3)</td>
<td>120.4 (12.2)**</td>
</tr>
<tr>
<td>SHW</td>
<td>208.7 (28.5)</td>
<td>128.6 (22.2)</td>
</tr>
</tbody>
</table>

* n=29, ** n=28; SHW = super heavy weight
Table 2. Weight lifted in kilograms and z-scores for the squat, bench press, and deadlift of the top 50 men for each weight class.

<table>
<thead>
<tr>
<th>Weight class (kg)</th>
<th>Squat</th>
<th>Bench</th>
<th>Deadlift</th>
<th>Squat z-score</th>
<th>Bench z-score</th>
<th>Deadlift z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>56</td>
<td>187.8 (25.1)</td>
<td>137.6 (15.3)</td>
<td>214.1 (20.6)</td>
<td>2.73</td>
<td>3.15</td>
<td>3.44</td>
</tr>
<tr>
<td>67.5</td>
<td>222.6 (12.3)</td>
<td>160.7 (9.6)</td>
<td>254.2 (14.9)</td>
<td>2.40</td>
<td>3.30</td>
<td>4.15</td>
</tr>
<tr>
<td>75</td>
<td>253.8 (18.1)</td>
<td>181.0 (11.9)</td>
<td>280.5 (15.3)</td>
<td>3.77</td>
<td>3.84</td>
<td>2.92</td>
</tr>
<tr>
<td>82.5</td>
<td>277.8 (14.9)</td>
<td>199.3 (11.4)</td>
<td>300.6 (12.7)</td>
<td>3.12</td>
<td>4.64</td>
<td>4.60</td>
</tr>
<tr>
<td>90</td>
<td>306.0 (16.3)</td>
<td>213.9 (12.8)</td>
<td>321.3 (15.9)</td>
<td>3.59</td>
<td>3.32</td>
<td>4.29</td>
</tr>
<tr>
<td>100</td>
<td>345.0 (21.1)</td>
<td>230.4 (9.7)</td>
<td>341.7 (17.2)</td>
<td>3.32</td>
<td>3.61</td>
<td>3.90</td>
</tr>
<tr>
<td>110</td>
<td>347.1 (15.5)</td>
<td>242.1 (12.8)</td>
<td>351.1 (14.9)</td>
<td>2.57</td>
<td>4.35</td>
<td>3.44</td>
</tr>
<tr>
<td>125</td>
<td>361.4 (17.7)</td>
<td>260.3 (13.0)</td>
<td>357.1 (13.5)</td>
<td>4.28</td>
<td>3.53</td>
<td>3.15</td>
</tr>
<tr>
<td>140</td>
<td>372.5 (22.0)</td>
<td>261.5 (14.9)</td>
<td>354.0 (15.9)</td>
<td>4.32</td>
<td>3.79</td>
<td>3.04</td>
</tr>
<tr>
<td>SHW</td>
<td>386.9 (27.7)</td>
<td>271.5 (17.0)</td>
<td>364.9 (16.6)</td>
<td>2.64</td>
<td>3.30</td>
<td>2.86</td>
</tr>
</tbody>
</table>

SHW = super heavy weight