Single Versus Multiple Sets for Strength: A Meta-Analysis to Address the Controversy

Matthew R. Rhea, Brent A. Alvar, and Lee N. Burkett

Key words: resistance training, weight training

Weight training can be used as a stressor to overload the neuromuscular system and develop strength. A number of weight training variables can be manipulated to overload this system. The main variables include volume, intensity, and the amount of rest allowed between sets or workouts. Of these variables, volume has received much interest among researchers and professionals, and considerable debate has arisen (Byrd, 1999) regarding an increased strength benefit of multiple-set programs.

The debate over the amount of volume needed to elicit maximal strength gains has continued in recent years with several narrative reviews (Carpinelli & Otto, 1999; Feigenbaum & Pollock, 1999) of research literature comparing single and multiple sets of training. These reviews have determined that single-set training programs elicit similar strength increases or health benefits (especially in untrained individuals) compared to multiple-set programs due to the inability of most of these studies to identify a statistical difference at the .05 level.

The reliance on probability values (p) places considerable importance on statistical power. If statistical power is low, the possibility of committing a Type II error (failing to reject the Null hypothesis despite a true difference existing) is increased. With this in mind, a power analysis (Cohen, 1988) was performed on a random sample of 10 studies from the literature comparing single and multiple sets. The mean power was calculated to be 0.36. Cohen (1988) reported that a power below 0.80 would incur too great a risk of a Type II error. Statistical significance is also heavily affected by large variance and small sample sizes. This can result in large differences between groups being deemed nonsignificant or in small differences reaching statistical significance solely based on sample size (and statistical power).

The body of research comparing single and multiple sets for maximal strength gains contains many studies performed with small sample sizes and low power; therefore, reliance solely on p values may be misleading. Completing a meta-analysis could be a valuable asset in this situation. This procedure, popularized by Glass (1982), combines the results of independent studies, sums sample sizes across studies, and increases statistical power. It involves calculating an effect size (ES), which represents the magnitude, in standard deviation units, of a treatment. The ES is an objective measure of meaningfulness, which when combined with statistical significance, provides a more accurate picture of treatment effects.

A meta-analysis has two major advantages over the narrative literature reviews previously performed in this body of research (Thomas & French, 1986). Procedures for decision making are provided and reported in a meta-analytical procedure, whereas narrative reviews rely solely on the reviewer’s jurisdiction. It also provides a quantitative method for analyzing research findings through calculating and comparing the ES. The purpose of this research was to systematically examine studies comparing single- and triple-set training programs for strength by calculating and examining ESs.

Submitted: October 29, 2001
Accepted: February 19, 2002

Matthew R. Rhea, Brent A. Alvar, and Lee N. Burkett are with the Department of Exercise and Wellness at Arizona State University.
Method

Literature Search

Searches were performed for all English language studies from 1962-2000 comparing single- and triple-set training programs. Computer searches of Science Citation Index, National Library of Medicine, Sport Discus, ERIC, and Medline were performed. Hand searches of relevant journals and reference lists obtained from articles were conducted.

Inclusion Criteria

Studies were included with both single- and triple-set training groups and compared identical strength measures among the groups. In the event that necessary statistical or descriptive information was deficient, solicitations were made from the primary investigators. A total of 16 studies were found (Berger, 1962; Capen, 1956; de Hoyos et al., 1998; de Hoyos, Herring, Garzarella, Werber, & Brechue, 1997; Haas, Garzarella, de Hoyos, & Pollock, 1998, 2000; Jacobson, 1986; Kraemer, 1997; Kraemer et al., 1997; Leighton, Holmes, Benson, Wooten, & Schmerer, 1967; Messier & Dill, 1985; Reid, Yeater, & Ulrich, 1987; Rhea, Alvar, Ball, & Burkett, 2002; Sanborn et al., 2000; Silvester, Suggings, McGown, & Bryce, 1982; Starkey et al., 1996).

Coding of Studies

The primary investigator read and coded each study to identify descriptive information and program design. Once the studies were coded, five ES were required for each level or variable to be examined statistically. Program methodology was examined. To be classified as a controlled study, both groups must have trained at equal intensity with the same variation in programs (i.e., if the multiple-set programs were periodized, the single-set program must be periodized in the same manner). The participants' status was divided into trained and untrained classifications. Participants must have been part of a weight-training program for at least 1 year prior to the study to be considered as trained. The length of the training programs were divided into categories of 1–5, 6–10, 11–15 weeks, and so on. Coder drift was assessed (Orwin, 1994) by randomly selecting 10 studies for recoding. Per case agreement was determined by dividing the variables coded the same by the total number of variables. A mean agreement of 0.90 was required for acceptance.

Statistical Data

Effect sizes were calculated using means and pooled standard deviations and weighted by the inverse of the variance (Hedges & Olkin, 1985), with single set training groups set as the control group. Tests for homogeneity were also conducted (Hedges & Olkin, 1985). Analysis of variance with confidence intervals identified differences among variables. Rosenthal's (1979) formula was used to calculate the tolerance of the overall ES to negative experimental results. This formula calculates the estimated number of studies demonstrating null results that would need to exist (that were not included in the analysis) to bring the overall ES to nonsignificance.

Results

A total of 16 relevant studies were coded, yielding 93 ESs (see Table 1). Coder drift was assessed to be 0.96. Thus, the coding process used in this analysis was found to be consistent. The overall ES of all studies included in the analysis was calculated to be 0.28 (SD = 0.36). The mean ES of those studies demonstrating equated intensity and controlled variation was calculated to be 0.70 (SD = 0.92). In those studies failing to maintain this control, the ES was calculated to be 0.25 (SD = 0.70). This difference approached significance, F(1, 55) = 2.55, p = .12. The test for homogeneity was significant, Q (92) = 222.9, p < .001, warranting examination of potential moderating variables or variables that resulted in ESs of differing magnitude.

Training status was found to be a significant variable. Participants who were classified as trained exhibited an ES (0.55; SD = 0.59) that was significantly greater than the ES of untrained participants, (0.23; SD = 0.54), F(1, 55) = 4.03, p < .05. Only three lengths of training categories produced enough ESs to be analyzed (see Table 1). Training programs of 11–15 weeks and 21–25 weeks resulted in higher ESs (0.39 and 0.33, respectively) than 6–10-week programs (0.17), but this difference was statistically nonsignificant (p = .26), possibly due to the lack of available ESs for longer training programs. More research is needed using training programs longer than 15 weeks. The tolerance of these results to null experimental results was calculated to be 571. Thus, 571 ESs demonstrating the null hypothesis would have to exist to bring the current ES to nonsignificance.

Discussion

The results of this meta-analysis demonstrate that three sets of training increase strength to a greater degree than single-set training (ES = 0.28). Studies that maintained stringent methodological control of train-
ing intensity and variation exhibited a larger difference between training programs (ES = 0.70). The differences in ESs between those studies that controlled training intensity and program variation demonstrate an important aspect of research in strength improvements. To investigate the impact of volume on strength development it is important to hold all other variables constant. Because of the greater methodological control, the ES of 0.70 is most likely a more accurate measure of the greater strength gains that are achieved with multiple set training.

Another finding of this analysis is that the difference in strength gains is more pronounced in trained individuals. While there is a significant difference between the gains for untrained individuals favoring multiset programs, this difference increases as the individual becomes more accustomed to strength training. Theoretically, as the individual becomes more accustomed to a weight training program, more volume is needed to maximally overload the neuromuscular system (Fleck & Kraemer, 1997).

As previously mentioned, a power analysis identified low statistical power among the individual studies included in this analysis. This lack of power appears to have resulted in the inability to detect the differences between single and multiset programs for strength despite its existence (Type II Error), as only a few studies (Berger, 1962; Kraemer, 1997; Kramer, Stone, O’Bryant et al., 1997; Kramer, Stone, O’Bryant, Conley et al., 1997) reached the .05 levels of significance. By combining the results of these studies, power was increased and the ES identified a significant difference between single- and multiple-set training protocols for strength gains. The magnitude of the ES is influenced by participants’ training status, length of training program, and methodological control of the study.

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>ES</th>
<th>SD</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>93</td>
<td>.28</td>
<td>.56</td>
<td>.22</td>
<td>.32</td>
</tr>
<tr>
<td>Controlled</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>methodology</td>
<td>53</td>
<td>.70</td>
<td>.92</td>
<td>.31</td>
<td>1.10</td>
</tr>
<tr>
<td>Uncontrolled</td>
<td>40</td>
<td>.25</td>
<td>.70</td>
<td>-.16</td>
<td>.65</td>
</tr>
<tr>
<td>Training status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trained participants</td>
<td>38</td>
<td>.46</td>
<td>.50</td>
<td>.40</td>
<td>.51</td>
</tr>
<tr>
<td>Untrained</td>
<td>19</td>
<td>.19</td>
<td>.51</td>
<td>.11</td>
<td>.28</td>
</tr>
<tr>
<td>Length of training (weeks)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6–10</td>
<td>37</td>
<td>.17</td>
<td>.69</td>
<td>-.06</td>
<td>.40</td>
</tr>
<tr>
<td>11–15</td>
<td>48</td>
<td>.39</td>
<td>.56</td>
<td>.23</td>
<td>.56</td>
</tr>
<tr>
<td>21–25</td>
<td>8</td>
<td>.33</td>
<td>.79</td>
<td>-.32</td>
<td>.69</td>
</tr>
</tbody>
</table>

Note. n = number of effect sizes; ES = effect size; SD = standard deviation.

References

References marked with an asterisk indicate studies included in the meta-analysis.


RQES: December 2002


**Authors’ Note**

Please address all correspondence concerning this article to Matthew Rhea, Department of Exercise and Wellness, Arizona State University, East Campus, 6113 S. Kent, CLRB, Mesa, AZ 85212.

E-mail: matthew.rhea@asu.edu