JEPonline
Journal of Exercise Physiology**online**

Official Journal of The American Society of Exercise Physiologists (ASEP)

ISSN 1097-9751
An International Electronic Journal
Volume 7 Number 5 2004

Review: Resistance Training

Meta-Analyses Do Not Support Performance of Multiple Sets or High Volume Resistance Training

Richard A. Winett

Center for Research in Health Behavior, Department of Psychology, Virginia Tech, Blacksburg, VA, 24061-0436, USA

ABSTRACT

META-ANALYSES DO NOT SUPPORT PERFORMANCE OF MULTIPLE SETS OR HIGH VOLUME RESISTANCE TRAINING. **Richard A. Winett JEP**online. 2004;7(5):10-20. Four recently published meta-analyses claim their results show that multiple-set resistance training protocols (higher volume) are superior to a single set of each exercise (lower volume) for producing strength gains in experienced trainees. This critique examines the framework, logic, procedures, statistics, results and interpretations of the four meta-analyses and shows that these studies did not follow many of the recognized guidelines for meta-analysis. There was very little support for any of the purported claims or conclusions. In fact, this critique suggests that simple, time efficient, single-set, lower volume protocols appear to be just as effective as multiple-set, higher volume protocols for increasing muscular strength regardless of goals or training status.

Key Words: Weight lifting; Volume; Intensity; Adaptation

INTRODUCTION

A long-term debate in resistance training is whether performing multiple sets of each exercise or a greater volume of exercise per muscle group results in greater strength gains than compared to a single set of each exercise. There is minimal evidence to support the greater efficacy of multiple sets (1), and a recent reexamination of a 40-year old study, which was perhaps the genesis of the belief in multiple sets for resistance training, revealed very little supporting evidence for the use of multiple sets (2). Nevertheless, the American College of Sports Medicine's (ACSM) Position Stand for resistance training recommends multiple sets for experienced trainees and competitive athletes (3).

Inconsistencies in results and data interpretations between different research studies can often arise due to problems with statistical power in some or all of the research at question. In an effort to overcome the statistical power limitations of prior research, researchers and academics often revert to using a meta-analysis. A meta-analysis is a quantitative way to integrate the results of empirical studies in one field, such as from resistance training studies. The purpose of a meta-analysis is to compare the outcomes produced

across studies by different treatments or protocols such as strength outcomes from training with single or multiple sets. In order to be included in a meta-analysis for a given field, each study needs to have a common metric. For example, in a meta-analysis for resistance training, each study would need to have a measure of strength as an outcome of the resistance training protocol.

Each of these included studies would contribute an effect size to the meta-analysis. An effect size is a standard, statistical measure to represent the degree or amount of 'outcomes', 'impacts', or 'effects', and is quantified as a mean difference divided by the standard deviation. An effect size of 0.00 to 0.32 is considered 'small', an effect size of 0.33 to 0.55 is considered 'medium', and an effect size of 0.56 to 1.20 is considered 'large' (4). There is relevance for such categorization, and the larger the effect size the better.

Effect sizes can be calculated from each study on muscular strength produced as a result of a treatment or protocol such as training with multiple sets or training with single sets. In a meta-analysis, for example, an overall mean for all the effect sizes from studies with multiple set protocols can be compared to an overall mean for all the effects sizes for single set protocols to determine if there is a statistically significant difference between them. However, meta-analyses are only viable if they follow established methodological, statistical, and interpretive guidelines. The conclusions drawn from a meta-analysis also are only as good as the criteria used to select the prior research used in the analysis.

Four recent meta-analyses (5-8) claim their results provide convincing evidence that when the data from strength training studies are properly aggregated, the effect size for strength outcomes are considerably greater with multiple sets, or a greater volume of exercise per muscle group, compared with a single set of each exercise. It is hypothesized that these four meta-analyses are flawed and that none of these meta-analyses, logically, theoretically, or statistically provides any real evidence for the superiority of training with multiple sets per exercise or a greater volume of exercise per muscle group. It also is hypothesized that none provides clear evidence of the importance of any specific parameter of training or protocol because the four meta-analyses (5-8) generally failed to follow a number of key guidelines for meta-analyses noted by Lipsey and Wilson (9) including:

1. separating studies with differentially rigorous research designs (i.e., randomized control group designs versus nonrandomized design and other studies),
2. testing for homogeneity to determine if there is more variation in a group of categorized research studies (e.g., four sets versus two sets per muscle group) than variation expected from sampling error (indicating that the studies should not be so grouped and categorized),
3. using one mean effect size from a study and not multiple, non-independent effect sizes,
4. identifying outlier studies and effect sizes and either moderating or excluding outliers,
5. acknowledging large variation of effect sizes within a group and not just focusing on mean effect size values,
6. reporting results using confidence intervals to express a probable range of outcomes in a population, and
7. having theory guide study hypotheses with results conforming to theory or biological plausibility.

Meta-analyses have become widely used in many fields and researchers have pointed to related issues and limitations of meta-analyses. These include the need to critically examine each study in a meta-analysis to ensure the quality of each study and the accuracy of its conclusions (10); assessing and reporting the sources of bias in the studies that are included in a meta-analysis and reporting systematic bias that may exist across studies (11,12); understanding and reporting on variation of treatment effects (13-15); including moderators and mediators of treatment effects (16); exercising caution when reporting meta-analyses done with few studies and effect sizes (17) or where studies in the meta-analysis have small differences (15), and not reducing the complexity of a field and treatment effectiveness to the 'Holy Grail' of one mean effect size (12, 18).

The purpose of this document is to review each of these meta-analyses in detail while providing an ongoing critique of each study.

META-ANALYSES

Rhea et al. (5)

Rhea and colleagues (5) performed a meta-analysis that they claimed included all published and unpublished studies that were strength training interventions, though with different experimental designs, and where there was a pre- and post-training strength measurement. An examination of their extended reference list indicated that at least 26 studies meeting these criteria were not included in their analysis. These studies are cited in a recent critique of the ACSM's Position Stand (19); 24 of these 26 excluded studies showed no significant difference between single and multiple sets per exercise (19; see p. 17 Table 4). The exclusion of specific studies can create a bias in the outcomes of the meta-analysis, particularly if the reported results are consistently at odds with the conclusions of the meta-analysis.

For each study in the meta-analysis, Rhea et al. (5) reported effect sizes (*ES*) for so-called intensity (% 1 RM), frequency (number of days per week participants trained a particular muscle group), and volume (number of sets per muscle group during each session) in previously untrained and experienced (>1 year) subjects. Rhea et al. (5) reported statistically significant differences in *ES* between previously untrained and experienced trainees but it was not clear if those statistically significant differences were for intensity, frequency, or volume of training. The use of sets per muscle group compared to sets per exercise is a departure from how these studies have been analyzed and debated. However, the article contained no information about how exercises were classified by muscle groups and only one person coded sets per muscle group with, hence, no available reliability statistics. Their coding is not straightforward. For example, it is not specified whether the squat exercise is categorized as affecting the major muscle groups involved in performing the squat (gluteus, quadriceps, hamstrings, gastrocnemius, and spinal erectors), or just the quadriceps muscle.

For each study, and where possible for each variable, an effect size was generated based on the difference between a post-training and pre-training strength score divided by the pre-training strength score's standard deviation, thereby indicating the effectiveness of a given variable or treatment. Hundreds of effect sizes were generated. The result is that some studies inappropriately contributed many effect sizes while other studies contributed few effect sizes, which creates potential bias (9).

Because studies had different numbers of participants, Rhea et al. (5) used a weighting factor for the number of participants. In order for a variable to be considered, there needed to be at least 10 effect sizes from the many studies included in the meta-analysis. It also was possible to look at effects for previously untrained (<1 year training) and experienced (>1year training) participants. Rhea et al. (5) calculated an overall mean effect size for each variable for trained and untrained participants across studies.

A strength of the methodology by Rhea et al. (5) is including published (though with some published studies excluded; see above) and unpublished studies to limit the 'file draw effect' (9). File draw effects mean that studies without statistically significant outcomes are relegated to the file draw and not published. If those studies are not included, the result is that the overall effect sizes are inflated. A weakness of their methodology is not giving greater weight to studies that used sounder methodology such as randomization to conditions. Another problem is that their data presentations did not independently examine effects of one variable, e.g., sets. The sets/muscle group variable for any study from a statistical standpoint is confounded with intensity and frequency. Either joint effects should have been examined or other variables (intensity, frequency) held constant to examine one variable (sets). Coding should have been performed for percentage of 1 RM to allow for use of this variable as a moderator in the analysis.

As expected, the effect sizes for previously untrained participants were larger than those for experienced participants simply because strength increases are larger and more rapidly produced in novice trainees. Overall, the meta-analysis as reported by Rhea et al. (5), presented primarily in tables and figures with limited analyses, suggests that for experienced trainees the most effective parameters for producing strength gains are training at 80% of 1 RM (about 8 repetitions using a conventional repetition duration; see later),

training muscle groups twice/week (three times/week was less effective), and performing four sets/muscle group. According to Rhea et al. (5), previously untrained people apparently produced better results training at 60% 1 RM, three times/week, and for four sets/muscle group.

The forgoing conclusions were simply based on the pattern of results and 'trend plots' (graphic representation of the data), with apparently no other main analyses reported. Such trend plots can lead to unsupportable conclusions. For example, while the effect size for experienced people using four sets/muscle group (1.17 ± 0.81) was greater than for using two sets/muscle group (0.92 ± 0.52), I performed a simple *t*-test using the study's data and found that this difference was not significant ($t<1$). Thus, the actual data from the meta-analysis indicated that there is no significant difference between training with two or four sets per muscle group.

A close inspection of the data presented in the meta-analysis reveals some additional points. In Table 1 (p. 458) there are data on training at different percentages of 1 RM and effect sizes for strength increases that need explanation. The effect sizes for previously trained study participants were 0.70 ± 0.65 for 70% 1 RM, 0.74 ± 0.99 for 75% 1 RM, 1.80 ± 1.30 for 80% 1 RM and 0.65 ± 0.77 for 85% 1 RM. The relatively large SD's at a given percentage of 1 RM indicate that the studies and their individual outcomes within that percentage of 1 RM differ from each other in some meaningful way. This suggests some artifact or important moderator variable that needed to be noted or analyzed by Rhea et al. (5), but was overlooked.

There also is no current theory or identifiable physiological mechanism that explains why 75% and 85%1 RM are effective stimuli, but 80% is extremely effective. What this presumably means is that if for an exercise a 1 RM is 100 kg then performing a specific number of repetitions to fatigue with 75 or 85 kg for a specific exercise is effective but performing repetitions to fatigue with 80 kg is much more effective. There is no known physiological mechanism that explains why that would that be the case. Rhea et al. (5) failed to address this issue.

Research shows there is a considerable range of repetitions that can be performed at a given RM by different individuals and with different types of resistance training movements (19). For example, in a recent training study, Chromiak and colleagues (20) found that when using 85 % of 1 RM participants were able to perform a mean of 4.5 ± 1.40 and 4.7 ± 1.70 repetitions in the bench press but 8.8 ± 4.50 and 10.8 ± 6.1 repetitions in the leg press at pre-training and post-training respectively. The different number of repetitions and the different standard deviations in different exercises at the same percentage of 1 RM call into question, as noted elsewhere (19), training models such as prescribed in the Position Stand (3) that use a specific percentage of 1 RM for a specific number of repetitions. If the data for 80% of 1 RM in the Rhea et al. (5) meta-analysis do represent some artifact, then training at *any* RM from 70% to 85% 1 RM is about equally effective. The pattern of results reported by Rhea et al. (5) makes no theoretical sense and lacks physiological plausibility.

Rhea et al. (5) did point out that the results of their meta-analysis only suggest the parameters of the most optimal protocol for strength development. Based on training status and personal goals plus other physical activities, a person could choose to perform a lower volume of training. For example, the number of sets/muscle group or training frequency could be reduced if a person has reached a point of diminishing returns as far as additional strength or muscle mass. However, there are no data presented by Rhea et al. (5) that support any particular training protocol.

In summary, the results from Rhea et al. (5) are questionable because they excluded certain published studies, used many non-independent effect sizes, focused on mean effect sizes with less attention to the large variation within a given training category, and could not explain the pattern of outcomes with any theory or physiological mechanism.

Wolfe et al. (6)

Wolfe and colleagues (6) claimed their meta-analysis showed the superiority of using multiple sets per exercise to increase strength. They performed a meta-analysis of published studies that researched the impact of single vs. multiple sets training, met at least minimal methodological criteria, and from which effect sizes could be extracted. Wolfe et al. (6) controlled for some important variables within the selected study sample such as publication bias, quality of the research design, number of people in each study, the number of effect sizes generated, and outlier status. After examination by Wolfe et al. (6), however, it was noted in their narrative that no studies so identified were excluded as outliers.

Wolfe et al. (6) approached the problem of multiple non-independent effect sizes by first including all effect sizes from one study and then also including a mean effect size from a study with multiple effect sizes. It is not clear how this approach by Wolfe et al. (6) corrected for the problem.

Surprisingly, after four decades of research in this area, Wolfe et al. (6) noted there were only 16 studies that met acceptable criteria to be included in their analyses. However, Wolfe et al. in Table 1 (pp 38-42) indicated 20 studies they included in their meta-analysis. *Only six of those studies included subjects with previous training experience.* One problem is simply the very limited number of studies from which to draw conclusions. Using a small number of studies in a meta-analysis can result in 'second order sampling error'; this is sampling error at the meta-analysis level (9).

Wolfe et al. (6) indicated that the *ES* for trained subjects was significantly greater as a result of training with multiple sets (0.70) compared with a single set (0.29), but not significantly different in untrained subjects (*ES* = 1.73 and 1.69, multiple sets and 1-set, respectively). There are though a number of other problems with the meta-analysis. Wolfe et al. (5) did not define trained. Wolfe et al. (6) noted in Table 3 that a study by Kraemer et al. (22) involved trained subjects. However, as indicated in a prior critique (19), the subjects were previously untrained collegiate female tennis players. Wolfe et al. (6) incorrectly cited the Kraemer et al. (22) publication as *Medicine and Science in Sports and Exercise*, when in fact it was published in the *American Journal of Sports Medicine.* Wolfe et al. (6) also claim that with only one exception, every study included in their analysis involving trained subjects reported that multiple sets were superior (p. 44). In a later section, the effect sizes from the studies noted by Wolfe et al. (6) are delineated and this claim is shown to be incorrect. What may arguably be small errors nevertheless question the accuracy of investigators who are attempting a much more sophisticated challenge such as a meta-analysis yielding 103 effect sizes.

Wolfe et al.'s (6) results included six studies that reported findings for experienced trainees. As noted above, the results showed that the effect size for the single set protocols was 0.29. For the multiple sets protocols, the effect size was 0.70 and there was a significant difference between these effect sizes. These data apparently show that multiple sets per exercise can be far more effective for experienced trainees. However, the studies with experienced trainees included in this meta-analysis warrant closer examination (6).

A series of experiments by Kraemer (21) contributed 10 effect sizes because data were drawn from three experiments (#2-4) in just one article. In 1997 Kraemer reported on experiments that he performed about 15 years earlier as a football coach. In experiment #2, Kraemer (21) reported that players training with single sets showed small gains expressed either as effect sizes or percent change for bench press (0.06; 4%) and leg press (0.13; 4%). Those players training with multiple sets showed greater gains in the bench press (0.13; 13%) and leg press (0.77, 19%). For experiment #3, the strength gains for the bench press for single sets (0.07; 3%) were considerably smaller than for multiple sets (0.24; 11%); similar outcomes were reported for the hang clean. For experiment #4, the strength increase in leg press (the only exercise reported with a 1RM) for single sets (.24; 6%) also was considerably smaller than for multiple sets (0.64; 18%). Such large increases in strength using multiple sets for this group of high-level strength athletes are unusual, if not questionable, and the results should be treated as outliers (9).

One other study showed differences between single set and multiple set training (22). However, as noted above, the training status of the participants at the start of the study was incorrectly designated by Wolfe et

al. as experienced trainees. An examination of the study's protocol indicated that the female participants needed a period to become familiar with resistance training.

Two other studies (25, 26) with experienced trainees in the analysis are of interest. In a 13-week study where the participants trained three times/week, Hass and colleagues (25) compared one set and three sets on a training circuit of nine strength exercises. Data were presented on outcomes for five exercises. I performed a calculation of effect sizes for the two conditions in the study that showed about the same outcomes except for leg curls (0.33 for single sets and 1.0 for multiple sets). Strength gains for the experienced trainees were approximately 10% regardless of the protocol used (one or three sets). There was a significant increase across all exercises as a result of performing one set or three sets of each exercise, with no reported significant difference in strength outcomes between one set and three set groups.

Ostrowski and colleagues (26) compared protocols involving one, two, or four sets per exercise. Each exercise was performed once/week and the participants trained four times/week using a split routine. Participants performed three exercises/muscle group so that some people were performing as few as three sets/muscle group/week (1-set group) while others were performing up to 12 sets/muscle group/week (4-set group). At the end of the study, there were no significant differences in outcomes among the groups. The strength gains were about 6%, more or less what can be expected with experienced trainees, for the bench press and squat exercises. Ostrowski et al. (26) concluded that reaching a *minimum* of volume and frequency of training is all that is required to produce positive outcomes.

Interestingly, the outcomes from Hass et al. (25) and Ostrowski et al. (26) were similar even with different exercise modalities, volume, and frequency of training. Despite what appear to be large differences in protocols, they all commonly provided an adequate stimulus and then recovery time for adaptation. The one set/exercise routines used by Hass et al. (25) and Ostrowski et al. (26) demonstrate that experienced trainees can benefit from simple, time efficient resistance training routines.

The difficulty in drawing definitive conclusions about single versus multiple sets with experienced trainees, particularly given the small number of requisite studies, is further illustrated by the following analyses I performed using the data presented in Wolfe et al. (6). Effect sizes from each of four studies (23-26) represented the mean effect size across different exercises performed for either single or multiple sets in experienced trainees. Each study provided just two effect sizes, one for single and one for multiple sets as is appropriate (9), and not the multiple non-independent effect sizes from each study of the six studies (21-26) as in the meta-analyses by Rhea et al. (5) and Wolfe et al (6). From the Ostrowski et al. study (26), I used the mean effect size for one and four sets because the mean effect size for four and two sets was about the same. I represented the data using confidence intervals. Given the data from the studies, confidence intervals indicate the range within which the population mean is likely to lie. I used a 95% confidence interval indicating that there is a 95% probability that the mean is within the ranges noted. The mean effect size and confidence range from these four studies (23-26) were for single sets, 0.44 ± 0.22 ($p = 0.05$) and for multiple sets, 0.64 ± 0.25 ($p = 0.05$). Expressed in confidence ranges ($p = 0.05$), these four studies suggest an effect size range of 0.22 to 0.66 for single sets and 0.39 to 0.89 for multiple sets. These are overlapping distributions. When the mean effect size for the single set (0.14) and multiple sets outcomes (0.45) from Kraemer's (21) three experiments were added into the pool, the overall single set effect size was then 0.38 ± 0.27 ($p = 0.05$) with a range of 0.11 to 0.65, and the multiple set effect size was 0.60 ± 0.26 ($p = 0.05$), with a range of 0.34 to 0.86, which again suggests overlapping distributions. When the mean single set effect size (0.28) and the multiple set effect size (1.27) from the study (22) where the training status of the participants was incorrectly noted as experienced were added, the overall mean effect size for single sets was 0.36 ± 0.26 ($p = 0.05$) with a range of 0.10 to 0.62, and for multiple sets 0.71 ± 0.46 ($p = 0.05$), with a range of 0.25 to 1.17, and remain overlapping distributions. Which studies are included, whether a mean effect size from each study or multiple non-independent effect sizes are drawn from each study, and how analyses are conducted impact conclusions that are drawn. The aforementioned confidence intervals (9) do not indicate the superiority of multiple sets.

The results from Wolfe et al. (6) are questionable because only a small number of studies involved experienced trainees (21-26), Wolfe et al. did not critically examine each study, and multiple non-independent effect sizes were used.

Rhea et al. (7)

With very little reported data, Rhea et al. (7) claimed that their meta-analysis demonstrated that 3-set training was superior to 1-set training in previously untrained and previously trained participants. They did not define how they coded trained and untrained participants. Standard deviations were large (i.e., overall ES = 0.28, SD = 0.56), indicating a large variation of effects from training with either one or three sets. Rhea et al. (7) used 16 studies in their analysis; 12 of them are cited in a prior critique (19), and nine out of those 12 studies reported no significant difference in strength gains as a result of single or multiple sets (Table 4, p. 17). One of the other three studies is the previously discussed outlier by Kraemer (21). Of the four other references used by Rhea et al. (7), one (27) was published after the Position Stand (3) and reported that the strength gains in the leg press exercise were significantly greater in a 3-set group compared with a 1-set group, with no significant differences in bench press strength, lean body mass, percent body fat, or chest and thigh circumference. The other three references (28-30) were abstracts that all reported no significant differences in strength gains as a result of training with one set or three sets. Therefore, 12 of the 16 studies in the meta-analysis (7) reported no significant difference between single and multiple sets. These points alone raise questions about this meta-analysis that claims training with three sets is superior to training with one set per exercise.

Rhea et al. (7) generated an effect size estimate by considering the one set group as the ‘control group’. The outcomes of this meta-analysis, as noted, suggested the superiority of using three sets/exercise compared to one, with a mean effect size of 0.28 ± 0.56. Rhea et al. (7) also compared mean effect sizes that had equated intensity and controlled variation (0.70 ± 0.92) to those that did not equate intensity and variation (0.2 ± 0.70) and found that this difference approached significance ($p = 0.12$). The large standard deviations and a significant test for homogeneity indicate other variables impacted outcomes and the likely presence of outliers. There is some question about interpreting mean effect sizes in the presence of a significant test for homogeneity because this means that there is a great deal of variability associated with the effects of a treatment (9).

One of the major problems noted in the review of the prior Rhea et al. (5) meta-analysis pertains here as well. A total of 93 effect sizes were generated from only 16 studies in this second meta-analysis by Rhea et al. (7). Given these problems in their meta-analysis, Rhea et al. (7) actually have no clear support for multiple sets.

Peterson et al. (8)

Peterson et al. (8) reported 37 training studies that had pre- and post-training strength measures in competitive athletes. The studies were apparently a subset of studies from the Rhea et al. (5) meta-analysis and the methodological problems noted above for Rhea et al. (5) pertain to Peterson et al. (8) as well. Effect sizes were generated for every strength outcome, using the formula post-training mean minus pre-training mean divided by the pre-training standard deviation. Studies could have multiple effect sizes because they had multiple outcomes such as strength gains for squats, bench press, and pull-down exercises. As noted above, such use of many non-independent effect sizes is not appropriate (9).

Peterson et al. (8) categorized the training protocol in each study for number of sets per muscle group, training at a specific percentage of 1 RM, and training muscle groups two or three days/week. For each of these variables (sets, percentage of 1 RM, frequency of training), there were mean effect sizes from the different studies. For example, using four sets/muscle group, the mean effect size was 0.90 (±1.32; indicating a very good impact but with a large standard deviation and, thus, much variation of outcomes) derived from 119 effect sizes from the different studies. The mean effect size for five sets was 0.64 ± 0.73 from 37 effect sizes, and for six sets, 0.68 ± 0.74 from 26 effect sizes. The mean effect size for 8 sets was

1.22 ± 0.56 but that was based on only six effect sizes, possibly from the same (one) study. Effect sizes for 12 (0.69), 14 (1.06), and 16 (0.41) sets were smaller but with no consistent pattern.

Peterson et al. (8) claimed that their meta-analysis showed a continuum of strength increases elicited by a continuum of increased training volume, but the actual data from their meta-analysis do not support their claim. Peterson et al. (8) also claimed that their meta-analysis showed that competitive athletes should use eight sets/muscle group. An inconsistent pattern of results and the small number of effect sizes for eight sets also makes this conclusion questionable.

A somewhat similar, inexplicable pattern emerged in the analysis by Peterson et al. (8) for training at a specific percentage of 1 RM. The effect size for training at 70 % 1 RM was 0.07 ± 0.06, indicating there is *no* impact on strength, whereas for 75 % 1 RM, the mean effect size was 0.73 ± 0.87, for 80 % 1 RM, 0.57 ± 0.69), and for 85% 1 RM, 1.12 ± 1.35), which is a large effect size but with a very large SD. The differences in both effect sizes and their standard deviations in the training range of 70-85 % 1 RM are puzzling. There is no theory or physiological mechanism explaining why a 5 % change in resistance and doing a few less or a few more repetitions can result in such large differences in outcomes. This result also is subject to some questioning because, as previously noted, people can vary a great deal by how many repetitions they perform to exhaustion for a specific exercise with a given RM (20).

Peterson et al. (8) also indicated that the overall outcomes were moderated by use of creatine, periodized training, and training to failure. These variables were reported in the narrative as eliciting statistically significant greater strength gains but no specific data were presented.

Another limitation was how the data were analyzed. Peterson et al. (8) presented the graphic outcomes of the data with no other formal analyses. Unless separate analyses were done between, for example, every number of sets, it is not known where, if any, there was a significant difference. In the absence of any obvious pattern, no clear conclusion can be reached. Furthermore, given that there were only six effect sizes contributing to the mean for eight sets, with those effect sizes possibly from one study, any conclusion about the impact of eight sets compared to any number of other sets is highly suspect. In addition, the mean for four sets was compiled from 119 effect sizes. Therefore, the two mean effect sizes for eight versus four sets cannot be legitimately compared.

For the sake of clarity, however, using the data from Peterson et al. (8), I statistically compared the difference between four and eight sets/muscle. The difference is not significant ($t<1$).

The results from Peterson et al. (8) are highly questionable because of their use of multiple non-independent effect sizes, the few effect sizes available for the category claimed to produce superior results (8 sets), their focus on mean effect sizes and not the large variation, and the lack of any theory or a physiological mechanism explaining their pattern of outcomes.

CONCLUSIONS

Four recent meta-analyses claim that their results show the superiority of using multiple sets or using four or more sets per muscle group compared with a single set per exercise in experienced trainees or competitive athletes. The pattern of results, the statistics presented in the studies, the statistics I performed for three meta-analyses (Rhea et al; 5; Wolfe et al; 6; Peterson et al; 8), an examination of the studies in one meta-analysis (Wolfe et al; 6), and an absence of any explanatory theory or physiological mechanism, *all fail* to support high-volume resistance training protocols or recommendations.

Table 1 shows the recommended guidelines for meta-analyses and whether each meta-analysis followed or did not follow a specific guideline. It is readily apparent that most of the basic guidelines (9) and cautions and caveats (10-18) for performing meta-analyses and presenting their data were not followed in the aforementioned meta-analyses (5-8).

Table 1. Adherence to Specific Meta-Analysis Guidelines

Guidelines	Reference			
	Rhea et al. (5)	*Wolfe et al. (6)*	*Rhea et al. (7)*	*Peterson et al. (8)*
Categorize Studies By Quality	No	Yes	No	No
Examine Each Study's Results and Conclusions	No	No	No	No
Test For Homogeneity	No	Yes	Yes*	No
Include All Relevant Studies	No	Yes	Yes	?
Control For Publication Bias	Yes	Yes	?	?
Use One Mean Effect Size Per Study	No	No	No	No
Precise Coding of Effects of Exercise or Trainee Status	No	No	No	No
Identify and Exclude Outliers	No	No	No	No
Acknowledge And Report Large Variation	No	No	No	No
Use Formal Statistics	No	Yes	Yes	No
Use Confidence Intervals to Report Results	No	No	No	No
Use Moderators in Addition to Gender, Age, Trainee Status	No	Yes	Yes	Yes
Caution in Conclusions When Using Few Studies	No	No	No	No
Use Theory or Physiological Mechanisms To Explain Results	No	No	No	No

? = Unclear.; *Significance test for homogeneity completed but no results presented.

The major questions are 'how' and 'why' these meta-analyses (5-8) were published with such obvious flaws. One answer to the 'how' question is that across disciplines, researchers have become enamored with meta-analyses that seem to be able to reduce the complexity of a field to a single number (9,18). This critique of these four meta-analyses shows that each meta-analysis by itself can be complex and when guidelines (9-18) for meta-analyses are not followed, the results can be incorrect and misleading.

One answer to the question of 'why' these studies were published is that their unchallenged claims support one facet of the current paradigm delineated in the ACSM Position Stand (3), which is that a greater volume of exercise elicits a greater increase in strength. However, the actual data and some of the studies within these meta-analyses indicate otherwise. Contrary to the prevailing, yet unsubstantiated belief about higher volume training (3), lower volume, time efficient protocols, such as a single set of nine exercises three times/week (25), appear effective regardless of training status and goals.

ACKNOWLEDGEMENTS

I appreciate the feedback and help with drafts of this article from John Donovan, Ph.D., Kevin Davy, Ph.D., Jack Finney, Ph.D., and Mike Houston, Ph.D. The writing of this paper was supported by an ASPIRES grant from Virginia Tech.

Address for correspondence: R.A. Winett, Ph.D., CRHB, Department of Psychology, Virginia Tech, Blacksburg, VA, USA, 24061-0436. (540-231-8747); Email: rswinett@vt.edu

REFERENCES

1. Carpinelli RN, Otto RM. Strength training: Single versus multiple sets. ***Sports Med*** 1998; 26: 73-84.
2. Carpinelli RN. Berger in retrospect: effect of varied weight training programs on strength. ***Brit J Sports Med*** 2002; 36: 319-24.
3. American College of Sports Medicine. Position Stand: Progression models in resistance training for healthy adults. ***Med Sci Sports Exerc*** 2002; 34: 364-80.
4. Lipsey MW. ***Design Sensitivity*** Newbury Park, CA; Sage. 1990.
5. Rhea M, Alvar B, Burkett L, Ball S. A meta-analysis to determine the dose-response relationship for strength. ***Med Sci Sports Exerc*** 2003; 35: 456-64.

6. Wolfe BL, LeMura LM, Cole PJ. Quantitative analysis of single- vs. multiple-set programs in resistance training. ***J Strength Cond Res*** 2004; 18: 35-47.
7. Rhea MR, Alvar BA, Burkett LN. Single versus multiple sets for strength: A meta-analysis to address the controversy. ***Res Quart Exer Sport*** 2002; 73: 485-88.
8. Peterson MD, Rhea MR, Alvar BA. Maximizing strength development in athletes: A meta-analysis to determine the dose-response relationship. ***J Strength Cond Res*** 2004; 18: 377-82.
9. Lipsey MW, Wilson DB. ***Practical Meta-Analysis*** Thousand Oaks, CA; Sage. 2001.
10. Klein DF. Flawed meta-analyses comparing psychotherapy with pharmacotherapy. ***Am J Psychiatry*** 2000; 157: 1204-11.
11. Lau J, Ioannidis LJ, Schmid CH. Quantitative synthesis in systematic reviews. ***Ann Intern Med*** 1997; 127: 820-26.
12. Shapiro S. Meta-analysis/Shmeta-analysis. ***Am J Epidemiol*** 1994; 9: 771-78.
13. Sterne JA, Schulz KF, Altman DG, Bartlett C., Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. ***Stat Med*** 2002; 21: 1513-24.
14. Glasziou PP, Sanders SL. Investigating causes of heterogeneity in systematic reviews. ***Stat Med*** 2002; 21: 1503-11.
15. Victor N. The challenge of meta-analysis: Discussion. Indications and contra-indications for meta-analysis. ***J Clin Epidemiol*** 1995; 48: 5-8.
16. Matt GE, Navarro AM. What meta-analyses have and have not taught us about psychotherapy effects: a review and future directions. ***Clin Psychol Rev*** 1997; 17: 1-32.
17. Flather MD, Farkouh ME, Pogue JM, Yusuf S. Strengths and limitations of meta-analysis: larger studies may be more reliable. ***Control Clin Trials*** 1997; 18: 568-79.
18. Spitzer WO. The challenge of meta-analysis. ***J Clin Epidemiol*** 1995; 48: 1-4.
19. Carpinelli RN, Otto RM, Winett RA. A critical analysis of the ACSM Position Stand on Resistance Training: Insufficient evidence to support recommended training protocols. ***JEPOnline***. 2004; 7: 1-60.
20. Chromiak JA, Smedley B, Carpenter W, Brown R, Koh YS, Lamberth JG Joe LA, Abadie BR, Altorfer G. Effect of a 10-week strength training program and recovery drink on body composition, muscular strength and endurance, and anaerobic power and capacity. ***Nutrition*** 2004; 20: 420-27.
21. Kraemer WJ. A series of studies—the physiological basis for strength training in American football: fact over philosophy. ***J Strength Cond Res*** 1997; 11: 131-42.
22. Kraemer WJ, Ratamess N, Fry AC Triplett-McBride T, Korzis LP, Bauer JA, Lynch JM, Fleck SJ. Influence of resistance training volume and periodization on physiological and performance adaptations in collegiate women tennis players. ***Am J Sports Med*** 2000; 28: 626-33.
23. Kramer JB, Stone MS, O'Bryant HS, Conley MS,Johnnson RL, Neiman DC, Honneycutt DR, Hoke TP. Effect of single vs. multiple sets of weight training: impact of volume, intensity, and variation. ***J Strength Cond Res*** 1997; 11: 143-47.
24. Schlumberger A, Stec J, Schmidtbleicher D. Single- vs. multiple-set strength training in women. ***J Strength Cond Res*** 2001; 15: 284-89.
25. Hass CJ Garzarella L, De Hoyos D, Pollock ML. Single versus multiple sets in long-term recreational weightlifters. ***Med Sci Sports Exerc*** 2000; 32: 235-42.
26. Ostrowski KJ, Wilson GJ, Weatherby R, Murphy PW, Little AD. The effect of weight training volume on hormonal output and muscular size and function. ***J Strength Cond Res*** 1997; 11: 148-54.
27. Rhea MR, Alvar BA, Ball SD, Burkett LN. Three sets of weight training superior to 1 set with equal intensity for eliciting strength. ***J Strength Cond Res*** 2002; 16: 525-9.
28. DeHoyos DV, Herring D, Garzarella L, Werber G, Brechue WF, Pollock ML. Effect of strength training volume on the development of strength and power in adolescent tennis players. ***Med Sci Sports Exerc*** 1997; 29 (suppl): S164.
29. DeHoyos D, Abe T, Garzarella L, Hass C, Nordman M, Pollock M. Effects of 6 months of high- or low-volume resistance training on muscular strength and endurance. ***Med Sci Sports Exerc*** 1998; 30 (Suppl): S165.

30. Hass CJ, Garzarella L, DeHoyos DV, Pollock ML. Effects of training volume on strength and endurance in experienced resistance trained adults. ***Med Sci Sports Exerc*** 1998; 30 (Suppl): S115.