Single vs. Multiple Sets of Resistance Exercise for Muscle Hypertrophy: A Meta-Analysis

James W. Krieger

Journal of Pure Power, Colorado Springs, CO

Abstract

Krieger, JW. Single vs. multiple sets of resistance exercise for muscle hypertrophy: a meta-analysis. *J Strength Cond Res* 24(4): 1150–1159, 2010–Previous meta-analyses have compared the effects of single to multiple sets on strength, but analyses on muscle hypertrophy are lacking. The purpose of this study was to use multilevel meta-regression to compare the effects of single and multiple sets per exercise on muscle hypertrophy. The analysis comprised 55 effect sizes (ESs), nested within 19 treatment groups and 8 studies. Multiple sets were associated with a larger ES than a single set (difference = 0.10 ± 0.04; confidence interval [CI]: 0.02, 0.19; $p = 0.016$). In a dose–response model, there was a trend for 2–3 sets per exercise to be associated with a greater ES than 1 set (difference = 0.09 ± 0.05; CI: −0.02, 0.20; $p = 0.09$), and a trend for 4–6 sets per exercise to be associated with a greater ES than 1 set (difference = 0.20 ± 0.11; CI: −0.04, 0.43; $p = 0.096$). Both of these trends were significant when considering permutation test p values ($p < 0.01$). There was no significant difference between 2–3 sets per exercise and 4–6 sets per exercise (difference = 0.10 ± 0.10; CI: −0.09, 0.30; $p = 0.29$). There was a tendency for increasing ESs for an increasing number of sets (0.24 for 1 set, 0.34 for 2–3 sets, and 0.44 for 4–6 sets). Sensitivity analysis revealed no highly influential studies that affected the magnitude of the observed differences, but one study did slightly influence the level of significance and CI width. No evidence of publication bias was observed. In conclusion, multiple sets are associated with 40% greater hypertrophy-related ESs than 1 set, in both trained and untrained subjects.

Key Words metaregression, effect size, lean body mass, volume

Introduction

Resistance training improves musculoskeletal strength, muscle mass, bone mass, and connective tissue thickness (22,41). The design of a resistance training program requires appropriate manipulation of numerous variables, including the frequency, intensity, and volume of the program (12). For general fitness purposes, the American College of Sports Medicine has recommended a program of 1 set of 8–10 exercises covering all major muscle groups (1). Multiple sets are recommended for athletic populations (21). In the past, some authors have argued that a single set per exercise is all that is necessary for all populations and that further gains are not achieved by successive sets (8). However, a large number of studies performed over the past decade have demonstrated greater strength gains with multiple sets per exercise (6,11,16,18–20,26–27,30,32,35–36). Also, a recent meta-analysis clearly showed multiple sets to be associated with 46% greater strength gains in both trained and untrained subjects (23).

The reason for the greater strength gains with multiple sets is not well established. Strength training is associated with both neural and structural adaptations that enhance force production (22). It is not clear whether the greater strength gains observed with multiple sets are because of greater neural adaptations, greater hypertrophy, or both. Some studies have shown greater hypertrophy with multiple sets (26,35), whereas others have not (11,27,30–32,40). Measures of muscle hypertrophy are highly variable and insensitive. Changes in muscle size are smaller and slower than changes in strength (28). Many resistance training studies are short in duration, and subject numbers tend to be small. Because of all these reasons, the risk of a Type II error is high. For example, McBride et al. (27) reported greater strength gains in a group performing 6 sets per exercise compared with a group performing 1 set per exercise. There were no significant differences in changes in lean mass between the groups. However, the study only lasted 12 weeks, and there were only 9 subjects per group. The mean change in leg lean mass was nonsignificantly greater in the multiple-set group compared with the single-set group (0.86 kg vs. −0.05 kg, respectively). A difference of 0.9 kg in 12 weeks is a meaningful difference in leg lean mass. Given the sample size

and the reported *SD*s, the estimated statistical power to detect this difference, using a 2-tailed test and an α of 0.05, is only 12%. Thus, only 12 of 100 studies would detect a significant difference, if each study only had 9 subjects per group. If this 0.9-kg difference represents a true difference between populations, then a Type II error has occurred. In fact, using an estimated *SD* of the difference, the study would need 75–175 subjects per group to detect this 0.9-kg difference with 80% power. Therefore, underpowered resistance training studies can potentially lead to incorrect conclusions regarding the effects of set volume on muscle hypertrophy, and these erroneous conclusions are only reinforced with the publication of more underpowered studies. Unfortunately, many resistance training studies do not report power analyses.

Another problem with determining the effects of set volume on hypertrophy is the many ways in which hypertrophy can be measured. Studies have used whole-body lean mass (11,26), regional lean mass (27,35), muscle thickness (31,40), muscle cross-sectional area (31,35), or muscle circumference (30–32) to measure hypertrophy. Different regions of a particular muscle may also be measured (40). Thus, comparisons across studies can be difficult. The calculation of a standardized effect size (ES) can aid in the comparison across studies (3). A meta-analysis of these ESs can allow for the identification of trends among conflicting and/or underpowered studies (45). Meta-analyses regarding the effects of set volume on strength have been published (23,33–34,44), but none of these analyses looked at measures of hypertrophy.

The purpose of this paper was to use meta-analysis to compare the effects of single and multiple sets per exercise on muscle hypertrophy. A second purpose was to establish a dose–response effect of set volume on hypertrophy. The hypothesis was that multiple sets would be associated with greater hypertrophy compared with single sets.

METHODS

Experimental Approach to the Problem

Studies comparing single with multiple sets per exercise, with all other variables being equivalent, were eligible for inclusion. This helped eliminate confounding effects of other training variables that may affect hypertrophy. To account for nonindependent ESs and the variation between between studies, between treatment groups, and between ESs within each treatment group, multilevel statistical models were used for the analysis. The dependent variable was the pre to posttraining change in muscle size. The primary independent variable was the number of sets per exercise.

Procedures

Study Selection. Searches were performed of PubMed, SPORT Discus, and CINAHL for English-language studies published between January 1, 1960, and October 15, 2009. A sample of keywords and phrases used in searches included "*resistance training*," "*strength training*," "*resistance exercise*," "*sets*," "*single*," "*multiple*," and "*volume*"; Boolean operators such as AND, OR, and NOT were used to help narrow searches. Hand searching and crossreferencing were performed from the bibliographies of previously retrieved studies and from review articles. Studies were selected if they met the following criteria: (a) resistance exercise program lasting a minimum of 4 weeks; (b) training on at least one exercise for at least one major muscle group; major muscle groups included the quadriceps, hamstrings, pectoralis major, latissimus dorsi, biceps, triceps, and deltoids; (c) adults ≥ 19 years; (d) comparison of single to multiple sets per exercise, with all other training variables being equivalent; (e) subjects free from orthopedic limitations that could affect progress on a resistance exercise program; (f) pre and posttraining determination of at least one measure of muscle hypertrophy; these measures included lean body mass, regional lean mass, muscle cross-sectional area, muscle circumference, and muscle thickness; (g) sufficient data to determine sets per exercise and to calculate ESs; and (h) published studies in English-language journals.

Data Abstraction. Data were tabulated onto a spreadsheet using Microsoft Excel (Microsoft Corp., Redmond, WA). Each row represented a specific ES for a treatment group. If there were multiple ESs for a particular treatment group (i.e., a treatment group was subjected to multiple measures of hypertrophy), then each ES was coded in a separate row.

Variables abstracted from each study were the following: authors, year, research design (randomized trial, nonrandomized trial, or randomized crossover), *n*, quality score, sex (male, female, or mixed), age (19–44 or ≥ 45 years), baseline body mass (kg), resistance exercise experience (<6 or ≥ 6 months), training program duration (weeks), average repetitions per set, training frequency ($d \cdot wk^{-1}$), sets per exercise, supervised training (yes/unspecified/no), pre and posttest means for hypertrophy measures, and pre and posttest *SD* for those measures. The study quality score was the sum of 2 scores used in previous reviews to rate the quality of resistance training studies: a 0–10 scale-based score used by Bågenhammar and Hansson (2) and a 0–10 scale-based score used by Durall et al. (10). For the average repetitions per set, if a range of repetitions was reported (e.g., 8–12 repetitions), the midpoint of the range was used (e.g., 10 repetitions).

For each measure of hypertrophy in each treatment group, an ES was calculated as the pretest–posttest change, divided by the pretest *SD* (29). Becker (3) recommended the ES for the control group be subtracted from the experimental group ES; however, numerous studies in this analysis did not include a control group. Because it is important to define the ES in a standard way across all studies (29), the control ES was assumed to be 0 in all studies and was not subtracted from the experimental ES. To test this assumption, the mean control ES was calculated among all studies that had a control group; the mean ES was 0.0 ± 0.03, which was not

TABLE 1. Studies included in the analysis.

References	Age (y)	Sex	Status*	Length (wks)	Type of measure	Sets	ES†	SV‡	Study ES§
Galvão and Taaffe (11)	≥45	M, F	U	20					0.03
$n = 16$					Lean body mass	1	0.05	0.021	
$n = 16$						3	0.08	0.012	
Marzolini et al. (26)	≥45	M, F	U	24					0.06
$n = 19$					Lean body mass	1	0.11	0.004	
$n = 18$						3	0.18	0.004	
McBride et al. (27)	19–44	M, F	U	12					0.12
$n = 9$					Leg LM‖	1	−0.01	0.040	
$n = 9$						3	0.23	0.033	
					Arm LM	1	0.00	0.040	
						3	−0.01	0.040	
Munn et al. (30)	19–44	M, F	U	6					0.02
$n = 23$					Arm circumference	1	0.24	0.006	
$n = 23$						1	0.09	0.006	
$n = 23$						3	0.31	0.006	
$n = 23$						3	0.07	0.005	
Ostrowski et al. (31)	19–44	M	T	10					0.20
$n = 9$					Tricep thickness	1	0.25	0.029	
$n = 9$						2	0.40	0.034	
$n = 9$						4	0.50	0.029	
					Rectus Femoris CSA¶	1	0.30	0.037	
						2	0.29	0.021	
						4	0.78	0.027	
					Rectus Femoris Thickness	1	0.27	0.040	
						2	0.14	0.035	
						4	0.73	0.027	
Rhea et al. (32)	19–44	M	T	12					0.44
$n = 8$					Chest circumference	1	0.50	0.015	
$n = 8$						3	0.62	0.012	
					Leg circumference	1	−0.13	0.043	
						3	0.64	0.059	
Rønnestad et al. (35)	19–44	M	U	11					0.24
$n = 10$					Quadriceps CSA	1	0.56	0.032	
$n = 11$						3	0.67	0.026	
					Hamstring CSA	1	0.27	0.032	
						3	0.56	0.026	
$n = 5$					Lower body LM	1	0.14	0.189	
$n = 5$						3	0.36	0.189	
$n = 5$					Upper body LM	1	0.25	0.189	
$n = 5$						3	0.45	0.189	
$n = 11$					Trapezius CSA	1	0.28	0.026	
$n = 10$						3	0.68	0.032	
Starkey et al. (40)	19–44	M, F	U	14					0.07
					Thigh thickness 20% Anterior				
$n = 18$						1	0.05	0.008	
$n = 20$						3	0.15	0.008	
					40% Anterior	1	0.14	0.009	
						3	0.13	0.007	
					60% Anterior	1	0.14	0.009	
						3	0.15	0.008	
					Medialis	1	0.15	0.010	
						3	0.14	0.007	
					Lateralis	1	−0.07	0.006	
						3	−0.08	0.008	
					20% Lateral	1	0.03	0.009	
						3	0.37	0.003	

40% Lateral	1	0.10	0.009
	3	0.21	0.007
60% Lateral	1	0.25	0.009
	3	0.22	0.008
40% Posterior	1	0.30	0.009
	3	0.43	0.007
60% Posterior	1	0.29	0.010
	3	0.34	0.008

*U = Untrained (<6 mo resistance exercise experience); T = Trained (≥6 mo experience).
†Effect size.
‡Sampling variance.
§Mean study-level effect size (mean multiple-set ES − mean single-set ES).
‖Lean mass.
¶Cross-sectional area.

significantly different from 0 ($p = 0.94$) when compared using a one-sample t-test. The sampling variance for each ES was estimated according to Morris and DeShon (29). Calculation of the sampling variance required an estimate of the population ES and the pretest–posttest correlation for each individual ES. The population ES was estimated by calculating the mean ES across all studies and treatment groups (29). The pretest–posttest correlation was calculated using the following formula (29):

$$r = (s_1^2 + s_2^2 - s_D^2)/(2s_1 s_2),$$

where s_1 and s_2 are the *SD*s for the pre and posttest means, respectively, and s_D is the *SD* of the difference scores. Where s_2 was not reported, s_1 was used in its place. Where s_D was not reported, it was estimated using the following formula:

$$s_D = \sqrt{((s_1^2/n) + (s_2^2/n))}.$$

Statistical Analyses

Meta-analyses were performed using multilevel linear mixed models, modeling the variation between studies as a random effect, the variation between treatment groups as a random effect nested within studies, the variation between ESs as a random effect nested within treatment groups, and group-level predictors as fixed effects (15). The within-group variances were assumed known. Observations were weighted

TABLE 2. Full model with all covariates.

Predictor	Coefficient ± *SE**	95% CI	*p* Value
Multiple sets per exercise			
No	0		
Yes	0.11 ± 0.04	(0.02, 0.19)	0.016
Intercept†	0.45 ± 0.10	(0.26, 0.64)	<0.0001
Sex			
M	0		
M, F	−0.31 ± 0.08	(−0.53, −0.09)	0.017
Training duration (wk)	0.00 ± 0.01	(−0.02, 0.01)	0.70
Training experience			
<6 mos	0		
≥6 mos	−0.06 ± 0.09	(−0.31, 0.19)	0.54

*Positive values for coefficients represent an increase in overall effect size (ES). Negative values represent a decrease in overall ES. Coefficients of 0 represent the default categories in the model. Coefficients for other categories within the same variable represent the difference from the default category.
†Intercept of the model produced by hierarchical regression.

TABLE 3. Reduced model.

Predictor	Coefficient ± SE*	95% CI	p Value	Permutation p value
Multiple sets per exercise				
No	0			
Yes	0.10 ± 0.04	(0.02, 0.19)	0.016	0.0
Intercept†	0.39 ± 0.05	(0.29, 0.49)	<0.0001	N/A‡
Sex				
M	0			
M, F	−0.28 ± 0.05	(−0.40, −0.16)	0.0012	0.0

*Positive values for coefficients represent an increase in overall effect size (ES). Negative values represent a decrease in overall ES. Coefficients of 0 represent the default categories in the model. Coefficients for other categories within the same variable represent the difference from the default category.

†Intercept of the model produced by hierarchical regression.

‡N/A = Not available; permutation p values were only calculated for covariates.

by the inverse of the sampling variance (29). An intercept-only model was created, estimating the weighted mean ES across all studies and treatment groups. A full statistical model was then generated. Because of the small number of studies identified for this analysis (Table 1), the number of predictors that could be included in the full statistical model was small. A binary variable (multiple or single sets) was included as a predictor in the model. Other predictors chosen for the full model were based on predictors observed to show weak relationships ($p < 0.30$) to strength in a previous metaregression that used an identical statistical model (23). The predictors selected were sex, training duration, and training experience. Although age showed a weak effect ($p = 0.24$) in the previous metaregression (23), it was not chosen for this model as 6 of the 8 studies in this analysis involved subjects <44 years of age. The full model was then reduced by removing one predictor at a time, starting with the most insignificant predictor (7). The final model represented the reduced model with the lowest Bayesian information criterion (BIC) (37) and that was not significantly different ($p > 0.05$) from the full model when compared with a likelihood ratio test (LRT). Model parameters were estimated by the method of restricted maximum likelihood (REML) (43); an exception was during the model reduction process, in which parameters were estimated by the method of

maximum likelihood (ML), as LRTs cannot be used to compare nested models with REML estimates. Denominator df for statistical tests and confidence intervals (CIs) were calculated according to Berkey et al. (5) The multiple-sets predictor was not removed during the model reduction process. Because metaregression can result in inflated false-positive rates when heterogeneity is present and/or when there are few studies (13), a permutation test described by Higgins and Thompson (13) was used to verify the significance of the predictors in the final model; 1,000 permutations were generated. To examine the relationship between set volume and treatment effect, a dose–response model was created by replacing the multiple-sets predictor with a categorical predictor representing the number of sets
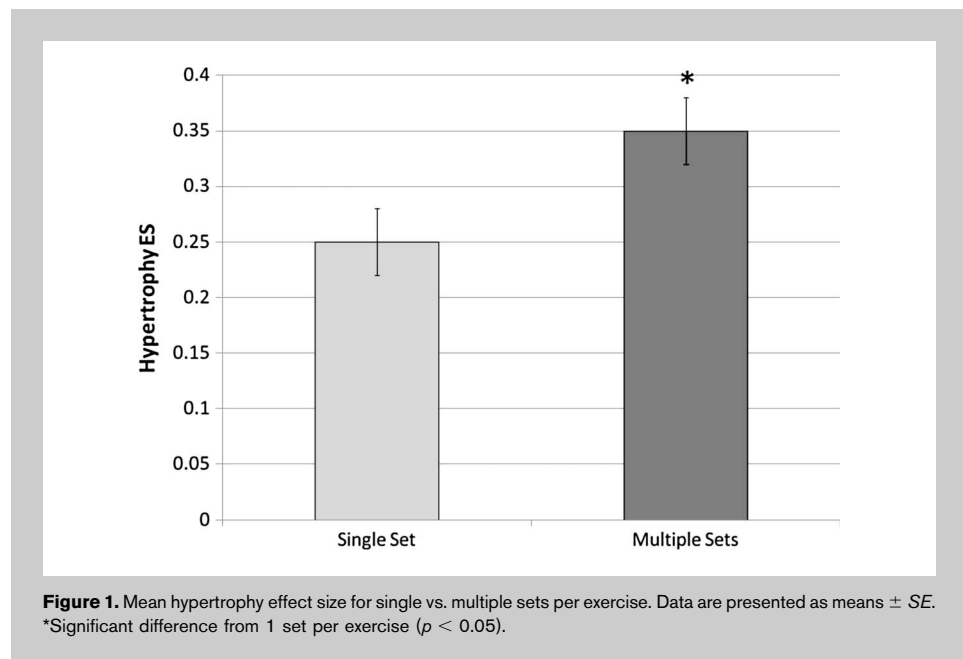

Figure 1. Mean hypertrophy effect size for single vs. multiple sets per exercise. Data are presented as means ± SE. *Significant difference from 1 set per exercise ($p < 0.05$).

performed per exercise: 1 set, 2–3 sets, and 4–6 sets. Adjustment for post hoc multiple comparisons among set categories were performed with a Hochberg correction (14). Histograms of residuals were examined to identify major departures from normality; no departures from normality were found. Publication bias was assessed via a funnel plot regression method described by Macaskill et al. (25).

To identify the presence of highly influential studies that may have biased the analysis, a sensitivity analysis was carried out by removing one study at a time and then examining the multiple-sets predictor. Studies were identified as influential if their removal resulted in

Figure 2. Dose–response effect of set volume on hypertrophy. Data are presented as means $\pm$ *SE*. ES = effect size. *Trend toward difference from 1 set per exercise according to Hochberg-adjusted standard *p* value ($p < 0.10$). †Significantly different from 1 set per exercise according to Hochberg-adjusted permutation *p* value ($p < 0.01$).

a change of >1 *SE* in the multiple-sets coefficient. All analyses were performed using S-PLUS version 8.0 (Insightful, Seattle, WA). Effects were considered significant at $p \leq 0.05$. Trends were declared at $p \leq 0.10$. Data are reported as means ($\pm SEs$) and 95% confidence intervals (CIs).

RESULTS

Study Characteristics
The analysis comprised 55 ESs, nested within 19 treatment groups and 8 studies (Table 1). The weighted mean ES across all studies and treatment groups was 0.25 ± 0.06 (CI: 0.13, 0.37).

Full Model
Results for the full model with all predictors are shown in Table 2. There was a significant effect of sets per exercise while controlling for all other covariates, with multiple
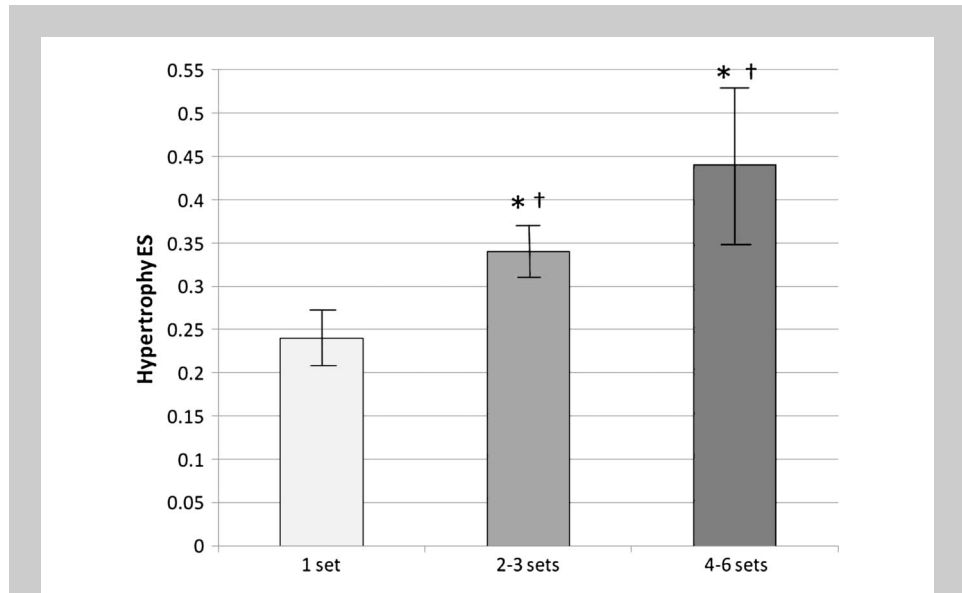
sets being associated with a larger ES than a single set (difference = 0.11 ± 0.04; CI: 0.02, 0.19; $p = 0.016$).

Reduced Model
Results for the reduced model are shown in Table 3. After the model reduction procedure, only the sex (male or mixed) of the treatment groups remained as a covariate. The BIC decreased from 8.9 in the full model to -9.7 in the reduced model. The reduced model was not significantly different from the full model ($p = 0.73$). In the reduced model, multiple sets were associated with a larger ES than a single set (difference = 0.10 ± 0.04; CI: 0.02, 0.19; $p = 0.016$; Table 3). The mean ES for a single set was 0.25 ± 0.03 (CI: 0.18, 0.32; Figure 1). The mean ES for multiple sets was 0.35 ± 0.03 (CI: 0.29, 0.41; Figure 1).

TABLE 4. Sensitivity analysis.

Study removed	Coefficient*	95% CI	p Value	Permutation p value
None	0.10 ± 0.04	(0.02, 0.19)	0.016	0.0
Galvão and Taaffe (11)	0.11 ± 0.04	(0.02, 0.19)	0.014	0.0
Marzolini et al. (26)	0.11 ± 0.04	(0.02, 0.20)	0.019	0.0
McBride et al. (27)	0.10 ± 0.04	(0.02, 0.19)	0.020	0.001
Munn et al. (30)	0.12 ± 0.05	(0.02, 0.22)	0.016	0.001
Ostrowski et al. (31)	0.10 ± 0.04	(0.01, 0.18)	0.028	0.0
Rhea et al. (32)	0.09 ± 0.04	(0.01, 0.18)	0.033	0.001
Rønnestad et al. (35)	0.09 ± 0.04	(-0.01, 0.19)	0.062	0.001
Starkey et al. (40)	0.12 ± 0.06	(0.00, 0.25)	0.043	0.001

*Coefficient $\pm$ *SE*. Value represents difference in effect size (ES) between single and multiple sets per exercise.
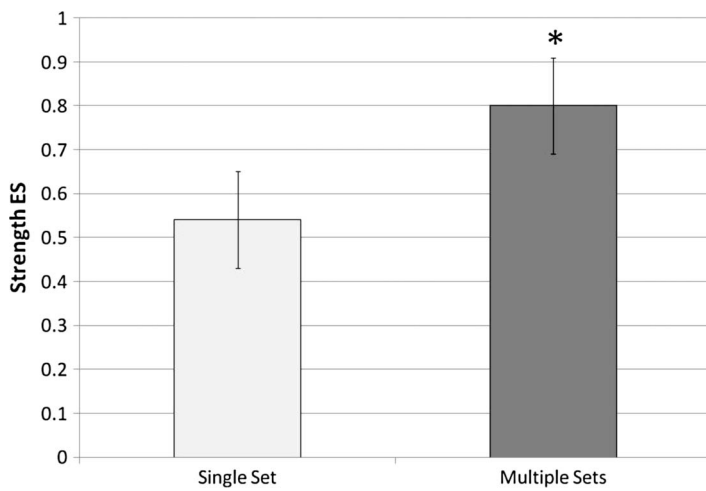
Figure 3. Mean strength effect size for single vs. multiple sets per exercise from Krieger (23). Note similarity to hypertrophy response in Figure 1. Data are presented as means ± *SE*. *Significant difference from 1 set per exercise ($p < 0.05$).
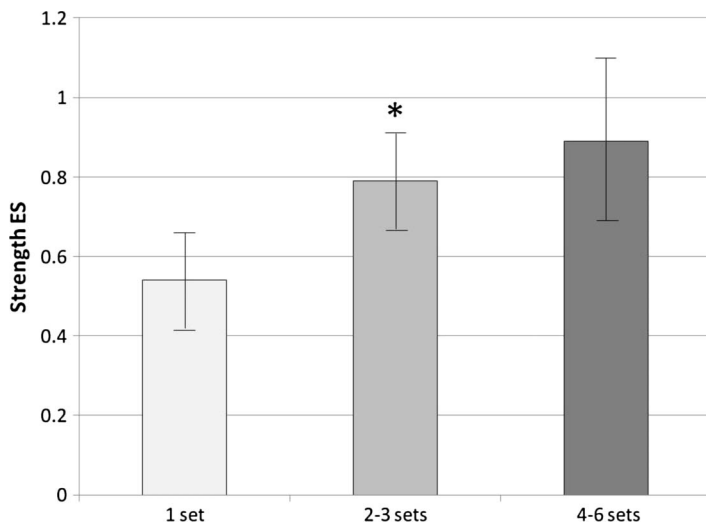

Figure 4. Dose–response effect of set volume on strength from Krieger (23). Note similarity to dose–response effect for hypertrophy in Figure 2. Data are presented as means ± *SE*. ES = effect size. *Significantly different from 1 set per exercise ($p < 0.001$).

Dose–Response Model

In the dose–response model, there was a trend for 2–3 sets per exercise to be associated with a greater ES than 1 set per exercise (difference = 0.09 ± 0.05; CI: −0.02, 0.20; $p = 0.09$). The difference was significant when considering the Hochberg-adjusted permutation test p value ($p = 0.009$). There was also a trend for 4–6 sets per exercise to be associated with a greater ES compared with 1 set per exercise (difference = 0.20 ± 0.11; CI: −0.04, 0.43; $p = 0.096$). The difference was significant when considering the Hochberg-adjusted permutation test p value ($p = 0.008$). There was no significant difference between 2–3 sets per exercise and 4–6 sets per exercise (difference = 0.10 ± 0.10; CI: −0.09, 0.30; $p = 0.29$). There was a tendency for increasing ESs for an increasing number of sets. The mean ES for 1-set per exercise was 0.24 ± 0.03 (CI: 0.18, 0.31; Figure 2). The mean ES for 2–3 sets per exercise was 0.34 ± 0.03 (CI: 0.27, 0.41; Figure 2). The mean ES for 4–6 sets per exercise was 0.44 ± 0.09 (CI: 0.26, 0.62; Figure 2).

Sensitivity Analysis

Results for the sensitivity analysis are reported in Table 4. The difference in ES between single and multiple sets was not affected by more than 1*SE* for any study removed. However, the removal of the study by Rønnestad et al. (35) changed the p value from 0.016 to 0.06. The CI was widened to (−0.01, 0.19). The p value from the permutation test remained significant ($p = 0.001$).

Publication Bias

There was no significant relationship between treatment effect and sample size (slope of line = $−0.002 \pm 0.002$; $p = 0.32$), indicating no evidence of publication bias.

DISCUSSION

The purpose of this meta-analysis was to determine whether multiple sets per exercise are associated with greater muscle hypertrophy than a single set per exercise in a resistance training program. Multiple sets per exercise were associated with significantly greater ESs in both the full and reduced statistical models. The mean ES for a single set per exercise was 0.25, whereas the mean ES for multiple sets was 0.35. Thus, multiple sets were associated with 40% greater hypertrophy-related ESs than a single set. According to

Cohen's classifications for ESs (<0.41 = small; 0.41–0.70 = moderate; >0.70 = large) (9), both estimates are consistent with small treatment effects. In a previous meta-analysis on strength using an identical statistical model (23), 1 set per exercise was associated with a moderate treatment effect (mean ES = 0.54), whereas multiple sets were associated with a large treatment effect (mean ES = 0.80; Figure 3). The differences in ES estimates for strength vs. hypertrophy are consistent with the observation that changes in muscle size are often smaller and slower than changes in strength (28), particularly in untrained subjects (6 of the 8 studies in the current analysis involved untrained subjects). The observed ES difference for sex (a decrease of 0.28 for mixed groups compared with male groups) is consistent with the observation that women experience smaller changes in muscle size compared with men (17).

In a previous meta-analysis on strength using an identical statistical model, a 46% greater ES was observed for multiple sets compared with single sets (23) (Figure 3). A 40% greater ES was observed in this study. This indicates that the greater strength gains observed with multiple sets are in part because of greater muscle hypertrophy. It is known that mechanical loading stimulates protein synthesis in skeletal muscle (39), and increasing loads result in greater responses until a plateau is reached (24). It is likely that protein synthesis responds in a similar manner to the number of sets (i.e., an increasing response as the number of sets are increased, until a plateau is reached), although there is no research examining this. The results of this study support this hypothesis; there was a trend for an increasing ES for an increasing number of sets. The response appeared to start to level off around 4–6 sets, as the difference between 2–3 sets and 4–6 sets was smaller than the difference between 1 set and 2–3 sets. Also, the difference between 1 set and 2–3 sets was nearly significant (and the permutation test p value was significant), whereas the difference between 2–3 sets and 4–6 sets was not. However, only 2 studies in this analysis involved 4–6 sets per exercise. Thus, the statistical power to detect differences is low, and definitive conclusions cannot be made. These results are similar to a previous meta-analysis on strength, where there was an increasing response to an increasing number of sets, with an apparent plateau around 4–6 sets per exercise (23) (Figure 4).

It has been proposed that the majority of initial strength gains in untrained subjects are because of neural adaptations rather than hypertrophy (28). The results of this analysis suggest that some of the initial strength gains are because of hypertrophy. Given the insensitivity and variability of hypertrophy measurements, it is likely that hypertrophy occurs in untrained subjects but is difficult to detect. This is supported by research that shows increases in protein synthesis in response to resistance training in untrained subjects (24). Recent evidence also shows measurable hypertrophy after only 3 weeks of resistance exercise (38).

To examine the effects of potential outliers on the outcome, a sensitivity analysis was performed. The magnitude of the difference between single and multiple sets was consistent regardless of which study was removed. However, the removal of the study by Rønnestad et al. (35) affected the width of the CI, and the significant effect of multiple sets turned into a strong trend. However, this is likely because of loss of statistical power, given that the magnitude of the estimate remained similar, the permutation test p value remained significant, and the analysis consisted of only 8 studies.

Publication bias represents the problem where studies showing statistically significant results are more likely to be published than studies that fail to show significant results (e.g., studies showing a significant difference between 1 set and multiple sets per exercise may be more likely to be published) (4). Thus, meta-analyses of published studies may overestimate the magnitude of a treatment effect (4). Analyses can be performed to detect the presence of publication bias; one analysis involves examining the relationship between sample size and treatment effect (25). The existence of a significant relationship suggests that publication bias may be present. However, no such relationship was observed in the current study. Two previous meta-analyses on the effects of multiple vs. single sets on strength also failed to observe any evidence of publication bias (23,44). Also, only 2 of the 8 studies in this analysis reported significant differences in hypertrophy-related measures when comparing single with multiple sets (26,35). This strongly suggests that publication bias is not present, because if it were, most of the studies would report significant differences. In fact, even though only 2 of the 8 studies reported significant differences, the mean study-level ES favored the multiple-set group in all 8 studies (Table 1). This indicates that many of these studies are underpowered to detect differences.

There are a number of strengths to the current study design. First, strict inclusion criteria were used; only studies comparing single with multiple sets while holding all other variables constant were included. Second, the multilevel model allowed for the simultaneous modeling of the variation between studies, between treatment groups, and between ESs within each treatment group. Third, both standard and permutation test p values were used to protect against spurious findings, a common problem with metaregression (13). Finally, a sensitivity analysis was performed, and this indicated the mean difference between single and multiple sets to be reasonably consistent across the removal of individual studies.

A primary limitation of this analysis is the small number of studies. Thus, the statistical power of the analysis is limited. This was evident as the removal of the study by Rønnestad et al. (35) affected the p value and CIs. This was also evident by the observed trends that did not quite reach statistical significance (although they were significant according to permutation tests). The small number of studies also limited

the number of predictors that could be included in the statistical model. Thus, interactions between set volume and factors such as training experience could not be explored, as had been done in a previous meta-analysis on strength (23). Also, the majority of studies in this analysis compared 1 set with 3 sets per exercise; only 2 studies in this analysis incorporated ≥4 sets per exercise. This limits the statistical power to compare 3 sets with greater set volumes, as the *SE* for the 4–6 set category was large. Given that the ES for 4–6 sets (0.44) is considered a moderate effect, whereas the ES for 2–3 sets (0.34) is considered a small effect according to Cohen's classifications (9), more research involving ≥4 sets is needed to clarify whether this is a chance difference or a true difference. Another limitation is that meta-regression, like epidemiological research, can only support observational associations and cannot demonstrate causation (42). A final limitation is the availability of data (42). Some studies, despite meeting the design criteria (comparison of single vs. multiple sets while keeping other variables constant), were excluded because hypertrophy was not measured. Because an analysis can only be undertaken for trials where all information is available, bias can be introduced in the results (42). However, most of the excluded studies reported greater strength gains in the multiple-set groups. Given the relationship between strength and muscle size, the consistency of the mean difference during the sensitivity analysis, the fact that the study-level ES favored the multiple-set group in all 8 studies, and the lack of evidence of publication bias, it is unlikely that the addition of more studies would alter the results, other than improving statistical power.

PRACTICAL APPLICATIONS

Multiple sets per exercise were associated with significantly greater changes in muscle size than a single set per exercise during a resistance exercise program. Specifically, hypertrophy-related ESs were 40% greater with multiple sets compared with single sets. This was true regardless of subject training status or training program duration. There was a trend for an increasing hypertrophic response to an increasing number of sets. Thus, individuals interested in achieving maximal hypertrophy should do a minimum of 2–3 sets per exercise. It is possible that 4–6 sets could give an even greater response, but the small number of studies incorporating volumes of ≥4 sets limits the statistical power and the ability to form any definitive conclusions. If time is a limiting factor, then single sets can produce hypertrophy, but improvements may not be optimal. More research is necessary to compare the effects of 2–3 sets per exercise to ≥4 sets. Future research should also focus on the effects of resistance training volume on protein synthesis and other cellular and molecular changes that may impact hypertrophy. Finally, resistance training studies comparing hypertrophic responses between treatments should include sufficient numbers of subjects to obtain adequate statistical power to detect differences; studies should also report power analyses.

REFERENCES

1. American College of Sports Medicine Position Stand. The recommended quantity and quality of exercise for developing and maintaining cardiorespiratory and muscular fitness, and flexibility in healthy adults. *Med Sci Sports Exerc* 30: 975–991, 1998.

2. Bågenhammar, S and Hansson, EE. Repeated sets or single set of resistance training: A systematic review. *Adv Physiother* 9: 154–160, 2007.

3. Becker, BJ. Synthesizing standardized mean-change measures. *Br J Math Stat Psychol* 41: 257–278, 1988.

4. Begg, CB and Berlin, JA. Publication bias and dissemination of clinical research. *J Natl Cancer Inst* 81: 107–115, 1989.

5. Berkey, CS, Hoaglin, DC, Mosteller, F, and Colditz, GA. A random-effects regression model for meta-analysis. *Stat Med* 14: 395–411, 1995.

6. Borst, SE, De Hoyos, DV, Garzarella, L, Vincent, K, Pollock, BH, Lowenthal, DT, and Pollock, ML. Effects of resistance training on insulin-like growth factor-I and IGF binding proteins. *Med Sci Sports Exerc* 33: 648–653, 2001.

7. Burnham, KP and Anderson, DR. *Model Selection and Inference: A Practical Information-Theoretic Approach*. New York, NY: Springer-Verlag, 2002.

8. Carpinelli, RN and Otto, RM. Strength training: single versus multiple sets. *Sports Med* 26: 73–84, 1998.

9. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.

10. Durall, CJ, Hermsen, D, and Demuth, C. Systematic review of single-set versus multiple-set resistance-training randomized controlled trials: implications for rehabilitation. *Crit Rev Phys Rehab Med* 18: 107–116, 2006.

11. Galvão, DA and Taaffe, DR. Resistance exercise dosage in older adults: single- versus multiset effects on physical performance and body composition. *J Am Geriatr Soc* 53: 2090–2097, 2005.

12. Hass, CJ, Feigenbaum, MS, and Franklin, BA. Prescription of resistance training for healthy populations. *Sports Med* 31: 953–964, 2001.

13. Higgins, JT and Thompson, SG. Controlling the risk of spurious findings from meta-regression. *Stat Med* 23: 1663–1682, 2004.

14. Hochberg, Y. A sharper Bonferonni procedure for multiple tests of significance. *Biometrika* 75: 800–802, 1988.

15. Hox, JJ and De Leeuw, ED. Multilevel models for meta-analysis. In: *Multilevel Modeling Methodological Advances, Issues, and Applications.* Reise, SP and Duan, N, eds. Mahwah, NJ: Lawrence Erlbaum Associates, 2003. pp. 90–111.

16. Humburg, H, Baars, H, Schröder, J, Reer, R, and Braumann, K-M. 1-set vs. 3-set resistance training: A crossover study. *J Strength Cond Res* 21: 578–582, 2007.

17. Ivey, FM, Roth, SM, Ferrell, RE, Tracy, BL, Lemmer, JT, Hurlbut, DE, Martel, GF, Siegel, EL, Fozard, JL, Metter, Je, Fleg, JL, and Hurley, BF. Effects of age, gender, and myostatin genotype on the hypertrophic response to heavy resistance strength training. *J Gerontol A Biol Sci Med Sci* 55: M641–M648, 2000.

18. Kelly, SB, Brown, LE, Coburn, JW, Zinder, SM, Gardner, LM, and Nguyen, D. The effect of single versus multiple sets on strength. *J Strength Cond Res* 21: 1003–1006, 2007.

19. Kemmler, WK, Lauber, D, Engelke, K, and Weineck, J. Effects of single- vs. multiple-set resistance training on maximum strength and body composition in trained postmenopausal women. *J Strength Cond Res* 18: 689–694, 2004.

20. Kraemer, WJ. The physiological basis for strength training in American football: Fact over philosophy. *J Strength Cond Res* 11: 131–142, 1997.

21. Kraemer, WJ, Adams, K, Cafarelli, E, Dudley, GA, Dooly, C, Feigenbaum, MS, Fleck, SJ, Franklin, B, Fry, AC, Hoffman, JR, Newton, RU, Potteiger, J, Stone, MH, Ratamess, NA, and Triplett-McBride, T. American College of Sports Medicine position stand. Progression models in resistance training for healthy adults. *Med Sci Sports Exerc* 34: 364–380, 2002.

22. Kraemer, WJ, Ratamess, NA, and French, DN. Resistance training for health and performance. *Curr Sports Med Rep* 1: 165–171, 2002.

23. Krieger, JW. Single versus multiple sets of resistance exercise: A meta-regression. *J Strength Cond Res* 23: 1890–1901, 2009.

24. Kumar, V, Selby, A, Rankin, D, Patel, R, Atherton, P, Hildebrandt, W, Williams, J, Smith, K, Seynnes, O, Hiscock, N, and Rennie, MJ. Age-related differences in the dose–response relationship of muscle protein synthesis to resistance exercise in young and old men. *J Physiol* 587: 211–217, 2009.

25. Macaskill, P, Walter, SD, and Irwig, I. A comparison of methods to detect publication bias in meta-analysis. *Stat Med* 20: 641–654, 2001.

26. Marzolini, S, Oh, PI, Thomas, SG, and Goodman, JM. Aerobic and resistance training in coronary disease: single versus multiple sets. *Med Sci Sports Exerc* 40: 1557–1564, 2008.

27. McBride, JM, Blaak, JB, and Triplett-McBride, T. Effect of resistance exercise volume and complexity on EMG, strength, and regional body composition. *Eur J Appl Physiol* 90: 626–632, 2003.

28. Moritani, T and deVries, HA. Neural factors versus hypertrophy in the time course of muscle strength gains. *Am J Phys Med* 58: 115–130, 1979.

29. Morris, SB and Deshon, RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychol Methods* 7: 105–125, 2002.

30. Munn, J, Herbert, RD, Hancock, MJ, and Gandevia, SC. Resistance training for strength: effect of number of sets and contraction speed. *Med Sci Sports Exerc* 37: 1622–1626, 2005.

31. Ostrowski, KJ, Wilson, GJ, Weatherby, R, Murphy, PW, and Lyttle, AD. The effect of weight training volume on hormonal output and muscular size and function. *J Strength Cond Res* 11: 148–154, 1997.

32. Rhea, MR, Alvar, BA, Ball, SD, and Burkett, LN. Three sets of weight training superior to 1 set with equal intensity for eliciting strength. *J Strength Cond Res* 16: 525–529, 2002.

33. Rhea, MR, Alvar, BA, and Burkett, LN. Single versus multiple sets for strength: A meta-analysis to address the controversy. *Res Q Exerc Sport* 73: 485–488, 2002.

34. Rhea, MR, Alvar, BA, Burkett, LN, and Ball, SD. A meta-analysis to determine the dose response for strength development. *Med Sci Sports Exerc* 35: 456–464, 2003.

35. Rønnestad, BR, Egeland, W, Kvamme, NH, Refsnes, PE, Kadi, F, and Raastad, T. Dissimilar effects of one- and three-set strength training on strength and muscle mass gains in upper and lower body in untrained subjects. *J Strength Cond Res* 21: 157–163, 2007.

36. Schlumberger, A, Stec, J, and Schmidtbleicher, D. Single- vs. multiple-set strength training in women. *J Strength Cond Res* 15: 284–289, 2001.

37. Schwarz, G. Estimating the dimension of a model. *Ann Stat* 6: 461–464, 1978.

38. Seynnes, OR, de Boer, M, and Narici, MV. Early skeletal muscle hypertrophy and architectural changes in response to high-intensity resistance training. *J Appl Physiol* 102: 368–373, 2007.

39. Spangenburg, EE. Changes in muscle mass with mechanical load: possible cellular mechanisms. *Appl Physiol Nutr Metab* 34: 328–335, 2009.

40. Starkey, DB, Pollock, ML, Ishida, Y, Welsch, MA, Brechue, WF, Graves, JE, and Feigenbaum, MS. Effect of resistance training volume on strength and muscle thickness. *Med Sci Sports Exerc* 28: 1311–1320, 1996.

41. Stone, MH. Implications for connective tissue and bone alterations resulting from resistance exercise training. *Med Sci Sports Exerc* 20: S162–S168, 1988.

42. Thompson, SG and Higgins, JT. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 21: 1559–1573, 2002.

43. Thompson, SG and Sharp, SJ. Explaining heterogeneity in meta-analysis: A comparison of methods. *Stat Med* 18: 2693–2708, 1999.

44. Wolfe, BL, Lemura, LM, and Cole, PJ. Quantitative analysis of single- vs. multiple set programs in resistance training. *J Strength Cond Res* 18: 35–47, 2004.

45. Zwahlen, M, Renehan, A, and Egger, M. Meta-analysis in medical research: Potentials and limitations. *Urol Oncol* 26: 320–329, 2008.